VisGets: Coordinated Visualizations for Web-based Information Exploration and Discovery

Marian Dörk, Sheelagh Carpendale, Christopher Collins, and Carey Williamson

Abstract—In common Web-based search interfaces, it can be difficult to formulate queries that simultaneously combine temporal, spatial, and topical data filters. We investigate how coordinated visualizations can enhance search and exploration of information on the World Wide Web by easing the formulation of these types of queries. Drawing from visual information seeking and exploratory search, we introduce *VisGets* – interactive query visualizations of Web-based information that operate with online information within a Web browser. VisGets provide the information seeker with visual overviews of Web resources and offer a way to visually filter the data. Our goal is to facilitate the construction of dynamic search queries that combine filters from more than one data dimension. We present a prototype information exploration system featuring three linked VisGets (temporal, spatial, and topical), and used it to visually explore news items from online RSS feeds.

Index Terms—Information visualization, World Wide Web, information retrieval, exploratory search, visual information seeking.

◆

1 INTRODUCTION

Searching for information on the World Wide Web is a fundamental task undertaken daily by millions of people. Furthermore, this computer use is expanding into more and more aspects of society, as people increasingly turn to the Web as an immediate source for news, references, and entertainment.

While the seemingly-endless information space of the Web contains diverse rich-media content, finding information on the Web is still generally done using an ordinary text query. This approach is demonstrably useful, in that people routinely find *something* useful, even if it is not exactly what they were seeking. However, searching can be frustrating when queries return thousands of hits, many of which are extraneous. Difficulties are even more likely when a person's information need is only vaguely defined. Choosing the right keywords for the search query may be difficult, and the text-based result list itself provides little contextual overview to promote general understanding. A common tactic is to issue multiple (slightly different) queries, and to look mainly at the first few items in the list of results. While this can lead to success, it just as often leads to long and frustrating searches with imperfect results [19].

Today's Web features thriving online communities, rich media content, maturing Semantic Web standards, and an increasing number of geographically referenced resources. This structure is quite different from the early days of the Web, which was dominated by unstructured textual information. Information published on today's Web is becoming ever more complex: it includes not only multimedia, but rich links between information fragments indicating semantic, social, and spatial relationships. Unfortunately, the traditional Web search processes do not reflect these advances. Search remains primarily textual, and result lists are still quite unstructured.

In this paper, we introduce VisGets as a means to support visual information exploration on the Web. VisGets are information visualization widgets whose manipulation constructs a Web query. The general goal is to support the information seeker in gaining casual insight into large collections of Web resources. While the concept of filtering Web-based information through interactive visualization is quite broad, we chose to start by considering three increasingly common dimensions of Web-based information: temporal, spatial, and topical (as evidenced by tags). Drawing upon the concept of dynamic queries [29]—which has been successfully applied to local, static databases [2], GISVis systems [8, 10], and StatsVis software [16, 34]—VisGets extend and adjust the basic concept applying it to dynamic, distributed, Web-based data. These extensions include weighted brushing, developed to reveal the degree of relatedness between data elements, and delta queries, which provide information about the change in query results and improve interactivity through minimizing data movement. Map visualizations of photos [1] and tags [30, 41] combine spatial and topical dimensions into one view and thus allow initial exploration of multi-faceted information. In contrast VisGets provide separate, yet linked visualizations that are used for query formulation along multiple dimensions.

To assess people's response to this type of information access, we conducted an exploratory user study, looking at both specific and more open-ended search tasks.

2 RELATED WORK

A wealth of literature on navigation and exploration of information spaces appears within and across several research communities. In this section, we highlight some of the key ideas underlying our work.

