

VAiRoma: A Visual Analytics System for Making Sense of Places, Times, and Events in Roman History

Isaac Cho, Wewnen Dou, Derek Xiaoyu Wang, Eric Sauda and William Ribarsky

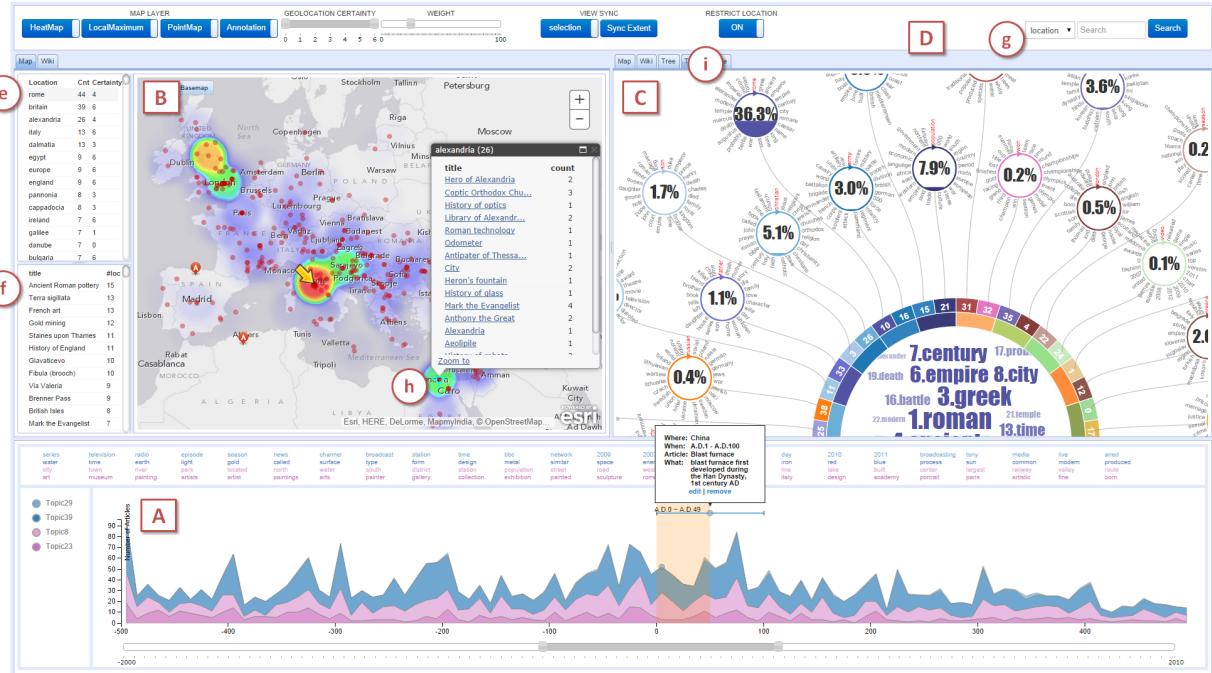


Fig. 1. Overview of VAiRoma Interface. The interface has three main views: Timeline view (A), Geographic view (B) and Topic view (C). A user-generated annotation is shown in the Timeline view.

Abstract—Learning and gaining knowledge of Roman history is an area of interest for students and citizens at large. This is an example of a subject with great sweep (with many interrelated sub-topics over, in this case, a 3,000 year history) that is hard to grasp by any individual and, in its full detail, is not available as a coherent story. In this paper, we propose a visual analytics approach to construct a data driven view of Roman history based on a large collection of Wikipedia articles. Extracting and enabling the discovery of useful knowledge on events, places, times, and their connections from large amounts of textual data has always been a challenging task. To this aim, we introduce VAiRoma, a visual analytics system that couples state-of-the-art text analysis methods with an intuitive visual interface to help users make sense of events, places, times, and more importantly, the relationships between them. VAiRoma goes beyond textual content exploration, as it permits users to compare, make connections, and externalize the findings all within the visual interface. As a result, VAiRoma allows users to learn and create new knowledge regarding Roman history in an informed way. We evaluated VAiRoma with 16 participants through a user study, with the task being to learn about roman piazzas through finding relevant articles and new relationships. Our study results showed that the VAiRoma system enables the participants to find more relevant articles and connections compared to Web searches and literature search conducted in a roman library. Subjective feedback on VAiRoma was also very positive. In addition, we ran two case studies that demonstrate how VAiRoma can be used for deeper analysis, permitting the rapid discovery and analysis of a small number of key documents even when the original collection contains hundreds of thousands of documents.

Index Terms—Visual Analytics, Text Analytics, Wikipedia

1 INTRODUCTION

Learning and gaining knowledge on places and events related to Roman history is a common interest shared by multiple cohorts of users

including students and interested citizens at large. For example, students in an architecture class are interested in learning about Roman cities, buildings, and piazzas through textual sources, since in addition to the architectural structures that make the places important, key events that occurred at these places throughout history enhance the understanding of their use, spatial meaning, and development. Another cohort of users are citizens who are interested in identifying relevant textual sources and developing useful knowledge on places, events, certain time periods during Roman history, and more importantly, the relationships between them (which may be hard to pull out in literature that either concentrates on one or another aspect of the history or

- Isaac Cho is with UNC Charlotte. E-mail: icho1@uncc.edu.
- Wenwen Dou is with UNC Charlotte. E-mail: wdou1@uncc.edu.

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvvcg@computer.org. Digital Object Identifier no. 10.1109/TVCG.2015.2467971

discusses it at too high a level).

Within the realm of Roman history, the task of identifying relevant textual information on places, time periods, people, and events is challenging since much information on locations and times as well as their relationships are described in many separate texts with no clear linkage or structure. For instance, the Roman Empire and Han Dynasty existed during the same era as the two most powerful empires in the western and eastern parts of the world. Relationships and comparisons between the two empires appear in many Wikipedia articles with each one covering one aspect of the relationship. Yet there is no simple way to gather all the relevant wiki articles, especially when the relationship between the two empires was only mentioned in one section of the article. This is one of many examples highlighting the need for organizing and representing the textual information in a way that can lead to the identification of previously unseen connections.

To this aim, we propose a visual analytics approach to help people gain knowledge through making connections on places, times, and events. The application focuses on constructing a narrative of the whole Roman history from ancient times, through the Empire, to modern times. The narrative is established through connecting the geolocations (ranging from countries to cities to significant landmarks), time information (ancient and modern dates), and topics extracted from the Wikipedia textual collection and representing them in an interactive web-based visual interface. Note that the locations, time, and topic information are all automatically extracted from the content of the Wikipedia articles through our data analytics process. As a result, in geographic terms, users can explore events and associated topics occurring in the Mediterranean and Western Europe during the time of Roman domination, but then they can follow these regions as they develop towards modern states. More importantly, users can discover previously unseen connections between space, time, and events thus generating new knowledge. To help users externalize and record the knowledge and insights gained from interacting with VAIroma, we have carefully designed and implemented annotation support within the system. Not only do we provide an annotation template to facilitate the knowledge construction; we also store users' annotations so that they can later access, share, and update previous findings.

As illustrated in our user study, VAIroma enables participants to find more relevant connections and relationships when compared to online search. The user study provides evidence that through extracting, organizing and visually representing information on place, time, and events in an associative way, VAIroma enables users to discover connections and relationship they would not find otherwise. The main contributions of the paper are the followings:

- We present a data-driven + visual analytics approach to construct a narrative on Roman history using the content of Wikipedia articles. VAIroma exceeds textual content exploration alone, as it permits users to compare, make connections, and externalize the findings all within the visual interface. VAIroma goes beyond previous systems in supporting a good part of the updated sensemaking model [34] for visual analytics systems.
- We design VAIroma for educated members of the general public who are interested in gaining knowledge on Roman history. It is implemented as a Web-based user interface for easy dissemination. We have made VAIroma a public service¹ for students and citizens to learn more about the whole or part of Roman history.
- We present the rationale behind the design of VAIroma and how we have leveraged and advanced the state-of-the-art interactive visualization and computational methods. We evaluated VAIroma with 16 students from an architecture class. The results demonstrate VAIroma enables the students to discover more articles and relationships relevant to their interests.

