FEATURE ARTICLE

News Globe: Visualization of Geolocalized News Articles

Nicholas Ingulfsen, *ETH Zurich, 8092, Zurich, Switzerland*

Simone Schaub-Meyer, *Technical University Darmstadt, 64289, Darmstadt, Germany*

Markus Gross, *ETH Zurich, 8092, Zurich, Switzerland*

Tobias Günther, *Friedrich-Alexander-University Erlangen-Nürnberg, 91058, Darmstadt, Germany*

The number of online news articles available nowadays is rapidly increasing. When exploring articles on online news portals, navigation is mostly limited to the most recent ones. The spatial context and the history of topics are not immediately accessible. To support readers in the exploration or research of articles in large datasets, we developed an interactive 3D globe visualization. We worked with datasets from multiple online news portals containing up to 45,000 articles. Using agglomerative hierarchical clustering, we represent the referenced locations of news articles on a globe with different levels of detail. We employ two interaction schemes for navigating the viewpoint on the visualization, including support for hand-held devices and desktop PCs, and provide search functionality and interactive filtering. Based on this framework, we explore additional modules for jointly exploring the spatial and temporal domain of the dataset and incorporating live news into the visualization.

News readers are faced with the challenge of keeping an overview of a flood of information stemming from an increasing number of news articles. More often than not, this causes readers to only focus on the newest headlines instead of browsing through less recent articles, which reduces an article's potential lifespan. This is a missed opportunity to discover relationships between news or to see how stories around a given topic evolve spatially and temporally. News articles typically contain metadata that is not visible to the reader on regular news sites, including simple information such as publication date, news categories, links to related articles, but also more content-specific information such as named entities[10] or geographic locations. This information can be used to structure data based on several criteria and enables focusing only on a subset of articles relevant to the reader. Visualizing news in a geospatial context can increase the accessibility of the data.

0272-1716 © 2021 IEEE
Digital Object Identifier 10.1109/MCG.2021.3127434
Date of publication 11 November 2021; date of current version 15 July 2022.

However, visually organizing a large amount of unstructured article data is challenging, as it requires showing only the most relevant data to the user while still providing further detail on demand. Interactively exploring the whole dataset on a global scale requires data structures that support efficient operations for the specific use-case. To implement the visualization on tablets with restricted hardware, careful optimizations are required.

In order to demonstrate the utility of the tool, we incorporate data from multiple sources of online news articles, including BBC, Reuters, and a local news portal. For building these datasets, we develop a pipeline for loading articles from the news websites and extracting metadata and georeferences for our visualization. To explore the news articles, we employ several methods of metadata visualization. We represent articles with geographic annotations as locations on a globe and cluster them interactively based on the user's viewpoint. By combining several filters, the user can explore and focus on an article subset of interest. We incorporate a replay mode, which allows the user to enter at any point in time within the range of the dataset and observe where articles appear as time progresses in fast-forward. This visualization mode combines the spatial and temporal structure of the dataset and reveals geographic hotspots of activity in news reports over time. An interface that can connect to a news feed and directly insert the newest articles into the existing dataset and visualization gives the reader the possibility to relate current news reports to historical data. In summary, we contribute the following:

- a news article data model for the representation of georeferenced news articles, time stamps, and meta-information, as well as provider-specific pipelines to process news articles from various news agencies;
- an interactive visualization system for the exploration of large collections of past and live news articles through geospatial and temporal mapping, inter-article references, and supporting interactive filtering;
- acceleration data structures to support interactive performance on desktops and tablets.

RELATED WORK

This section explores previous work approaching news article visualization and examines techniques for representing geospatial data.

Visualization of news article data: News article visualization, as a data visualization problem,[13] is traditionally a topic of text visualization.[1] With the increasing digitalization of news, information visualization and infographics became a fundamental aspect of journalism.[14] Previous work on the visualization of news stories mainly focused on analyzing topics or named entities,[10] extracted from the documents using text-mining techniques. Grobelnik *et al.*[5] provided a graphical interface for exploring named-entities across a set of news stories, where the co-occurrence of entities within a document is visualized by drawing connections between those entities. Koivunen-Niemi *et al.*[8] developed a representation of news media with a focus on visualizing article-counts in time sets of occurrences and co-occurrences of specified words. Their visualization allows seeing temporal shifts in topic coverage by the explored set of news sources. There are a few examples of previous work that examined possibilities for visualizing geospatial news data. Rinner *et al.*[11] explored visualizations of news distributions on a local scale, representing locations corresponding to articles as points on a 2D city map. They merged multiple occurrences of the same place and indicated the count as a secondary attribute using the size of the points. Another work explores how a single news story evolves temporally and spatially with a Time-Ray Map,[15] where the temporal domain is represented with circular bars around a circularly projected 2D Map.