2.1 Picking, Foraging, and Exploring Information

Using metaphors from foraging behavior, the concepts of "berrypicking" and "information scent" describe navigation between numerous online resources. With berrypicking, the static search concept is replaced with a dynamic understanding of searching and browsing, where searchers roam somewhat randomly between different sets of documents. By gathering multiple information bits, the searcher learns about the content and consequently changes his/her information needs, and thus the subsequent queries [6]. An extension of berrypicking is the notion of *exploratory search*, which starts with a vague information need, and then leads to a learning process [22, 37]. Information scent is related to berrypicking, but puts less emphasis on the changing information needs. Instead, it focuses more on proximal cues that direct the searcher to more interesting information items or navigational elements [9, 39]. However, both berrypicking and information scent describe a type of low-level navigation in which individual resources or navigational elements may guide the information seeker, but they do not provide general overviews.

Faceted navigation utilizes the fact that information spaces can be organized along multiple dimensions or facets. Instead of employing only a single organization scheme, a faceted classification of information provides the searcher with multiple content-oriented facets that

---

- *Marian Dörk is a MSc student with the Computer Science Department at the University of Calgary, e-mail: mdoerk@ucalgary.ca .*
- *Christopher Collins is a PhD student with the Computer Science Department at the University of Toronto, e-mail: ccollins@cs.utoronto.ca .*
- *Sheelagh Carpendale and Carey L. Williamson are Professors with the Computer Science Department at the University of Calgary, e-mail: {sheelagh, carey}@cpsc.ucalgary.ca .*

Manuscript received 31 March 2008; accepted 1 August 2008; posted online 19 October 2008; mailed on 13 October 2008.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org .

can be explored without entering explicit search queries [11]. Facets can be understood as orthogonal, non-exclusive categories that describe multiple aspects of information sources. While faceted navigation provides a better overview on a information collection, the manual creation and actualization of faceted classifications are work-intensive and require domain knowledge.

2.2 Visualization for Information Seeking

Previous work has shown that interactive visualizations can considerably improve the exploration of data, suggesting that information seekers should be able to "rapidly, safely, and even playfully explore a database" [29]. Graphical user interface elements can enable visual information seeking via *dynamic queries* [2]. An extension of dynamic queries displays visual overviews beyond the current query constraints and thus provides cues for how changing a query may yield a satisfying set of results [31]. In offline data sets, manipulation of successive selections in a visualization can be used to consecutively reduce the data to subsets [16, 34] and dynamically adjusting the Boolean combination of text-based queries can be used to affect visualizations of relations among items of interest [32, 18]. To help the information seeker assess the relevance of search results numerous techniques have been developed that provide visualizations of document-query similarity and document characteristics [15, 20, 23]. Search results from multiple sources and of different media type can also be combined into a layout visualizing relationships with links and tags [33].

Coordinated views may yield deeper insights into these multiple dimensions or facets by showing the interdependencies among the information sources [5]. For example, geographic information can be explored in an interface where two linked visualizations are provided for spatial and conceptual domains [8]. Interactive visualizations featuring maps, topics, and timelines have been employed to explore video libraries [10], however, the used map does not display spatial distribution of items. Coordinated views in general have not been applied to Web-based information seeking.

2.3 Information Visualization on the Web

The Web can be seen as a medium for information visualization that may constitute both a data source and a delivery mechanism [27]. Improved support for Web standards increasingly allows simple interactive visualizations within the Web browser. For example, interactive widgets and visualizations [13] have been embedded into Web pages without requiring browser extensions, however the spatial distribution of items along the attributes that can be filtered with sliders are not visualized and the whole data necessary for the visualization has to be loaded at once, making exploration of larger information spaces impossible. A particular challenge is to provide a responsive and interactive interface even though the data to be explored is distributed over the network. Though research is progressing towards providing a responsive and interactive interface when the data is distributed over the Web —such as dividing query process into query preview and query refinement [25] and basing choices about placement on process on client- or server-side capabilities on the visualization's complexity [40]—this remains an open challenge.