2 RELATED WORK

VAIroma intersects multiple research areas, including sensemaking, visualization of history, as well as visual analysis of unstructured texts.

2.1 Sensemaking

Sensemaking plays an important role in visual analytics to enable people to create actionable knowledge from observed data [11, 33]. Dervin presented sensemaking methodology for information science [11, 12]. Pirolli and Card introduced a notation model of the sensemaking loop for visual analytics [30]. Although influential, this work did not discuss any aspect of implementation as a computer system. Most recently, Sacha et al. [34] presented a human-computer sensemaking model for visual analytics systems that details the connection between human and computational aspects and adds the knowledge generation loop. The support of knowledge generation is an important goal for all visual analytics system. VAIroma is designed and developed with this very goal in mind, with features (such as annotations) specifically designed for recording newly created knowledge.

2.2 Visualization of Topical Trends

Several research papers have developed visualization systems for topic summarization from large text corpora [1, 29]. One popular approach is using a river-like graph such as a stacked graph or ThemeRiver [19] to provide a good overview of topical themes over time. Examples are TIARA [26, 39], Textflow [9], Cloudlines [25], and EventRiver [27]. Similarly, VAIroma employs a stacked graph to identify key historical events of the Roman empire over a 3000 year time period described by 40 topics with individual timeline graphs for each group or topic to allow users to explore topics in detail.

Recently, visualization systems have used social media data to explore temporal events with geographic information, such as Blogscope [2], Tweetgeist [35], and TwitInfo [28], based on user's location information. In contrast, VAIroma maps temporal trends with geolocation that are automatically extracted from unstructured text corpora.

2.3 Visualization of Wikipedia Meta Data

Wikipedia data is rich and heterogeneous; multiple previous papers have focused on analyzing and visualizing the meta data of Wikipedia. Holloway et al. [20] presented a visual analytics application based on a semantic map for Wikipedia meta data for the identification of semantic structure and categories of Wikipedia collection. Wattenberg et al. [38] described a visualization technique, the chromogram, to explore Wikipedia administrator activity. Brandes and Lerner [3] presented a visual interface to show authorship of controversial topics of Wikipedia, such as important historical events and persons. Chan et al. [6] presented Vispedia, a web-based interface that generates visualizations of data integrated from Wikipedia. Chevalier et al. [8] introduced WikipediaVis, a dashboard skin for Wikipedia, for users to view the history of a Wikipedia article in order to improve the quality of articles. Riche et al. [32] presented iChase, a heatmap-based interactive visualization tool, that allows management of editing activities on Wikipedia. Some previous work used DBpedia data to address the problem of visualizing heterogeneous data [5] and Resource Description Framework [15]. In contrast, our work focuses on extracting meaningful spatial, temporal, and topical information from the collections of Wikipedia articles for the purpose of supporting the generation of new knowledge. To the best of our knowledge, none of the previous Wikipedia related visualizations have analyzed large collection in this manner.

2.4 Visualization of History

Other visualizations have focused on supporting the study of history. Kimura et al.[22] proposed a visualization method to extract relationships between people and Japanese historical documents by using person names and place names. Itoh and Akaishi [21] presented an iterative visualization framework for historical figures that are extracted from historical documents. Their work, however, treats historical events in a short time period (1560 to 1570) and is limited with a few keywords. In contrast, our work covers over 3000 years with 40 related topics from 189,000 Wikipedia articles.

Krishnan et al. [23] introduced POLIS, an interactive web-based application to support scholarly research of the ancient Greek world, based on Oxford Classical Dictionary data. While their work relies on

¹<https://vairoma.uncc.edu>

manually generated information by 2,300 people, our approach provides an automated method to extract events and their indicators as well as interactive visualization.

3 DESIGN RATIONALE

The design of VAiRoma stems from an interdisciplinary collaboration between researchers in computer science and four professors from the college of arts and architecture. To solidify the construction and representation of the narrative on Roman history, the team went through an iterative process of identifying the major components that contribute to the narrative. Three of the four professors are architects; the other one is an art historian who studies the history of ancient coins.

Task abstraction and design requirements The VAiRoma project aims to support the task of exploring Roman history and constructing a narrative using digital collections. During this exploration, users would need a way to record major insights they have discovered. Therefore, the design requirement includes the support of different ways users would *enter and carry out* the exploration, as well as the recording of insights and new knowledge generated during the often iterative exploration. We detail the 3 ways users with different expertise would enter the exploration phase below based on our discussions with the interdisciplinary team. The approach we use here is modeled after the overview design study [4].

Why show a map? The architects *start* the exploration from navigating within a world map or search for a location of interest. Therefore incorporating a map in the visual narrative representation supports the thinking process of architects. Showing places of interest on a map also supports other users who may not be familiar with certain places so that they can easily locate targets on the map. Further, showing different geo hotspots over time allows users to make sense of the changes in space and time.

Why show a timeline? The art historian *starts* by navigating along a historical timeline and identifying important time periods based on his prior knowledge. To construct a narrative, a timeline is needed to provide information on when important historical events occurred. Therefore, we develop a timeline visualization based on the dates extracted from Wikipedia articles, as introduced in section 4.1.4.

Why two representations for topic summary? The computer scientists who represent users who are less familiar with Roman history *start* the exploration by interacting with a summary of the digital collection. To best present the topic summaries, we develop two visualizations that present different characteristics of the topic summaries. The topic hierarchy view provides an intuitive display of topic groups in a horizontal tree structure, while the circular topic view emphasizes on the importance of different topics or topic groups so that users can choose to start investigating the topics with more weights.

Supporting exploration, verification, and knowledge building via user interaction. The team designed user interactions that are tailored to the needs of our targeted users. In addition to the capabilities that are commonly provided by visual analytics systems, such as zooming and filtering, VAiRoma permits search by location or articles, as well as comparisons between different time ranges and spatial areas. To further support knowledge generation, gathering, and recording, we designed several annotation features and an annotation template within VAiRoma. The recent Knowledge Generation Model for Visual Analytics [34] highlights the importance of supporting the knowledge generation loop in visual analytics systems. In particular, through our iterative development and evaluation process, the team discovered that it is critical to support annotations during the analysis process so that valuable insights and new knowledge generated will not be lost. In addition, to further ease the creation of the annotations, we designed an annotation template since previous studies have found that a template encourages users to record their insights and newly created knowledge [7].

Dissemination. Lastly, to facilitate access to and dissemination of the VAiRoma interface, the team collectively decided to implement the visual component of VAiRoma in a web-based environment. The web-based interface connects to a web server to retrieve data on the fly based on users' interactions, and stores the annotations the users make onto the server. When one logs in later using the same user ID, the previously made annotations can be retrieved and updated within the VAiRoma interface. The server was set up and tuned so that it could support many users simultaneously.

4 INTRODUCING VAIROMA

VAiRoma is a visual analytics system that consists of several data analysis processes and an interactive visual interface. An overview of the system architecture is shown in Figure 2. All of the data analysis and visual components are developed to address the design objective of supporting the creation of a historical narrative.

4.1 Data Processing and Organization

To start, we first collected all English Wikipedia articles [40]. The entire collection contains more than 4.5 million English articles. We developed a data parser to convert the content of the articles to plain texts. The cleaned data then underwent four processes, namely Keyword Filtering, Topic Modeling, Location Extraction, and Date Extraction before being visualized.

4.1.1 Keyword Filtering

We worked with our collaborators to develop a set of Rome relevant keywords to filter the Wikipedia collection to the most relevant subset. The Rome-relevant keywords include “Rome”, “Roma”, and “Roman” plus other words. As a result, the geo scope of VAiRoma focuses on the city of Rome, but goes beyond to cover places that the Roman Empire and the Italian State encompass. The resulting subset, through only about 5% of the full Wikipedia English collection, still amounts to about 189,000 full-text documents. Next, we employed analytical methods to extract information to support the construction of the three major components identified in the design process, namely topics/themes, locations, and timeline.