Geospatial visualization: Geospatial data visualization[12] is a well-studied research topic, which enjoys wide-spread application in navigation systems and geographical information systems. A number of geographic mappings have been used to represent geospatial data.[9] Depending on the type of visualization, some map types have advantages over others in terms of readability and speed at which a user can understand the conveyed information. A *flat map* presents information in 2D. Depending on the covered geographic space, distances may be visually distorted. A *2D curved/raised map* maps 2D data into 3D space. The map is either flat or curved and potentially draws some information, such as lines, in 3D space. A *3D exocentric globe* visualizes a 3D globe viewed from the outside and can be explored from all sides. Finally, a *3D egocentric globe* is a 3D visualization of the globe, where the viewer is located at the center and observes the surface of the globe from the inside. Yang *et al.*[20] studied which map representations work best in a virtual reality (VR) setup, where the exocentric globe performed best. Further research went into the type of visualization on a map, including annotations on point data.[2] Relations between points can be represented by line connections. Yang *et al.*[19] compared several techniques for visualizing line connections (origin–destination flows), such as 2D lines on flat maps, flows between two copies of a map, flows raised into 3D on a 2D map, or raised flows on exocentric globes. In terms of readability, the exocentric globe performed best, with flows that are raised proportionally to the distance of the endpoints. The performance was also best when many of those flows were present, as the third dimension provided additional space to reduce flow intersections and the users could examine the connections from any viewpoint. Further research went into data flow visualization with a focus on representing many-to-many flows.[18] With *Living Globe*, Duarte *et al.*[4] showed demographic information on the level of countries using textures and bar charts. Unlike our approach, they do not employ hierarchical representations of point data, have limited support for temporal filtering, and show only two data attributes on the map. Cho *et al.*[3] interactively analyzed a Wikipedia article collection on Roman history, allowing users to explore spatial and temporal information. For spatial mapping, they used heat maps on a 2D map and stacked timelines for temporal overview and filtering. In contrast, we encode information on a globe and show directional relationships between sites on the map. *Globe AR* by Asturiano[a] presents a layer-based approach in globe visualization. Similarly to *Living Globe*, it contains a map layer to apply a texture to the globe and a layer to draw bars with a specified height. In addition, it provides layers to draw line connections between globe locations and to tessellate parts of the globe surface into hexagons with specified color or height properties. The data representation has neither levels of detail nor does it support interactive filtering. While their approach concentrates on AR aspects, we focus on the interactive data analysis in the context of news articles.

REQUIREMENT ANALYSIS AND TASKS

News articles are commonly read in news applications that present newest articles first. We collaborated with a local news portal, which is interested in new forms of spatial and temporal news article exploration, such as the observation of the spatial distribution of articles on a specific topic over time, to engage readers. In the wake of the digital revolution, traditional news agencies compete for a digital audience. Hence, the users of our application are online news readers, which includes both a general audience who wants to read about current news and how they relate spatially and temporally, as well as journalists who might want to explore a collection of news articles in a more structured form. The focus of the article exploration tool is set by the local news agency. The goal is to extend article exploration beyond the limits of sequential browsing usually provided on news websites, i.e., to support exploring articles according to their spatial and temporal information, and their inter-connection through references. By providing new means to access existing news articles, and by tapping into previously unused potential, i.e., geospatial information and article cross-references, the collaborating news portal hopes for new impulses in audience engagement and market share. Due to the heterogeneity of news article data, and the desire to explore new approaches for spatial and temporal exploration, we derived the following requirements:

R1) support various input sources of past and live news articles, e.g., from the BBC, Reuters, and articles from the local news agency;
R2) provide spatial exploration of thousands of articles without clutter and overdraw;
R3) enable exploration of references among news articles;
R4) support temporal exploration to see how topics evolve over time;
R5) support interactive filtering based on location, tags, and article categories;
R6) provide access to article details (heading, categories, etc.);
R7) support desktop and tablet devices to enable multiple usage scenarios.

While the requirements R1 and R7 are technical, the others concentrate on the interactive visual data exploration (R2–R6). From these requirements, we derived the following tasks that our system should enable us to solve, where the enumeration maps directly from requirement to task:

T1) import from multiple sources into a common news article data model, including live feeds;
T2) group spatially close articles (dependent on the user location) and access group statistics;
T3) locate articles on a map, and follow their connectivity to other articles;
T4) perform temporal aggregation, access statistics about when articles occurred, and view replay animations;
T5) apply interactive filters (spatial, temporal, meta-information), and use tag search;
T6) access details of the selected article, as well as the closest related articles;
T7) use the system on a desktop or tablet device.

In the following, we describe the data processing pipeline and elaborate on the design of the interactive visualization system such that it supports solving the aforementioned tasks.

NEWS GLOBE

In order to meet T1, we developed a common news article data model that forms the basis for all further processing and visualization steps, including spatial georeferences, time stamps, article meta-information, as well as links to other articles. We developed the system using around 45,000 news articles of the year 2019 from a local news website, 6000 news articles from BBC, and 5000 from Reuters. The datasets were collected using a news service API.[b]

Our visualization application consists of two components: the globe visualization (in 3D) and the user

[a][Online]. Available: www.github.com/vasturiano/globe-ar

[b][Online]. Available: www.newsapi.org

FIGURE 1. Main Window Sections. (1) Globe visualization. (2) Timeline with histogram. (3) Filtering panel. (4) Article details. (5) Menu.

Entry	Example
Unique ID	15330235
Publication Date	1558210279
Headline	E-scooters in our cities
Sub-Headline	Plans and challenges
Article Summary	E-scooters are now [...]
URL	https://blick.ch/15330235
Related articles	[15303162, ...]
Country ID	Switzerland
Location ID	Basel
Category ID	Economy
Tag IDs	[Bikes, Social Media, ...]