As information on the Web becomes more structured and semantically organized [7], visualization may constitute a more intuitive way to access vast quantities of Web-based information. Publishing and interacting with structured information on the Web becomes easier [17]. Large-scale information repositories such as Wikipedia can be turned into semantic information repositories, which can be queried like a database [4]. The multitude of RSS (Really Simple Syndication) feeds of millions of Web pages already contain structured metadata on newly added posts or recent changes on blogs and news sites.

Two dominant visualizations on the Web are tag clouds and geographic maps providing visual access to the conceptual and spatial domain respectively. A *tag cloud* visualizes a list of words in which differing font sizes represent differing frequency in use. While tag clouds have been used to provide overviews of search results [21], studies have shown that tag clouds can support a range of information seeking tasks [26]. Examples of the Web becoming increasingly geospatial [28], include Flickr's photo map [1] and the use of location information to structure and explore multimedia [8]. It has been shown that a large photo collection featuring both keywords and spatial information may help to label locations on a map [1]. Furthermore, the tag clouds and maps have been combined into one view to support initial exploration of large, multi-faceted data sets [30, 41]. Up to now tag clouds and maps have not been used together to allow multidimensional query formulation for information items.

The recent interest in socially-oriented visualization on the Web is illustrated by systems like the NameVoyager[36], We Feel Fine [14], Exhibit [17], and ManyEyes [35]. These all provide interactive visualizations on the Web, however, the data sets are selected by the creator, except for ManyEyes where the users can upload their own data. WeFeelFine does use a creator selected set of data collected from the Web, however, while users can choose which subsets of this data they will see visually, there are still open questions about how to support visually based querying of Web-based information.

3 DESIGNING VISGETS

We extend the movement toward visual information exploration on the Web, by developing interactive visualizations that provide query functionality and update the visual control with information about the data retrieved by the current query settings. These are InfoVis query widgets (VisGets) that combine visual representations of the data being queried with retrieval of Web resources.

3.1 Design Goals

Our primary goal is to expand the possibilities for formulating queries by integrating VisGets into interactive online Web search. Use of visualization in formulating queries may ease the specification of some concepts that are difficult to express in text. For example, time range is difficult to specify in text without using several words: such as "during February or March last year" or "last summer", and might not be particularly effective as a text query in a search engine. However, time as a concept can be extremely useful for humans as a query filter. For this discussion, a *query* is a request for information from a large online information space. A query is composed of a set of *parameters*. A parameter is a piece of information within a query that clarifies an information need. This parameter can be thought of as a constraint by which the entire information space is filtered. It can also be thought of as an attractor based on similarity factors. Commonly, in Web searches, these parameters are words, and complex queries are built up through Boolean combinations of words.

Query parameter *dimensions* are an important feature in VisGets. A query parameter dimension is a concept that is prevalent enough in Web information to be effective as a filter and is based on what type of information is commonly accessible on the Web, and what types of information people commonly use to characterize items of interest. For instance, photos are often associated with time, location, and tags. Therefore, one should be able to retrieve photos, or other similarly organized Web-based information, along these dimensions. In making a query based on a dimension, one manipulates a *range* rather than using a word or set of words. Based on the related work and these considerations the specific design goals for VisGets are:

Enable casual formulation of complex queries. A VisGet should support casual exploration of large information spaces using sophisticated queries. Complex queries can be constructed in conjunctive form (e.g., Boolean AND) using a combination of multiple query tools. VisGets should allow the formulation of search queries based on parameters that are difficult to specify with textual queries.

Summarize information collections visually. The display should include a visual overview of the parameter dimension within the VisGet, as well as a clear indication of the currently selected and filtered items. The interface should allow interactive exploration and review of the interrelations between multiple query dimensions.

Visualize bounds of query dimensions. The information seeker should be able to visualize the currently selected range for each query dimension, as well as the full range available for each query dimension. It should be easy to adjust the selected range, and to switch between a selected range and the full view.

Visualize query changes. As query parameters are modified in a VisGet, the effect of these changes should be displayed within the VisGet itself, and simultaneously reflected in all the other VisGets. The set of information items in the results list may be updated incrementally, using transition animations to help the information seeker to understand the changes to the information set.