4.1.2 Topic Modeling

To extract meaningful topics that pervade the text collection, the filtered Wikipedia data went through the Topic Modeling process (Figure 2B) [37]. Forty topics are extracted to summarize the 189,000 Rome-related articles. To facilitate visual exploration and reduce clutter of the topic space, we employed the Topic Rose Tree [14] to organize topics into a hierarchical structure in which similar topics are in the same branch.

To extract locations (Figure 2C left) mentioned in the Wikipedia articles, we leveraged the Stanford NER (Named Entity Recognizer) [17]. Challenges introduced by working with an established NER include no well-established standards for evaluation [24], and the varying performance of various NER tools on different data sources. To help address the accuracy of the extracted locations, we added a certainty score for each location to indicate our confidence in the location (0 to 6, as seen in Table 1), with each score denoting a certain type of locations. This rating is later displayed in the visual interface and we allow users to filter locations based on the score.

Table 1. Geolocation certainty

Certainty	Description
6	Continent or oceans
5	Peninsulas, countries, rivers or mountains
4	Famous cities including capitals
3	Ancient cities
2	Manually corrected geolocations
1	Geolocations that match a title in the Wikipedia
0	Other geolocations

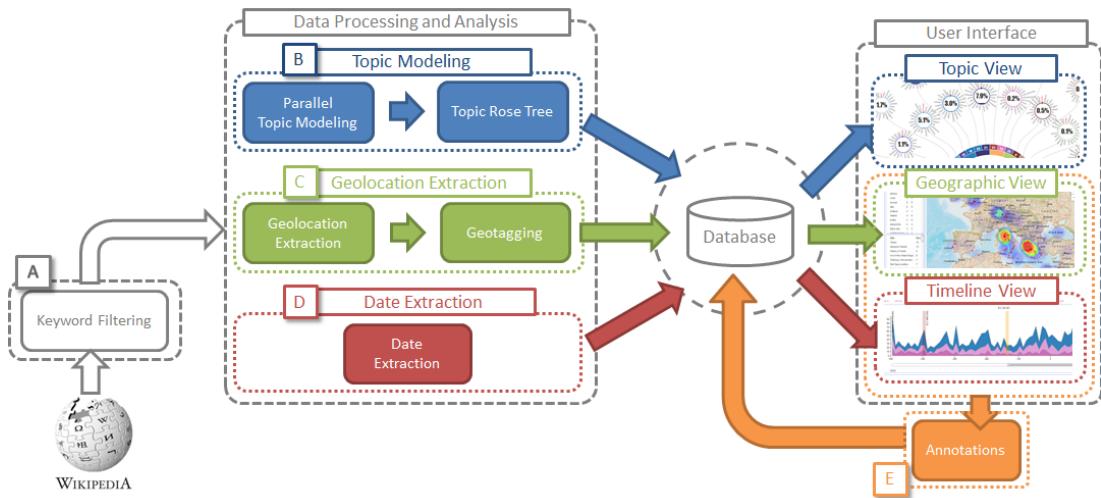


Fig. 2. System Architecture of VAIroma. Collected Wikipedia articles go to the keyword filtering process to filter articles about Roman history. Then filtered data goes to the Topic Modeling and Named Entity Extraction processes to extract topics, locations, and dates before being visualized. All results including user-generated annotations are stored in our database and available for use.

4.1.3 Location Extraction and Geotagging

After the location extraction, to pinpoint locations on the map in the visual interface, the extracted locations underwent geotagging (Figure 2C right) to identify geocoordinates. This process is done through GeoNames' REST web services [18] and Wikipedia web services. The GeoNames tool enables us to find ancient geolocations that no longer exist or changed to a different name in the modern era, although we cannot always find their coordinates from the modern database. Multiple candidates are available for geotagging one location, we choose the top ranked candidate for the location.

Although the geotagging process provides reasonable results including latitude and longitude for ancient names, some geolocations still need to be manually assigned due to ambiguities (e.g. Athens could be the capital of Greece or a city in Georgia). The authors went through several iterations to correct inaccurate geotags. Each author is assigned a continent and examined the locations over time. And we have also corrected wrongly tagged locations discovered by the participants during our user study (section 6)

4.1.4 Date Extraction

To extract temporal information (Figure 2D), we first used the Stanford NER tool. However, we quickly found that the NER tool performed far worse in extracting ancient dates/eras since it cannot detect dates that come with dating system abbreviations such as A.D. (Anno Domini) and B.C. (Before Christ). To address the problem, we used regular expressions to detect such dates. More specifically, we found abbreviations including A.D., B.C. and B.C.E. with and without periods. We then looked for the preceding or following numerical number as year information. If the preceding/following word is not a number, we checked whether the word refers to a century (e.g. 1st century B.C.) or a decade (e.g. A.D. 200s). To deal with event dates that are vaguely mentioned, such as *early* 16th century, we generate a year based on the description (early, mid, or late) to represent the time mentioned, since the year information is needed to construct the historic timeline. Coupling the NER extraction and our regular expression results, we achieved a fairly good coverage on both modern and ancient dates.

4.1.5 Correlating Locations and Dates

To support deep analysis and answering questions such as, for example, which locations in the world are the geo hotspot during 0 - 500 AD and how do the geo hotspots migrate over time, the correlation between a location and when it was mentioned needs to be established. This is not a trivial task, since numerous locations and times are often mentioned in a single Wikipedia article. For this reason, document-wide location and time correlation would result in false links since it

assumes all locations and dates in one article are correlated.

To make the correlations more precise and meaningful, we leverage the structure of the Wikipedia articles, with history related articles often following a chronological structure. Therefore, we establish the correlations between locations and time on a paragraph level. For instance, in the Wikipedia article titled "Roman-Persian Wars", a paragraph on an earlier war mentioned the attack against Tigranes in 69 BC, while another paragraph mentioned the line of stalemate shifted to northern Euphrates in 115 AD. Through correlating locations and dates on a paragraph level we established the links "Tigranes - 69 BC" and "Euphrates - 115 AD". This process is deployed to all articles in our Wikipedia collection, in order to find meaningful correlations between locations and dates.

4.2 VAIroma: The User Interface

VAIroma is a web-based application that is mainly developed using D3.js [10], ArcGIS [16], PHP and Javascript. As illustrated in Figure 1, VAIroma comprises 3 primary views: Timeline view, Geographic view and Topic view. Each view provides different features extracted from the Wikipedia collection and, more importantly, these views are interlinked to allow users to learn and create knowledge through analyzing a large amount of Wikipedia articles temporally, spatially and topically. In addition, VAIroma includes a Wikipedia Article viewer to allow users to read the formatted article.

4.2.1 Timeline View

The timeline view (Figure 1A) presents temporal topical trends of the Wikipedia collection over 4000 years (2000 BC to 2010 AD). The timeline is constructed based on topic and time, with each point representing the number of articles related to a certain topic that also mention the indicated time. Through user interaction, a user can choose to view the temporal trend of a topic, topic group, or all topics combined. When displaying multiple topics together, the timeline appears as a stacked graph with each topic as an individual ribbon.

To focus on a certain time range (e.g. the 500-year period for the Roman Empire), we added a slider at the bottom to permit zooming in to any time period of interest. Figure 1A illustrates a timeline between 500 BC and 500 AD selected through the slider. Each tick in the timeline represents a decade, so a user can select either a decade or multiple decades via mouse clicking or dragging. When a time (range) is selected, only the locations mentioned within the time range are shown in the geographic view.

As shown in previous work [13], peaks along the timeline are often indicators of one or more events that contribute to the peaks. Thus examining peaks for various topics or groups of topics is one way to

start the exploratory analysis, yielding major events described in the Wikipedia articles. As illustrated in our case study (section 5), the users were able to identify major events for the Roman Empire by perusing peaks along the timelines.