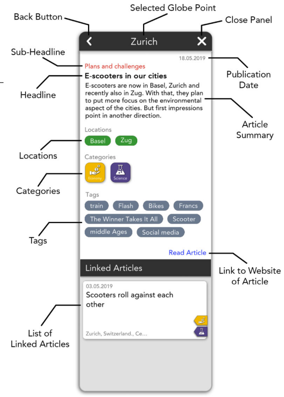

FIGURE 2. Left: Listing of the meta-information of an article. Right: The article detail view showing all information about a single article.

interface (2D layer). Figure 1 gives an overview of the system. (1) The middle of the screen contains the globe visualization itself. To meet T2, we perform a hierarchical clustering of news articles to reduce visual clutter, and explore the hierarchical representation of news articles on the globe. Connecting lines show the references to other articles, fulfilling T3. We explain the globe visualization in the "Globe Visualization" section. (2) A timeline expands along the bottom of the screen, enabling T4. It allows the user to select a specific time range as a filter for the articles and displays a histogram of articles to aid the selection. More details are provided in the "Temporal Information" section. (3) The left-hand side area either shows settings or the filtering panel for categories and tags, enabling T5. The filtering requires efficient data structures for interactive exploration and is explained in the "Interactive Filtering" section. (4) To address T6, the space on the right-hand side of the screen contains all information on selected articles, such as categories, tags, referenced locations, linked articles, and the text itself. We cover the article UI in the "Article Details" section. (5) The menu bar on the left-hand side of the screen gives access to replays, live feed, and settings. UI elements and user interactions are adapted for both desktop and tablets to meet T7. Considerations for responsive device-agnostic design are discussed in the supplemental material. The elements 3) and 4) are hidden when not in use to provide more space for the globe view 1).

Data Pipeline

To support news data from multiple sources (T1), we implement a data processing pipeline that transforms the data to a common format. Due to a lack of standardization in the article metadata formats, each news API requires a separate data parser. Figure 2 (left) summarizes the data entries of our common article format, as required in our visualization with an example of a complete news entry. An entry needs to contain the article's headlines, summary, plain text, the publication date, the link to the website, and optionally references to other entries. These requirements are fulfilled by most articles on news websites and also by our three sample datasets. Some of the articles' properties are used directly without further processing. These include the headline, subheadline, and a leading summary text, as well as the metadata on links to related articles, the publication date, and a URL to the website, where the full article is accessible. If no related links are provided in the dataset, we extract the URLs from the website of the article to find linked articles in the dataset. For tagging and classification, the raw articles are processed using the natural language processing API *TextRazor*.[c] It classifies the text into a set of categories and subcategories from the *IPTC Media Topic NewsCodes taxonomy*. For our visualization purposes, we use the 17 top-level categories with the highest article counts, which account for 99.8% of all category assignments. We extend the list of categories with an additional *Others* category and assign it to all articles not belonging to any of the chosen top categories. The extracted tags from *TextRazor* are words or combinations of words that appear within the article and are classified into tag-categories. We map those tags to IDs and store them as a list in each article entry. In addition, we keep a list of all tags within the *location* tag-category for further

[c][Online]. Available: www.textrazor.com

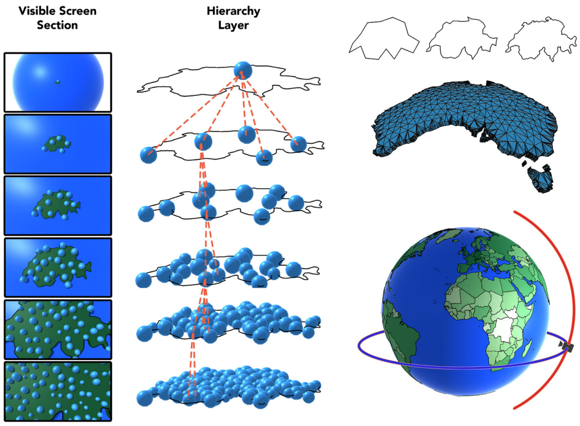

FIGURE 3. Illustration of the level-of-detail concept, showing (left) locations visible by the camera, (middle) the corresponding hierarchy level, (top right) level-of-detail borders, (middle right) country mesh, and (bottom right) camera navigation.

processing. These location tags have no geographic information assigned. To find geographic coordinates for these tags, which is commonly referred to as geocoding, we perform a lookup in the OpenStreetMap database,[d] which returns a list of matching location nodes together with their geographic coordinates. We merge location tags that map to the same location and assign each location a unique ID. In the article entry, we replace location tags with the IDs and export a separate map from location IDs to location name and coordinates. This location extraction method is not always accurate. For the purpose of this work, we assumed a sufficient data quality. Approaches to improve the data quality are discussed later. Besides locations, in general, we store with each article specifically to which countries it refers. To detect whether a certain location name is a country, we match it with a list of country names and variations of country names. The countries are also mapped to IDs and a list of country IDs is stored in each article entry.

Globe Visualization

Location hierarchy: Each article references a set of locations. Our largest dataset contains 17,343 unique locations. To add all locations, we incorporate an agglomerative hierarchical clustering[7] based on geodesic distance (T2). When merging locations, the higher level nodes in the hierarchy are represented by one of the merged locations. We perform the selection of the representative based on the highest article count of the individual subnodes. Figure 3 (left) shows selected cluster levels of locations in Switzerland. The clustering enables the exploration of various levels of details and is essential for the overall performance as it reduces the number of locations that need to be rendered and updated (T7).