Use integrated dynamic manipulation. A VisGet should provide interaction methods for adjusting query dimensions, and should provide responsive updates for the results of query adjustments. The interface should be inviting and intuitive, supporting interactive exploration of the relationships between multiple query dimensions.

Provide information drill-down. The interface should provide access to appropriate resolution in a VisGet's parameter dimension. The information seeker should be able to display detailed information for result items upon request (e.g., on-demand detail, or detail in context).

3.2 Choosing Information Dimensions for VisGets

Information resources published on the Web are increasingly often organized along three major dimensions—time, location, and tags—corresponding to when, where, and what [24].

Time. Practically everything on the Web has some kind of publication date. Examples include blog entries, recent updates on Wikis, and news feeds from friends on social networking sites. These mechanisms keep the information seeker up-to-date on current developments, either globally or in a closer social context. Being able to explore past conversations, or to see what is being discussed right now, makes time an useful dimension for Web-based exploration.

Location. With built-in GPS (Global Positioning System) capabilities now in many devices, more and more information published on the Web has geospatial information. Photos, news items, and encyclopaedia entries increasingly include geographic longitude and latitude, making location-based information exploration possible. Even information without explicit location information can be enhanced using natural-language processing, mashups, and geographic look-up services [38].

Tags. The topics associated with Web-resources are often made explicit through free-form keywords or tags. Tags can be added by the creators of Web pages intentionally to help people who are interested in specific topics to find Web resources such as photos and bookmarks. Tags are also often developed communally. As tag clouds become commonplace, they invite information seekers to explore and discover semantic and thematic relationships between resources on the Web.

While there are many other dimensions that can be used to categorize and structure information on the Web, for our initial VisGets we select these three dimensions as examples of searchable characterizations of Web resources, that are both widely used and are also often based on information that is fairly easy to extract.

3.3 Appearance and Functioning of VisGets

A VisGet is an information visualization widget that combines visual representation with interaction control with Web-query capabilities. A VisGet is composed of:

- a visualization of the Web-searchable information upon which it is based (in our examples time, location and tags),
- dynamic interactions that adjust and refine the range of the selected information in the VisGets' visualizations,
- search response that provides a list of information results, and
- coordinated interaction and adjustment with other VisGets.

For the following examples, information items were extracted during interaction from online RSS feeds from the Global Voices project [12], an editorial aggregation blog about blogs from around the world. Obtaining online data from RSS feeds is discussed in more detail later.

Time VisGet. For the temporal dimension, we used a simple bar chart (see Figure 1). The length of the base of the bar chart indicates the start and end of the time range of the searchable information. In Figure 1 this is from April 2006 to March 2008. There is a bar for each month in this range and the height of the bar indicates how many information items have been published in that month. Either range of months or a single month can be selected. The temporal selection can be changed by dragging the orange sliders along the horizontal axis. An individual month can be selected by narrowing the sliders, or by clicking on the month bar itself. When a single month is selected, an expanded bar chart for the days of the month is shown, allowing range filtering and selection at a finer granularity. At any point during interaction, any selection made through a VisGet can be reset to defaults by a single click.

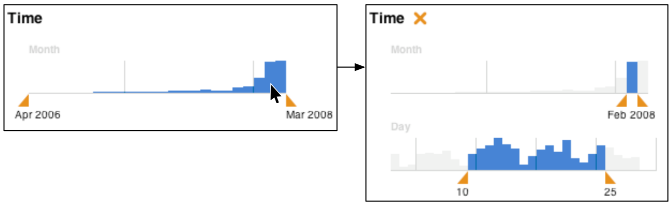


Fig. 1. Time VisGet: temporal bar charts represent the amount of information items published this month or day. The temporal range can be changed by selecting individual bars or dragging interactive sliders.