4.2.2 Geographic View

In the geographic view (Figure 1B) we chose ArcGIS [16] in order to provide several imagery maps at multiple scales, roads and topographic images. In addition, we added three feature layers: heatmap layer, points layer, and pin layer. The heatmap layer is designed to represent density of geolocations on the imagery map by color coding, with red areas indicating more frequently mentioned regions. The heatmap permits users to discover the geographic “hot spots” based on the selected topic and time. The points layer shows all geolocations as red circles on the geographic view. While the heatmap shows a good overview of geolocations of the selected time range, the points layer shows exact locations. Users can click on a red circle to see a list of Wikipedia articles (Figure 1h) that contain the geolocations and dates selected. The pin layer shows pins indicating selected locations. This layer is especially useful when several points are overlapped on the map due to the area having different names (ancient vs. modern). The user can click on a pin to invoke a tooltip showing the titles of Wikipedia articles mentioning the location (Figure 1h tooltip). In addition, this layer is used for indicating a geolocation that is found through the search function (Figure 1g).

4.2.3 Topic View

Two representations for displaying topic hierarchies, topic content, as well as topic weights (that denote the importance of various topics) were developed. Two main reasons led us to this decision based on an iterative design and evaluation process:

- 1) Since VAIroma is designed under the assumption that the users may not be familiar with either visualization or topic models, we want the topic visualizations to stay intuitive without overwhelming amount of information all displayed in one view.
- 2) At the same time, for users who have experience with topic models, we also want to present important information (output from the models) efficiently within limited screen space to guide the discovery of important topics.

In this section, we present the two topic representations and how they complement each other in presenting topic-related information.

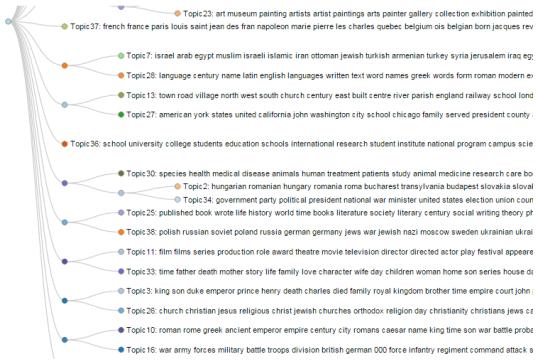


Fig. 3. The topic tree view: topics are displayed in a horizontal tree structure. Panning, zooming, and selection are supported.

Topic Tree View The topic tree view provides a hierarchical list of topics (Figure 3) that are extracted to summarize the main themes in the collection [14]. In this view, each topic has a unique number and color. Colors are mainly used to correlate a topic and its trend in the timeline. Colors are randomly assigned to each topic and the corresponding topic ribbons in the timeline view will be of the same color. We have a color palette of 20 unique colors, 1 out of the 20

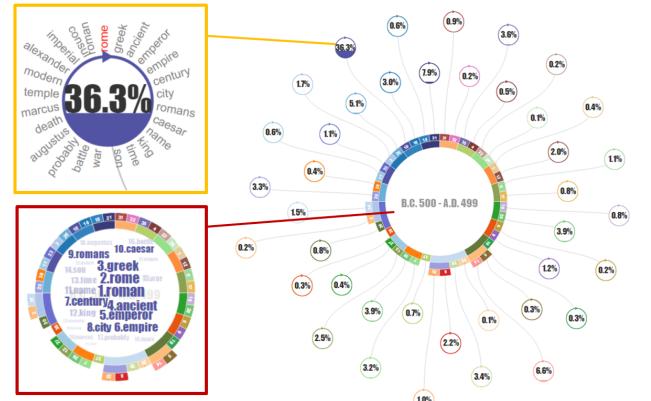


Fig. 4. The circular topic view: the sunburst in the center represents the hierarchical structure of the topics. The texts in the center shows either the currently selected time range or the topic keywords (shown in the red rectangle). Each outer circle represents a topic. Zooming permits topic keywords which are arrayed around the topic circle (orange rectangle) to be more easily read.

colors is randomly assigned to each topic and its temporal trend. We also made sure that two adjacent tree nodes are not assigned the same color so that their temporal trends can be easily distinguishable. The user can select a topic to filter the collection down to articles that are relevant to the topic (for example, the user might want to look into articles that are related to the topic on religion or another topic on wars and battles). When the user clicks on a topic, the timeline view changes to the timeline of the topic and the results in the Geographic view are updated to show locations that are mentioned in the articles that are related to the topic. The topic tree view is intuitive and users can start navigating through the topic space with no or little training.

Circular Topic View The circular view is developed after our initial deployment of the VAIroma interface with the topic tree view alone. Users with experiences in text summarization methods often want to start with identifying the important topics; such feedback led us to design a different topic representation. Two design requirements were gathered from the feedback: 1) the view needs to show the topic hierarchy clearly, 2) the view needs to present modeling related information, such as topic weights, how they change over time, as well as keywords of topics. A compact overview is desired to facilitate the comparison between topic weights without too much navigation.

Through several design iterations, we decided to adopt a circular layout for presenting the topics as opposed to a flat tree, since the circular layout uses space more efficiently. Our design is inspired by the sunburst graph [36], which is a compact design for presenting hierarchical information. In our design, the sunburst graph in the center circle represents the hierarchical structure of topics: the inner ring shows topics or topic groups that are directly connected to the root node, while the outer ring layers show child branches and leaf nodes. We then connect each leaf node in the circular topic view with its topic circle. The organization of the topic circles is cognizant of space, alternating their distance to the center ring. With each circle filled with its topic weight, users can compare all topic weights in a glance.

When zooming into a certain level, a topic circle shows the top 25 keywords (orange rectangle in Figure 4). The main keyword is highlighted in red and the other keywords are listed in clockwise order from the main keyword. Each circle is filled according to its weight within the selected time range, a percentage of the precise proportion is also shown in the circle. In addition, when hovering mouse over a specific topic circle or a sunburst node, a word cloud of the selected topic’s keywords is displayed inside of the center circle (red box in Figure 4). The numbers in the word cloud indicate importance ranking of keywords. When a user clicks on a topic circle, the timeline view displays the trend of the selected topic.

Overall, we discovered that the circular view’s compactness makes it superior for use on systems with smaller screen space, such as laptop.

topic	title	par	locations	date	count	weight	wiki
topic16	Italian and Swiss expedition (1799–1800)	6	Italy	-218	1	0.6492151897...	open
topic16	Italian and Swiss expedition (1799–1800)	6	manua	-218	1	0.6492151897...	open
topic16	Italian and Swiss expedition (1799–1800)	6	trebbia river	-218	1	0.6492151897...	open
topic16	Italian and Swiss expedition (1799–1800)	6	riviera	-218	1	0.6492151897...	open
topic16	Italian and Swiss expedition (1799–1800)	6	naples	-218	1	0.6492151897...	open
topic16	Battle of Baecula	5	guadalupe	-208	1	0.6104024032...	open
topic16	Watermill	7	ptolemaic egypt	-220	1	0.6156956157...	open
topic16	Watermill	7	perachora	-220	1	0.6156956157...	open
topic16	Watermill	7	greece	-220	1	0.6156956157...	open
topic16	Battle of Scarrifholls	11	welsttown	-216	1	0.55020752697...	open
topic16	Battle of Scarrifholls	11	carr	-216	1	0.55020752697...	open
topic16	Battle of Lake Trasimene	4	trebia	-217	2	0.54203915177...	open
topic16	Battle of Lake Trasimene	4	rome	-217	3	0.54203915177...	open

Show whole article Highlight keywords Highlight topic keywords Roma

Fig. 5. The tabular view shows a list of Wikipedia articles of a selected time range. In this figure, geolocation (Welsttown) is highlighted in yellow, date (216 BC) is highlighted in pink, filtered keyword (Roma) is highlighted in red, and keywords of topic 16 (army, military etc) are highlighted in blue.

tops. On the other hand, the topic tree view has the advantage of being more intuitive since users can dive right into the content of the topics.

4.2.4 Wikipedia Article View

This view simply shows a Wikipedia article (Figure 6C). Users can get to the article by either clicking on a location on the map then choosing an article mentioning the location, or directly clicking on an entry in the article table(Figure 1f). The Wikipedia article view resides in the same screen space as the Topic Tree View. Users can switch between views by clicking on the tab above the view (Figure 6D).