Globe: Several map representations are available for geospatial visualizations (cf., the "Geospatial visualization" section), each with their advantages and disadvantages for different use cases. A fundamental design decision is whether to use a 2D or 3D map representation. A 2D map has the advantage that there is no occlusion of the hemisphere compared to a 3D globe representation. Drawing line connections between locations on the map in the third dimension (2.5D and 3D) helps disambiguating and resolving line crossings when elevating the lines proportionally to the distance of the locations.[19] An advantage of 3D globe representations is that line connections are not distorted. This is contrary to 2D maps, where lines cannot always be drawn along the geodesic, e.g., across the poles. Based on these considerations, and given that our visualization should support line connections, we decided to use a 3D exocentric globe. Furthermore, we plan to add support for VR/AR in future development, for which a 3D exocentric view is preferable.[20] Our 3D globe visualization consists of a blue sphere that represents the water and individual country meshes that represent all land areas. To enable an efficient navigation around the 3D globe (T7), we created multiple levels of details for the country polygons, which are selected based on the camera position, as shown in Figure 3 (top right). To provide aggregated information on the level of countries (T2), we color each country mesh in a logarithmic scale based on the number of articles referring to it using maps from ColorBrewer,[e] and map the number of articles logarithmically to the point radius (T2), using the linearized mapping by Schumann and Müller.[16] Later, we add annotations to the points, showing the location name of the representative hierarchy node. Users can rotate the sphere around two axes and zoom the view; on a desktop PC via mouse interaction and scrolling, on handheld devices via movement and pinching (T7). Camera flights are animated when selecting a different location. Selected locations or countries are highlighted by color.

Categories: Each article in our dataset was assigned up to five out of the 18 categories.

[d] [Online]. Available: www.openstreetmap.org

[e] [Online]. Available: www.colorbrewer2.org

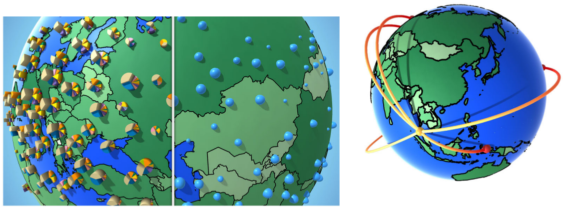

FIGURE 4. Depiction of visualization elements, including (left) pie charts or (middle) spheres, and (right) location references via arcs.

Throughout the application, we use for each category a combination of color[6] and icon[f] [see (3) in Figure 1]. We optionally encode the distribution of categories (T2) within a hierarchy node using pie charts, as shown in Figure 4 (left). Pie charts are a commonly used distribution representation that our users are familiar with. There are, however, well-known perceptual limitations[17] that limit the utility of pie charts considerably. Hence, we must be clear about what the visualization can do and what it cannot do. Angles are notoriously difficult to read correctly. Thus, the exact numeric values cannot be perceived. Instead, our goal is to provide an overview on which categories are most dominant in a certain region, which does not require reading the exact percentages of pie chart segments. To facilitate comparison between the segments, we use the third dimension to adapt the segment height proportionally to the segment size. This allows for the comparison of segment sizes when observing the charts from the side. To reduce visual clutter, the scale of the pie charts is adjusted based on the distance at which the visualization is observed and can be further adjusted by the user with a scale factor in the application settings. As an alternative to the pie chart representation, uniformly colored spheres can be used for less visual information [see Figure 4 (middle)].

Line connections: To visually indicate which locations are referenced from the selected article, we draw line connections to the respective locations (T3). The lines are drawn as tubes in 3D space and have a color gradient to indicate the origin and destination of the relation between the points on the globe. An animation starting from the origin conveys the line direction. Figure 4 (right) shows lines with the yellow point as the origin and the red points as destinations. To simplify distance comparisons between the endpoints, we increase the line elevation with the distance as proposed by Yang *et al.*[19]

[f][Online]. Available: www.flaticon.com

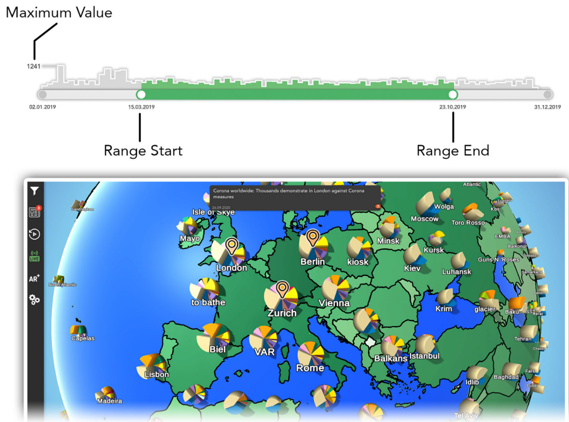

FIGURE 5. Top: Example of timeline with histogram for temporal filtering. Bottom: Markers indicating article appearance in live or replay mode.

Temporal Information

We implemented a timeline filter (T5) [see Figure 5 (top)], showing the date range of the articles, which expands along the bottom of the screen [see (2) in Figure 1]. To support the user in selecting a time range and to show the temporal distribution (T4) of articles, we display a histogram. For a responsive visualization on desktop and tablet devices, the granularity of the bins is chosen based on the screen size (T7).

Replay and live mode: The temporal evolution of article appearances can be visualized in the replay mode (T4), and new articles can be inserted automatically in the live mode (T1) into our data structures. Both modes are accessible through the menu [see (5) in Figure 1]. The appearance of an article is visualized with an animated location marker and the article title is displayed at the top of the screen, shown in Figure 5 (bottom). To show news appearing on the backside of the globe, the location markers are clamped to the globe horizon. When a marker is located behind the globe, we indicate the spatial information by reducing its scale depending on how far on the backside it is located. Like the globe points, the location markers are clustered to reduce clutter when many markers are present at the same time (T2). The brightness of the marker color indicates how many new articles are referencing a location within its hierarchy node, where darker colors indicate higher counts.