4.2.5 Tabular View

To help users analyze a selected set of Wikipedia articles in detail, the tabular view lists the articles and permits examination of keywords, locations, and times at a *paragraph* level. As seen in Figure 5A, the table shows information regarding the related topic, Wikipedia article titles, paragraph number, location, date, and count of the location appearing in this paragraph, etc. The list is sorted by topic weight in descending order as a default but a user can sort the list by any fields. When the user selects a single row, the corresponding paragraph is shown in the text field (Fig 5B). In the text field, geolocations and dates are highlighted in yellow and pink respectively by default. In addition, filtering keywords and topic keywords can be highlighted by toggling the blue buttons on the bottom. The view also shows full text of the selected Wikipedia article when the user checks the “show whole article” check box on the bottom left. We have found that the tabular view is quite useful because it supports quickly pinpointing where and how a place or time of interest is discussed.

Note that the tabular, topic, and Wikipedia views all occupy the same screen space at the top right corner of the interface, users can easily switch between different views by clicking on the tabs at the top of the view (Figure 1i).

4.3 VAIROMA: Interaction

VAiRoma provide rich user interactions to support fast exploration of Wikipedia articles and the discovery of previously unseen connections between space, time, and topics. In addition, VAiRoma supports user annotation so that the created knowledge can be stored and updated.

4.3.1 Selecting a time range

An important interaction supported in the VAiRoma interface is the selection of a specific time period on the timeline view in order to see the locations mentioned during the period. Once a time period is selected, a list of all related locations can also be found in the location table next to the Geographic view (Figure 1e). Simultaneously, a list of Wikipedia articles mentioning the selected time range is shown in the article table (Figure 1f). This table is useful for rapidly scanning titles of the Wikipedia articles to make sense of what happened during the selected time range. If the user wants to read a certain article, she can click the entry in the table to show it in the Wikipedia Article View.

Authorized licensed use limited to: CAMBRIDGE UNIV. Downloaded on October 20, 2024 at 16:07:26 UTC from IEEE Xplore. Restrictions apply.

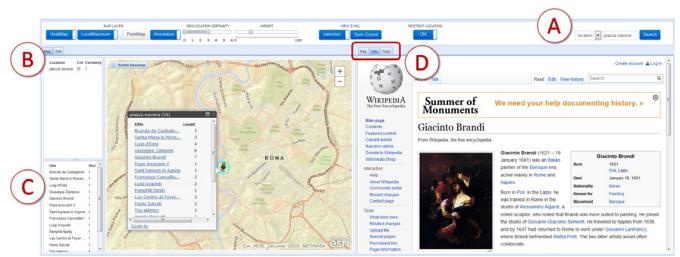


Fig. 6. Search results of “Piazza Navona” (A). Searched geolocations are shown in the location table (B) and related articles are shown in the article table (C). The green pin represents Piazza Navona in Roma.

4.3.2 Filtering

VAiRoma supports multiple filtering functions, including filtering by topics, time, and geo regions. By filtering with a combination of these, users can arrive at a much more focused set of articles to read and discover previously unseen connections in space, time, and events.

4.3.3 Search by Location or Article

Often a user may identify a place or event of interest during the VAiRoma analysis. To learn more, the user can search for locations or articles through the search box on the top right (Figure 1g). The user can choose either ‘location’ or ‘article’ by the dropdown menu and then type in a string. The search result is displayed in both location and article tables (Figure 1e and f), as well as in the Geographic view. More information on either a location or an article can be obtained through clicking on an entry in either table. An example is shown in Figure 6, a user searched for “Piazza Navona” using the “location” search (Figure 6A), with the time span set to 0 - 2010. As a result, “Piazza Navona” is shown in the location table (Figure 6B), while a list of 13 Wikipedia articles mentioning “Piazza Navona” are shown in the article table (Figure 6C). The list of articles includes churches, buildings, and personages related to Piazza Navona at certain time in history. Having these connections to Piazza Navona organized and displayed in one place through a simple search enables users to quickly make sense of the connections and derive new knowledge. This is a real use case discovered during the user study, and all participants agreed that the VAiRoma better supports discovery of relationships compared to Web searches.

4.3.4 Search All Locations in a Specific Area

Another function frequently requested by architects and historians is to search for articles mentioning locations and events occurring in close proximity to the location of interest. Such events, ancient or modern, contribute to the understanding of a particular place. VAiRoma supports such function by allowing users to view Wikipedia articles mentioning nearby locations when zooming into a specific region (areas less than 3 mile radius). Searching for all mentioned locations in an area can be achieved by zooming into a region and right clicking on the Geographic View to get “All locations in this area”. The resulting Wikipedia articles and locations are shown in the tables, so that users can peruse the results to discover new and sometimes unexpected relationships between different articles, events, and locations.

4.3.5 Comparison between Two Different Time Ranges

To support another important knowledge-fostering function, the VAiRoma interface permits the comparison of geographic patterns side by side to allow the user to answer question such as “How do geo centers in Europe migrate over time?”. Figure 7 shows a comparison between two major peaks (220 - 201 BC and 60 - 41 BC) of a topic centered on battles and wars. The Geographic view on the right (taking the place of the Topic view) is brought out by clicking on the “map” tab. The two geographic views show heatmap results for the 2 selected time ranges. The left view highlights several locations in Italy (Bruttium, Apulia, Capua, etc.) and Spain (Ebro), while the right view highlights Greece (Alesia) and Tunisia (Utica) and France (Georgovia). Furthermore, the article table of the left view shows many

October 20, 2024 at 16:07:26 UTC from IEEE Xplore. Restrictions apply.

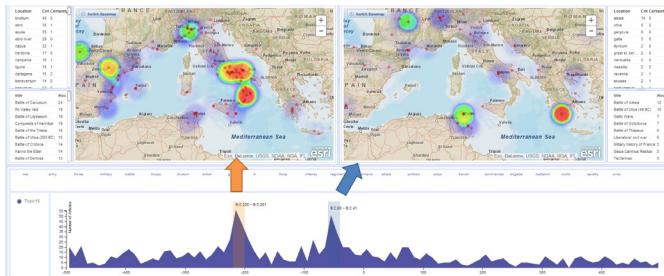


Fig. 7. Comparison between 220 - 201 BC (left) and 60 - 41 BC (right) in Topic 16 (war). Top 3 Wikipedia articles of the left view are “Battle of Causium”, “Po Valley raid” and “Battle of Lilybeum”. Top 3 articles of the right view are “Battle of Alesia”, “Battle of Utica (49 BC)” and “Gallic Wars”.

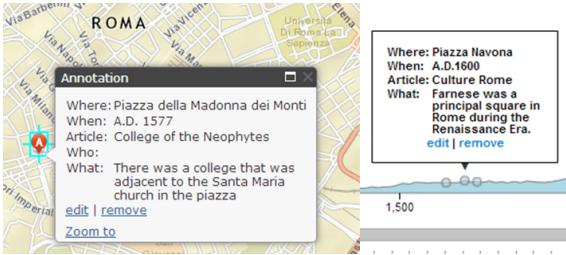


Fig. 8. An example of user's annotations on the map view (left) and the timeline view (right) from our user study. Users made annotations about “Piazza della Madonna dei Monti” in 1577 AD and “Piazza Navona” in 1600 AD. Both findings were not discover via Web searches.

Wikipedia articles about battles occurring during 220 - 201 BC (Hannibal and the Second Punic War), while the article table for the right view shows battles occurring during the later time. Hence, the user can quickly find and compare the changing foci of conquests or civil wars in history, including but not limited to where the battle sites were, when the battles occurred, and the scale and effect of different battles.

4.3.6 User Annotation

To help users retain their findings and newly created knowledge, VAIroma supports user annotation (examples shown in Figure 8). Users can make an annotation directly on the map or timeline view when they find something noteworthy. To further simplify the annotation process, VAIroma provides an annotation template to help users record the 4Ws (where, when, who, and what) and through interpretation to make headway towards the 5thW (why). Generated annotations can be updated or removed later. In the Geographic view, the annotation is denoted by a symbolic icon (red ‘A’ in Figure 8 left) at a specified location. Figure 8 left shows an annotation made by a participant in our user study, on findings regarding “Piazza della Madonna dei Monti”, a small piazza in Rome. Via VAIroma, the user found that there was a college in the piazza near the Santa Maria church from the article “College of the Neophytes”. It is worth noting that the user did not discover the relationship via Web searches.

In the timeline view, an annotation is represented by a circle on top of the graph (Figure 8 right). Users can view or hide the annotation by clicking on the circle. All annotations are stored in the database indexed by usernames. A user can access and edit previously made annotations whenever she logs on with the same username.

In summary, through carefully designed visual representations and user interactions, the VAIroma interface supports exploration, analysis, and more importantly knowledge generation based on the Rome-related Wikipedia article collection.

5 CASE STUDIES

In this section, we present two case studies conducted with real users. The case studies illustrate how VAIroma can help a user cohort, in this case educated citizens who are interested in but not experts on Roman

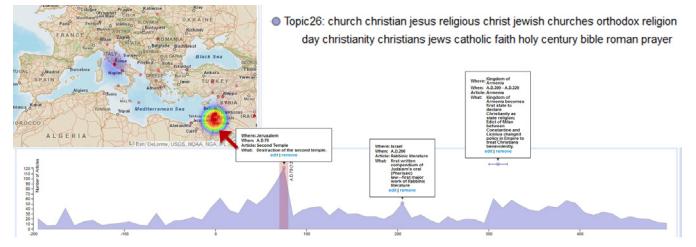


Fig. 10. Annotations on three main peaks of Topic 26 (religion). The main peak indicates the destruction of the Second Temple in Jerusalem (70 AD). The map view highlights Israel.

history, to gain knowledge on the subject through attaching meanings to aspects of the geotemporal topical analysis and interpreting relations among the events and topics that contribute to them. VAIroma enables users to identify and make sense of major events through perusing a small selection of highly relevant Wikipedia articles. This user cohort is represented by two professors of Computer Science and Architecture, neither of whom is an expert on Roman history.

5.1 Case study 1: Identifying and making sense of event peaks for the Roman Empire

The users started by concentrating on the eras of the Roman Empire between 27 BC - 476 AD by moving the sliders to the relevant range within the timeline view. They then mainly concentrated on the topic and timeline view to identify and annotate event peaks during the 500-year period. Interestingly, the users found that there were typically 2-4 main topics per event peak (Figure 9 bottom, next page) and that they could annotate the event peak by perusing around 10-20 documents from the top group listed in the article table. Thus the combination of temporal, topical, and document interactive analyses permits one to boil down 189,000 documents to around 200 key documents for this time range. This process took a period of hours, as the users selected peaks, then main topics, and perused the documents in the article table as well as observing the changes in geospatial patterns. Typically the users viewed articles for the 10 year period around the peak, or sometimes on either side depending on the width and position of the peak. The titles in the article table were useful in setting initial aspects of the topic for that event peak. Then the users could quickly read the key documents for more detail on the contributions from that topic. Obviously, this effort takes very much less time than any effort that starts with the original unorganized mass of relevant Wikipedia documents.

The users made annotations to record their findings throughout the analysis. Examples of the collected annotations are shown in Figure 9. Through interacting with VAIroma, the users were able to identify event peaks that can be attributed to major events in history: the first Caesars (41 - 54 AD), Roman conquest of Britain (43 AD), the Flavian Dynasty (60 AD), mounting troubles at the Empire’s borders and the tetrarchy established by the Diocletian that ended with his death (311 AD), the embracing of Christianity (300 - 319 AD), the rise to power of Constantine and the establishment of the Eastern Empire (320 - 335 AD), etc. When one compares the themes revealed in the annotations with an overview of Roman history during this period (compiled by the historians in our group) that summarizes standard sources, one finds that the annotations fit this overview quite well. Thus a small group of non-experts has established a reasonable narrative for Rome over 5 centuries and has also collected key document in the Wikipedia collection. They have done this in a small fraction of the time that would be necessary if one had to attack the whole collection.

5.2 Case study 2: Unveiling and Interpreting deeper relations through the use of VAIroma

A rich context is supplied that supports deeper analysis, sometimes via discovering unexpected relations. Figures 10 and 11 show two examples. Many of the topics are quite discriminating. For example, Topic 26, exemplified by the purple stream in Figure 10, deals with religion.

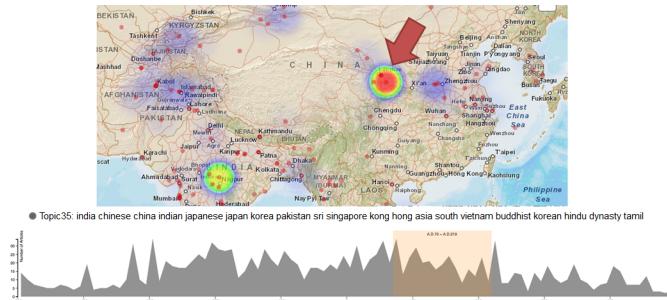


Fig. 11. Selecting between 70 - 219 AD on Topic 36 leads to geo hotspots in China during the Han Dynasty (red arrow).

Looking at this topic over the 5 century time period reveals the story of the Judeo-Christian religion during this time, and its connection to the Empire. Key events such as the destruction of the Second Temple (after which Jews were dispersed from Jerusalem and the center of the Christian church moved towards Rome) and the development of both Christian and Rabbinical doctrines are laid out in the annotations, which were derived from key documents. Figure 11 shows findings on Topic 35, which focuses on the Far East. The topic reveals a rich set of references to the Han Dynasty in China, which had many parallels to the Roman Empire during an overlapping time period, and the emergence of trade routes via India (which included trade between the empires). In this case, we found the time range depicted in Figure 11 by moving a time window on the timeline until a strong set of geographic hotspots showed up in China and India. Such interconnections automatically emerge and help one avoid an overly narrow (in this case Eurocentric) view of events. The organization and discrimination provided by individual topics also reveal more subtle references to plagues, droughts, other natural disasters, and social upheavals that can be drawn out and related to political, military, and social events. These examples also reveal a key aspect of the approach we developed. It is initially exploratory so that one does not need to do elaborate initial filtering of the document collection (or figure out what to filter, since it may not be at all clear). As a result, threads and interrelations such as these, which can be unexpected, will be kept and, just as importantly, can be explored more deeply on their own by users.

6 USER STUDY

In this section, we present a user study with Architecture students who were in Rome during the summer to study and learn about the roman piazzas' architecture as well as their history. When in Rome, the participants used web searches and books from a local library to learn about the piazzas. The goal of the study is to find out whether

VaiRoma can enable the participants to discover *new* findings than what they have gathered from the books and Web searches.

6.1 Participants and Experiment Design

Sixteen participants took the user study. 9 were male and 7 were female (7 master students and 9 undergraduate students). The age of the participants ranged from 19 to 35. Given the familiarity of the participants with the roman piazzas, we consider them as experts of the subject matter. However, none of them is an expert in either contextual Roman history or the visual interface. The user study was conducted in a computer lab in the Architecture department. The specifications of the specifications of the computers used for the user study is iMac21s (1920 × 1080 resolution), with i5-347s @2.90GHz, 8GB memory, GeForce GT 650M and Windows 7 64bit installed.

We designed the tasks together with the professor of the architecture class to best incorporate VAIroma into the learning objectives of the class. Since each of the 16 participants was assigned to learn about 1 out of 8 Roman piazzas during their 2-month stay in Rome, we designed the user study task to follow their workflow instead of having all participants investigating one piazza of our choice. Since the participants have accumulated lots of knowledge on their assigned piazzas by doing web searches in Rome, they can now make informed comparisons on discoveries using VAIroma to what they already knew.

Through our pre-questionnaire, we found that all had high daily Google map usage (5.1 out of 7 on a 7 point scale) and Wikipedia site usage (4.1 out of 7), though only one participant was familiar with text summarization methods. All participants were provided 15 minutes of training on the VAIroma interface. The participants were asked to find relevant Wikipedia articles related to their assigned piazza using VAIroma and were encouraged to make annotations about their findings. All participants successfully finished the user study within 60 minutes including a post-study questionnaire regarding subjective evaluations and feedback on the VAIroma interface.