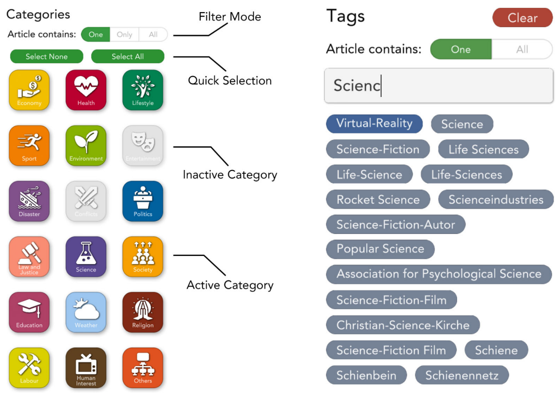

FIGURE 6. Left: Depiction of category filters with filter modes. Right: Tag filter list with best matches. Here, *Virtual-Reality* is already selected.

Interactive Filtering

To enable T5, several filters—including the temporal filter mentioned previously—can be applied. The filtering of articles is reflected in the globe points, pie charts, country color, and article lists.

Categories: Often, readers of news articles are interested in specific news categories only. For this reason, most online news portals display articles based on categories. Similarly, we implemented category filtering using the category annotations of the articles. The UI for this functionality is in the filter menu [see Figure 1 and Figure 6 (left)].

Tags: Compared to filtering with categories, searching for specific tags in a dataset gives a more precise control on the topic of the articles. The tag filter UI is located in the filter panel right below the category filters [see Figure 6 (right)]. Based on the search text input, a list of tags is proposed, from which the user can select tags for filtering. To match the search input to all tags in our dataset, we use the Levenshtein ratio and present the best matches on top of the list.

Performance Optimizations

Modifying a filter requires us to update all visuals presenting article data directly or in an aggregated form. As filters can change every frame, such as when dragging the handle of the timeline filter, we aim to achieve the necessary performance of these updates for interactive frame rates (T7).

All three filters described above are applied to the articles, either including or excluding them. To be shown in the visualization, the article needs to be included by all filters. Our implementation stores for each article an exclusion counter and for each filter the current inclusion status of all articles. When one of the filters changes, we compute for each article its new inclusion status of this filter and compare it to the previous one. If the inclusion status changed, we update the article's exclusion count. As long as the exclusion count is larger than zero, the article is excluded. When a filter changed, we store a filter ID with this newest filter operation. The different parts of the visualization that need to refresh on filter updates check with this ID whether they need to update. When a filter changes, only the visible points on the globe are updated. Each of them performs a lookup in the exclusion counts for its articles to update the point size and category counts for the pie charts. The partial update is the most critical optimization, as only a small fraction of all globe points is visible simultaneously. Whenever a new globe point appears, it checks with the filter ID whether there was an update and refreshes its appearance accordingly. By doing so, the cost of the filtering operation spreads over several frames, and intermediate updates on globe points that are not visible can be skipped. These performance improvements lead to about one to two orders of magnitude faster updates, depending on the filter settings, dataset size, and the viewpoint, compared to updating all points at once.

Article Details

The articles can be accessed through the locations they are referencing. When the user selects one of the globe points or a country, a list of all articles that reference a selected location or country is shown (T3) on the right-hand side of the screen (see Figure 1). To show all details on a specific article, the user can select one of the article cards from the list to obtain more details. The detail view is displayed in the same window as the article list on the right-hand side of the screen, providing more information (T6). Figure 2 (right) gives an example. At the top, the subheadline, headline, and summary text, as well as the publication date of the articles are shown. A clickable list of the linked locations is displayed below. Whenever an article is selected and shown in the article detail view, lines are drawn on the globe to all referenced locations of this article (see Figure 4). The origin of the lines is chosen to be the closest referenced location of the article to the currently selected location hierarchy. When selecting one of the locations from the clickable list, the camera plays a flight animation to this location (T3), allowing the user to associate the geographic location to the name. Further, the article detail view shows the list of categories and tags

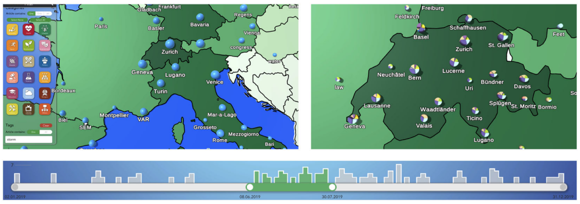

FIGURE 7. Use cases. Top left: Spatial distribution of articles filtered with the *storm* tag. Top right: Pie charts show the categorical distribution of articles. Bottom: Timeline shows the temporal distribution of articles and a selected time range (bottom).

belonging to this article. The link on the bottom right leads the reader to the website of the full article. Below the article details, the linked articles are listed. From there, the user can directly navigate to related articles, as well.

RESULTS

In this section, we demonstrate the geolocalized news article visualization on a use case that requires a system which enables tasks T1–T7. We analyze and improve our method with a user study and evaluate the performance and memory requirements. For further results, we refer to the additional material and the accompanying video, available online.

Use Case

In the following, we consider a use case where a user is interested in interactively exploring the spatial and temporal evolution of news articles on storms. Using our system, the relevant tags can be selected for filtering (T5). The globe then shows only the articles for the selected filters [see Figure 7 (top left)]. Through the grouping and aggregation of the articles depending on the viewpoint, the user can decide on the spatial granularity of the visualization (T2). When enabling the pie charts, the user can see which categories are most relevant for the selected tag filter, which in this case are mainly the weather and disaster category [see Figure 7 (top right)]. The timeline shows the distribution of articles for the selected tag filters (T4). This reveals hotspots of activity, which in our case are for the most part during summer time. From there, the user can select a desired time range, such as an active period during summer [see Figure 7 (bottom)], and observe in replay mode [see Figure 5 (bottom)] when and where news on this topic appeared. If the user is interested in an individual article within the selected range, the user can inspect its details [see Figure 2 (right)] (T6) by selecting it on the globe. This also shows line connections to related articles [see Figure 4 (right)] (T3) on the globe, allowing the user to further explore the topic.

User Study

To evaluate how well our tool meets the requirements and to evaluate whether the visualization enables the tasks, we conducted an informal user study with 10 participants performing an interactive exploration of news articles with the tool. Participants were asked to answer task-inspired questions during which we observed the participants. The observations and informal feedback given by the participants were used to further improve the tool. The participants hold a profession that requires daily work at a computer, their age ranges between 25 and 33, and they are readers of conventional news websites. One participant has a color vision deficiency, all others have normal or corrected to normal vision. They participated in the user study remotely and answered the questions on a laptop or desktop computer with screen sizes from 14 to 24 inches.

Procedure: The user study consisted of three parts. First, we explained the features of the tool and let the participants familiarize themselves with the system by freely roaming around the globe through filtering and navigation. Overall, this part lasted for about 10 minutes per participant. Second, we asked the participants to answer seven questions using our application. The order of the questions was randomized between participants to remove memory effects. For further evaluation, we recorded the audio and captured the screen during each session. The seven questions were designed in a way that answering them required the participants to use the main features of the tool related to the tasks T1–T7. Since answering the questions required a tool supporting all these tasks, none of the questions could be solved in a comparable time frame with a conventional news website. Answering the questions required the use and interpretation of the following features and visualizations [besides the technical related tasks (T1 and T7) which are necessary for all of them]: grouping and statistics (T2) for Q1–Q3, connectivity (T3) and article details (T6) for Q4 and Q5, and temporally aggregated statistics (T4) for Q6 and Q7, as well as filtering (T5) for all of them except for Q3. The full description of the questions can be found in the supplemental material, available online. Whenever a participant required a hint or a reminder, we provided help not earlier than 25 seconds after posing the question. The times when this was necessary are reported later along with the actual time it took the participants to answer. The third part

TABLE 1. Time (in seconds) required to answer the questions: (●) w/o help, (●) w/ help, (●) wrong answer.

T2			T3, T6		T4	
Q1	Q2	Q3	Q4	Q5	Q6	Q7
07	07	18	13	09	26	12
12	08	30	14	09	28	18
14	11	40	15	10	33	25
18	12	65	17	18	36	27
23	17	66	19	25	46	27
27	17	66	21	28	48	31
40	37	85	25	30	53	43
42	37	97	35	32	71	47
95	69	110	124	48	73	55

Notes: Columns are sorted by time, showing measurements for individual participants. Headline shows the connection between the questions and the tasks. In addition, all questions require T1 and T7, as well as T5 for all except Q3.

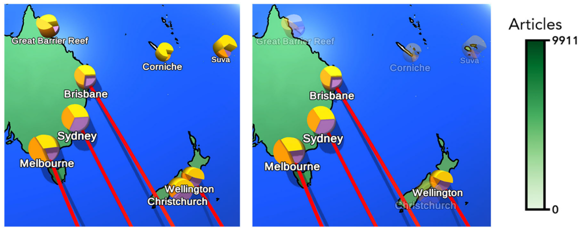

FIGURE 8. User study uncovered problems that we fixed. On the left, pie charts and labels are shown before the improvements. On the right, the updated transparent pie charts and labels are shown when an article is selected, as well as the country color legend that was added.

of the procedure took place during and after the practical sessions, when we asked the participants to report their feedback. This included observations made during the initial free roaming and contained their agreement with nine statements on a 5-point Likert scale.

Findings: Table 1 reports the time (in seconds) needed by the participants to answer the individual questions. The timings are grouped by task and are sorted by time for each question in increasing order. It can be seen that most questions have been answered correctly (shown in blue) in a short time. Once help was needed (shown in orange), the participants tend to take more time to answer. We hinted which feature to use earliest after 25 seconds and only when the user was lost deciding how to answer the given question. Besides two participants, everyone needed a reminder of the meaning of the country color to answer question Q3. We observed, that the participants required less help for questions related to task T4. Help was required related to zooming to the right hierarchy level (Q6). Furthermore, there are few wrong answers shown in red. Reasons for them are not knowing the location of a place on the globe (Q1) or wrong interpretation of the color (Q3). For question Q5, we observed that, once zoomed in, the participants did not pay attention to the lines showing linked locations and instead reported all visible cities [see Figure 8 (left)], leading to wrong answers. Their numbers are not correlated. On the other hand, most users were able to interpret the size of the points and height of pie chart segments (Q1 and Q2), to find articles at specific locations and to use the timeline to analyze and filter the temporal distribution (Q6 and Q7) correctly.

In summary, both, the evaluation of the answers to the questions and the provided feedback indicate that the temporal exploration (T4) worked well. The spatial exploration (T2 and T3), however, required some general knowledge about geographical locations to make use of the tool. Thus, some users had more difficulties to answer the spatial exploration questions. In terms of visualization, we indicated some features that are not directly understandable, such as the country color and the location references. In the following, we report improvements we made to address these findings.

Improvements: Based on the findings in our user study, we implemented two improvements. First, we added a legend of the country color to the overview as a visible explanation of the relationship of the country color with respect to the number of articles to replace the hint needed in question Q3. The legend is shown in Figure 8 (right). Second, to address wrong answers in Q5, we use not only lines to the referenced locations but additionally fade out nonconnected locations [see Figure 8 (left, middle)].

Performance

In this section, we report the performance and memory consumption of our Unity implementation. Both the memory and the performance are inherently dependent on the size and the structure of the dataset. We use for our measurements the largest news dataset described in the "News Globe" section. All the timing and memory measurements are taken on an Intel Core i7 3.5 GHz with four cores, 16 GB RAM, with an NVIDIA GeForce GTX 660 and on a Samsung Galaxy Tab S6.