Since we wanted the experiment to fit into the curriculum of the class and due to the constraint that the experimenters could not be in Rome when the students were asked to gather knowledge on the piazzas, the user study did not include a condition on performing the same task using web searches. Although a direct comparison between the VAIroma interface and web searches was unavailable, we explicitly asked the participants to highlight and annotate additional findings that they had not found through Web searches during the study.

6.2 Results

Overall, participants had a very positive experience with VAIroma. Figure 12 shows the participants' subjective ratings of the VAIroma interface. We employed a 7-point Likert scale on all questions. We organized the questions into three categories. The first relates to usability

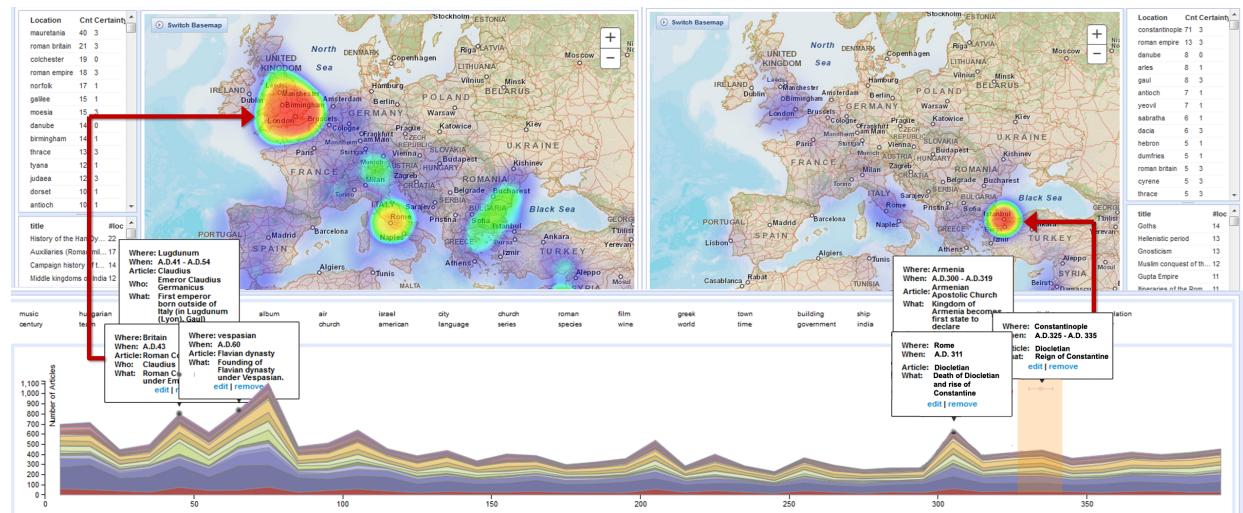


Fig. 9. Annotations of 6 major events in Roman Empire. The map on the left corresponds to the "Roman conquest of Britain", the map on the right describes "The Armenian War and the establishment of the Eastern Empire" at 16:07:26 UTC from IEEE Xplore. Restrictions apply.

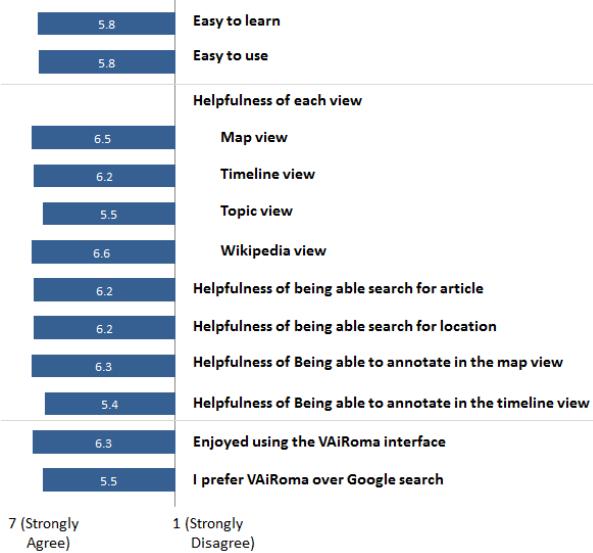


Fig. 12. Average users ratings on questions regarding VAiRoma.

of VAiRoma (ease of use, ease to learn). Most of the participants felt VAiRoma is easy to learn and use since they are familiar with most of the visual representations, but more importantly, because all of their desired functions are supported by VAiRoma. The second group of questions relates to whether the visualizations and user interactions (including annotation functions) are helpful for the exploratory task performed by the participants. All questions received favorable average ratings of 5.4 or above. The ratings on individual views and interactions provide insights on how the participants perceive and use different functions provided in VAiRoma. The comments from the users can also be used to further improve the current interface.

The last group of questions have to do with the overall exploration experience with VAiRoma and comparison to web searches. When asked how many articles can be found with VAiRoma that cannot be found via the web searches, the participants answered that they found on average 4 additional Wikipedia articles. Regarding the relevance ranking of the articles found via VAiRoma (1: *not relevant at all* through 7: *very relevant*), the average was 4.7. 10 out of 16 participants reported that they could not find some articles that they found via web searches, but these were mostly non-Wikipedia articles. Most participants reported that the interlinked visualizations showing relationship between space and time are advantages of VAiRoma. Some comments are “it helps you find connections to people and events that google does not always find” and “(VAiRoma provides) faster, easier and much clearer connections between different articles”. Some participants pointed out false positive results presented in the interface, namely few irrelevant articles. The possible improvement of such results are discussed in section 7. Overall, people favored VAiRoma. Some participants commented on features that enable them to find more relevant information through VAiRoma, including “the way information is organized in VAiRoma”, “the interface being interactive”, and “the combination of spatial, temporal, and topical patterns”.

6.3 Annotation

The 16 participants generated 44 annotations. All participants made annotations of new knowledge that they could not find via web searches (an example is shown in Figure 8). Note that some of the piazzas are small piazzas with no Wikipedia articles dedicated to them. However, the participants assigned to these piazzas were still able to identify relevant articles and events related to or occurring in the piazzas through the use of VAiRoma, largely because of the locations are extracted from the articles, not just titles. One participant was skeptical about what he could discover before the user study since he could not find much information via web searches on a small piazza. He was pleasantly surprised to find connections to old churches and sculptures

using VAiRoma. All answered on the post-questionnaire that they could find new Wikipedia articles mentioning certain aspects of the piazzas via the VAiRoma interface compared to their Web searches.

In summary, all participants agreed that the VAiRoma interface was easy to learn and use. They could find more relevant connections and relationships between the piazzas and Wikipedia articles compared to web searches or other references. In addition, VAiRoma enables them to discover historical events that they did not know before.

7 DISCUSSION AND CONCLUSION

In this paper, we present VAiRoma, a visual analytics approach to construct a data-driven view of Roman history based on a large collection of Wikipedia articles. In this section, we discuss the potential limitations of our approach and also its potential for expanded use.

One limitation might rise from the use of the English Wikipedia collection as our textual source. Although providing great coverage on all subjects, the descriptions provided by Wikipedia on a certain subject may not be very deep. We believe the current VAiRoma is sufficient for supporting general learning about Roman history by the general public. However, to support deeper scholarship, we plan to incorporate scholarly collections under the Open-Access License such as PLOS one [31]. Incorporating scholarly articles will enable experts on the subject to perform in-depth analysis and comparison of the textual sources, yielding deeper insights that can be shared with the general public. We also want to make the analysis of scholarly articles embodied in VAiRoma available to educated but non-expert citizens. We foresee that the wealth of contextual information plus the connection to more accessible collections such as Wikipedia will make the scholarly work more understandable to this cohort.

Another limitation may stem from the location and time extraction. As previous researches have pointed out, it is very difficult to evaluate the accuracy of our extraction results with no ground truth corpus. However, with the user interface being a public web-service, we can leverage the intelligence of the crowd to obtain more accurate geo and time extraction results. We are also working with a geographer to build an improved set of location and time terms.