Table 2 lists the frame render times. The *idle* time shows the performance when no user interactions take place and no animations are shown. The pie chart meshes, which refresh the category counts and animate the changes of the pie chart sections, are some of the most performance intensive parts.

TABLE 2. Rendering time per frame in milliseconds (averaged over 10 s) for different tasks on PC/tablet.

	Spheres		Pie charts	
Task	**PC [ms]**	**Tablet [ms]**	**PC [ms]**	**Tablet [ms]**
Idle	0.7	1.1	1.1	1.6
Zooming	3.8	13.7	18.3	29.8
Rotating	3.7	8.7	7.3	12.4
Filtering	31.1	99.2	36.4	119.2
Replay	2.4	10.3	2.4	10.3

Therefore, we perform the measurements with both the pie charts, and the more lightweight spheres. We can see the performance impact of the pie charts, which mainly stems from the segment animations when pie charts change their values, but also from the additional rendering cost for the more detailed meshes compared to the spheres. The performance influence of the pie chart updates can be regulated by the mesh resolution of the individual sections. When *zooming* with the camera closer to the globe, many globe points are loaded, and their sizes and meshes are updated simultaneously. Therefore, we can again see a higher impact when the pie charts are active. Pure *rotations* of the camera around the globe have a similar effect as the zooming, as new hierarchy nodes are loaded when showing new regions. However, the number of node updates, and hence the performance impact, is much lower compared to zooming. The *filtering* of articles is the most performance-intensive task. The cost mainly results from checking the inclusion for each of the articles with the updated filter. In addition, there is the update cost of the visible globe points, which is higher for the pie charts as described above. However, filtering is most often only performed during a single frame when a category or tag filter changes, or during a few frames when dragging a filtering handle on the timeline. Therefore, we found a reduction in the frame rate to around 10 FPS on tablet devices to be sufficient to conclude that filtering does not impede the article exploration experience. For bigger datasets, it is imaginable to spread the cost of filtering over several frames. The *replay* mode shows markers for the locations instead of the 3D point representations. Consequently, there is no difference between the point visualization modes. The markers are also more lightweight than the sphere and pie chart modes, as they only adapt their color to reflect the number of articles.

When developing for tablets, memory can be a limiting factor. We profiled the memory usage to get an understanding of which components require most resources. The 233 country meshes, the globe, and the pie chart meshes required about 109 MB. Rendering requires 88 MB including shaders, textures, and materials. The UI allocates 36 MB for 2D textures, including images and fonts. The data structures used for storing the articles and locations for efficient access and the location hierarchy and its state require 144 MB in total. The remaining memory usage of 36 MB mainly results from the Unity internal systems, such as the component and object system.

Discussion

Data quality: The georeferencing, i.e., the translation from text to geographical coordinates can contain errors, e.g., due to ambiguities, which may mislead the user when exploring the data. For better data quality, location-information could either be annotated manually when writing the article or possibly with the aid of tools that propose locations based on the text. There are also automatic location extraction solutions, such as geoparsepy[g] that extract and disambiguate locations from text. For this article, we assumed that more reliable geolocations will be available in the future, and opted for a mapping through the OSM database.

Static hierarchy: Our current approach builds a static data structure for the location hierarchy based on the collection of all locations appearing in the dataset, which proved sufficient for us. Adaptive clustering methods could yield better results in showing relevant locations when reducing the dataset through filtering or when inserting articles to the dataset with new locations. However, temporal coherence will then become an issue.

Clustered geolocations: We explored several visualization techniques to show different aspects of the articles and their geolocations. Through clustering, the exact locations of the articles are not always visible. An interesting extension for visualizing the article's exact location could be to explore the usage of heat maps, by coloring the globe based on the number of articles in a given area, which would show precise locations of hotspots of articles.

Pie charts: To display the category distribution at spatial locations, we chose 3D pie charts with a redundant height encoding. Pie charts are discouraged by visualization experts[17] due to inaccurate angle perception and under-representation of small portions. Thus,

[g] [Online]. Available: pypi.org/project/geoparsepy

pie charts can at best give a qualitative view on how the individual portions add up to one, where differences must be large enough to be distinguishable. The 3D pie chart further suffers from perspective distortion, changing the apparent size of slices. To compensate for this to some extent, we added the encoding by height. However, it is clear that this alone cannot solve all problems. If a precise encoding of numeric values is of importance, another visual encoding is needed.

CONCLUSIONS

In this article, we presented a method to visualize geolocalized information of news articles on a global scale. We developed a 3D globe visualization for article exploration. With agglomerative hierarchical clustering on the locations, the visualization can adapt its level of detail based on the user's viewpoint and provide details on a local, as well as on a global scale. Several options for narrowing the dataset to a specific topic, category, or time window through filtering enable the user to focus on an article subset of interest. Our method incorporates features for joint exploration of the spatial and temporal domain of the articles through a replay mode. We optimized the performance-critical parts of the application to achieve interactive frame rates on tablets. Based on three news datasets, we were able to design an application that fulfills the given requirements for geolocalized article exploration. The resulting method improves the accessibility of large article datasets, while also incorporating recent news in the visualization by loading live articles.

In the future, we would like to focus on visualizing how a news story evolves over several articles. By building directed links among the articles with related topics or locations that point forward in time, interactive storylines with branches through the article dataset could be built.

ACKNOWLEDGMENTS

This project was supported in part by Ringier, in part by TX Group, in part by NZZ, in part by SRG, in part by VSM, in part by viscom, and in part by the ETH Zurich Foundation. This article has supplementary downloadable material available at https://doi.org/10.1109/MCG.2021.3127434, provided by the authors.

REFERENCES

1. M. Alharbi and R. S. Laramee, "SoS TextVis: A survey of surveys on text visualization," in *Proc. Comput. Graphics Vis. Comput.*, G. K. L. Tam and F. Vidal, Eds., The Eurographics Association, 2019, pp. 143–152, doi: 10.2312/cgvc.20181219.
2. M. A. Bekos, B. Niedermann, and M. Nöllenburg, "External labeling techniques: A taxonomy and survey," *Comput. Graphics Forum*, vol. 38, no. 3, pp. 833–860, 2019.
3. I. Cho, W. Dou, D. X. Wang, E. Sauda, and W. Ribarsky, "Vairoma: A visual analytics system for making sense of places, times, and events in Roman history," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 210–219, Jan. 2016.
4. E. Duarte, P. Bordonhos, P. DIAS, and B. S. Santos, "Living globe: Tridimensional interactive visualization of world demographic data," in *Int. Conf. Human Interface and the Management of Information* (Lecture Notes in Computer Science 9734). Cham, Switzerland: Springer, 2016, pp. 14–24.
5. M. Grobelnik and D. Mladenic, "Visualization of news articles," *Informatica (Slovenia)*, vol. 28, no. 4, pp. 375–380, 2004.
6. P. Green-Armytage, "A colour alphabet and the limits of colour coding," *Color: Des. Creativity*, vol. 5, pp. 1–23, Aug. 2010.
7. S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.
8. L. Koivunen and M. Masoodian, "Visualizing narrative patterns in online news media," *Multimedia Tools Appl.*, vol. 79, no. 1/2, pp. 919–946, 2020.
9. S. Nusrat and S. G. Kobourov, "The state of the art in cartograms," *Comput. Graphics Forum*, vol. 35, no. 3, pp. 619–642, 2016.
10. D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Lingvisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.
11. C. Rinner, A. Komaromy, and A. Lindgren, "Geospatial tools for the visualization and analysis of local news distribution," *Future Local News: Res. Reflections, Ryerson Journalism Res. Centre*, pp. 1–22, 2018. Accessed: Dec. 07, 2021. [Online]. Available: https://futureoflocalnews.org/portfolio-item/geospatial-tools-for-the-visualization-and-analysis-of-local-news-distribution/
12. T. Rhyne and A. MacEachren, "Visualizing geospatial data," in *Proc. Int. Conf. Comput. Graphics Interactive Techn.*, Aug. 2004, pp. 1–141, doi: 10.1145/1103900.1103931.
13. R. Raghav, S. Pothula, T. Vengattaraman, and D. Ponnurangam, "A survey of data visualization tools for analyzing large volume of data in big data platform," in *Proc. Int. Conf. Commun. Electron. Syst.*, 2016, pp. 1–6.
14. G. Smit, Y. De Haan, and L. Buijs, "Visualizing news," *Digit. Journalism*, vol. 2, no. 3, pp. 344–354, 2014.
15. J. Sheidin, J. Lanir, P. Bak, and T. Kuflik, "Time-ray maps: Visualization of spatial and temporal evolution of news stories," in *EuroVis 2017 - Short Papers*. Norrköping, Sweden: The Eurographics Assoc., 2017.

16. H. Schumann and W. Müller, *Visualisierung - Grundlagen Und Allgemeine Methoden*. Berlin, Germany: Springer, 2000.
17. E. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT, USA: Graphics Press, 2001.
18. Y. Yang, T. Dwyer, S. Goodwin, and K. Marriott, "Many-to-many geographically-embedded flow visualisation: An evaluation," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 1, pp. 411–420, Jan. 2017.
19. Y. Yang, T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen, "Origin-destination flow maps in immersive environments," *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 693–703, Jan. 2019.
20. Y. Yang, B. Jenny, T. Dwyer, K. Marriott, H. Chen, and M. Cordeil, "Maps and globes in virtual reality," *Comput. Graphics Forum*, vol. 37, no. 3, pp. 427–438, 2018.

NICHOLAS INGULFSEN is currently a full stack Software Engineer at PHSG, St. Gallen, Switzerland. His research interests include computer animation, visualization, and real-time rendering. He received the M.Sc. degree in computer science from ETH Zurich, Zurich, Switzerland, in 2020. Contact him at nicholas.ingulfsen@gmail.com.

SIMONE SCHAUB-MEYER is currently a postdoctoral researcher at the Visual Inference Lab, Technical University of Darmstadt. Her research interests include video processing, motion representation, and video frame synthesis. She received the Ph.D. degree in computer science from ETH Zurich, Zurich, Switzerland, in 2018. Contact her at simone.schaub@visinf.tu-darmstadt.de.

MARKUS GROSS is currently a Professor of computer science at ETH Zurich, Zurich, Switzerland, the Head of the Computer Graphics Laboratory, and the Director of Disney Research Studios, Zurich. His research interests include physically based modeling, computer animation, immersive displays, and video technology. He received the Ph.D. degree in computer graphics and image analysis from Saarland University, Saarbrücken, Germany, in 1989. Contact him at grossm@inf.ethz.ch.

TOBIAS GÜNTHER is currently a Professor of computer science at the Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany. His research interests include visualization, light transport, and real-time rendering. He received the Ph.D. degree in scientific visualization from the University of Magdeburg, Magdeburg, Germany, in 2016. He is the corresponding author of this article. Contact him at tobias.guenther@fau.de.