In summary, we present a visual analytics approach to help people gain knowledge through making connections on places, times, and events. This approach shows the power of establishing a system based on the revised Sensemaking principles [34] and goes farther than previous implementations. VAiRoma focuses on constructing a narrative of the Roman history from ancient times, through the Empire, to modern times. We expect VAiRoma can contribute to knowledge generation by citizens and plan to keep improving VAiRoma based on feedback we gather. In addition, we will construct other long-timeline stories using the same approach.

The underlying techniques and approaches in VAiRoma are general, and they can be applied to other types of collections. To show this, we have started to work on a large collection containing the myriad aspects of climate change. With such a diverse subject as this, no one is an expert in all its manifestations, and the interrelations between topics, often hidden, are quite important. Yet this is an area of vital importance to both scientists and citizens. What we have learned from VAiRoma in terms of successfully informing non-experts is very useful here. This includes the fact that non-experts can correctly identify salient events and topics in long, complex narrative, that they can uncover unexpected (but real) relationships (for example, the sub-history of the Christian church as it relates to Rome or the trade relations between the Roman Empire and far-off civilizations), and that they can efficiently build relevant knowledge (filtering down to a small number of relevant documents for a given topic, time, and place from thousands of documents). These general capabilities are applicable to climate change and other complex narratives.

REFERENCES

- [1] E. Alexander, J. Kohlmann, R. Valenza, M. Witmore, and M. Gleicher. Serendip: Topic model-driven visual exploration of text corpora. In *Visual Analytics Science and Technology (VAST), 2014 IEEE Conference on*, pages 173–182, Oct 2014.
- [2] N. Bansal and N. Koudas. Blogscope: a system for online analysis of high volume text streams. In *Proceedings of the 33rd international conference on Very large data bases*, pages 1410–1413. VLDB Endowment, 2007.
- [3] U. Brandes and J. Lerner. Visual analysis of controversy in user-generated encyclopedias. *Information Visualization*, 7(1):34–48, 2008.
- [4] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *Visualization and Computer Graphics, IEEE Transactions on*, 20(12):2271–2280, Dec 2014.
- [5] M. Cammarano, X. Dong, B. Chan, J. Klingner, J. Talbot, A. Halevy, and P. Hanrahan. Visualization of heterogeneous data. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1200–1207, Nov 2007.
- [6] B. Chan, L. Wu, J. Talbot, M. Cammarano, and P. Hanrahan. Vispedia: Interactive visual exploration of wikipedia data via search-based integration. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1213–1220, Nov 2008.
- [7] Y. Chen, S. Barlowe, and J. Yang. Click2annotate: Automated insight externalization with rich semantics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 155–162. IEEE, 2010.
- [8] F. Chevalier, S. Huot, and J. Fekete. Wikipediaviz: Conveying article quality for casual wikipedia readers. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE*, pages 49–56, March 2010.
- [9] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2412–2421, 2011.
- [10] D3. D3.js. <http://www.d3js.org>. Accessed: 2015-03-30.
- [11] B. Dervin. An overview of sense-making research: Concepts, methods, and results to date. *International Communications Association Annual Meeting*, 1983.
- [12] B. Dervin. From the mind's eye of the user: the sense-making qualitative-quantitative methodology. *Qualitative research in information management*, 9:61–84, 1992.
- [13] W. Dou, X. Wang, D. Skau, W. Ribarsky, and M. Zhou. Headline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on*, pages 93–102, Oct 2012.
- [14] W. Dou, L. Yu, X. Wang, Z. Ma, and W. Ribarsky. Hierarchicaltopics: Visually exploring large text collections using topic hierarchies. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2002–2011, Dec 2013.
- [15] B. Dumas, T. Broché, L. Hoste, and B. Signer. Vidax: An interactive semantic data visualisation and exploration tool. In *Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12*, pages 757–760, New York, NY, USA, 2012. ACM.
- [16] esri. Arcgis. <http://www.arcgis.com>. Accessed: 2015-03-30.
- [17] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, ACL '05*, pages 363–370, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
- [18] GeoNames. Geonames. <http://www.geonames.org>. Accessed: 2015-03-30.
- [19] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *Visualization and Computer Graphics, IEEE Transactions on*, 8(1):9–20, 2002.
- [20] T. Holloway, M. Bozicevic, and K. Börner. Analyzing and visualizing the semantic coverage of wikipedia and its authors. *Complexity*, 12(3):30–40, 2007.
- [21] M. Itoh and M. Akaishi. Visualization for changes in relationships between historical figures in chronicles. In *Information Visualisation (IV), 2012 16th International Conference on*, pages 283–290, July 2012.
- [22] F. Kimura, T. Osaki, T. Tezuka, and A. Maeda. Visualization of relationships among historical persons from japanese historical documents. *Literary and Linguistic Computing*, 28(2):271–278, 2013.
- [23] M. Krishnan, J. Ober, and M. Pyzyk. Polis: Designing a visualization tool for the research of complex sociopolitical landscapes. 2014.
- [24] R. Krovetz, P. Deane, and N. Madnani. The web is not a person, berners-lee is not an organization, and african-americans are not locations: An analysis of the performance of named-entity recognition. In *Proceedings of the Workshop on Multiword Expressions: From Parsing and Generation to the Real World, MWE '11*, pages 57–64, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [25] M. Krstajic, E. Bertini, and D. A. Keim. Cloudlines: Compact display of event episodes in multiple time-series. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2432–2439, 2011.
- [26] S. Liu, M. X. Zhou, S. Pan, W. Qian, W. Cai, and X. Lian. Interactive, topic-based visual text summarization and analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pages 543–552. ACM, 2009.
- [27] D. Luo, J. Yang, M. Krstajic, W. Ribarsky, and D. Keim. Eventriver: Visually exploring text collections with temporal references. *Visualization and Computer Graphics, IEEE Transactions on*, 18(1):93–105, 2012.
- [28] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller. Twitinfo: aggregating and visualizing microblogs for event exploration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 227–236. ACM, 2011.
- [29] D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen. Comparative exploration of document collections: a visual analytics approach. *Computer Graphics Forum*, 33(3):201–210, 2014.
- [30] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, pages 2–4. Mitre McLean, VA, 2005.
- [31] Plosone. Plosone. <http://www.plosone.org>. Accessed: 2015-03-30.
- [32] N. H. Riche, B. Lee, and F. Chevalier. Ichase: Supporting exploration and awareness of editing activities on wikipedia. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 59–66. ACM, 2010.
- [33] D. M. Russell, M. J. Stefk, P. Pirolli, and S. K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, CHI '93*, pages 269–276, New York, NY, USA, 1993. ACM.
- [34] D. Sacha, A. Stoffel, F. Stoffel, B. Kwon, G. Ellis, and D. Keim. Knowledge generation model for visual analytics. *IEEE Transactions on Visualization & Computer Graphics*, 2014.
- [35] D. Shamma, L. Kennedy, and E. Churchill. Tweetgeist: Can the twitter timeline reveal the structure of broadcast events. *CSCW Horizons*, 2010.
- [36] J. Stasko and E. Zhang. Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on*, pages 57–65. IEEE, 2000.
- [37] X. Wang, W. Dou, Z. Ma, J. Villalobos, Y. Chen, T. Kraft, and W. Ribarsky. I-si: Scalable architecture for analyzing latent topical-level information from social media data. In *Computer Graphics Forum*, volume 31, pages 1275–1284. Wiley Online Library, 2012.
- [38] M. Wattenberg, F. B. Viégas, and K. Hollenbach. Visualizing activity on wikipedia with chromograms. In *Human-Computer Interaction-INTERACT 2007*, pages 272–287. Springer, 2007.
- [39] F. Wei, S. Liu, Y. Song, S. Pan, M. X. Zhou, W. Qian, L. Shi, L. Tan, and Q. Zhang. Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 153–162. ACM, 2010.
- [40] Wikipedia. Wikipedia. <http://en.wikipedia.org>. Accessed: 2015-03-30.