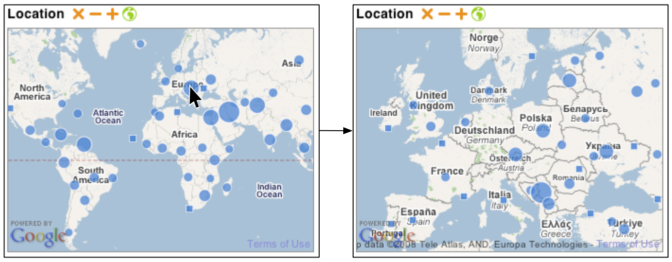


Fig. 2. Location VisGet: the size of a circle represents how many items refer to a region. Interacting with the map and the circles changes the spatial query parameter which is defined by the map's bounds.

Location VisGet. The spatial VisGet is based on a map of the world, upon which superimposed squares and circles represent information items (see Figure 2). Squares are used to mark the location of individual items, while the circles are used to represent aggregated items. The size of the circle reflects the number of items in the aggregation it represents. The amount of the map that is viewable in the window, or the map's display boundaries, serve as the spatial query parameter. That is, the results list will contain the information items associated with the markers (circles and squares) that are spatially within the area currently shown in the VisGet. Thus the spatial query can be changed and filtered by zooming and panning the map. Zooming in and out can be done either via the scroll-wheel, by double-clicking the left or the right mouse button, or by using '+' and '-' buttons at the top of the VisGet. Furthermore, it is possible to select individual circles or squares, to expand the aggregated information represented by the circles, and zoom into the map to show a more detailed spatial distribution of the selected information items. In the current implementation, based on Google Maps API, the location VisGet uses a Mercator projection that distorts the map increasingly towards the poles. A promising direction for future research would be to improve the map projection and the methods for location-based aggregation and filtering [1, 3, 8, 10, 24, 30, 41].

Tag VisGet. The tag VisGet features an alphabetically sorted tag cloud that provides a topical overview of the information collection. The font size of each tag represents how often it appears among the information items. The overall bounds of the topic dimension range are based on the tags in use within the whole information collection. When a tag is selected, it is coloured orange and becomes a tag filter, limiting the information in the results list to those that include this tag (see Figure 3). Multiple tags can be selected as filters concurrently.

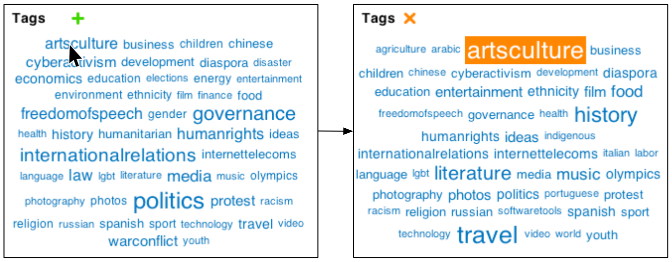


Fig. 3. Tag VisGet: the size of a tag (word) in the tag cloud represents the relative frequency of tags associated with information items. Selecting a tag activates it as a filter.

Overcrowding of the tag cloud is avoided by limiting the amount of tags displayed to those most frequently used. More seldom used tags can be shown either by selecting the '+' button on the top of the tag VisGet, or alternately, by narrowing down the result set by either selecting a tag as a filter or refining the query setting the other VisGets' (Time and Location) parameters.

Taiwan: Religious buildings–a way to explore how cultures mix

The popular image of multi-culturalism as a mosaic, a salad bowl in which different cultures mix but keep their integrity, is misleading. Cultures are more like soups, flavored with many ingredients, some identifiable. –From the book 'Cultures and societies in a changing world,' written by Wendy Griswold. Taiwan has a long history of diasporas and colonization. Even the so-called 'aboriginal people' in Taiwan drifted there several thousands years ago. About four hundre…

Mar 26, 2008 by I-fan Lin on Global Voices Online

Fig. 4. Hovering over a result displays detailed information as an overlay.

Results. The result list depicts the information items that comply to all query parameters. When queries change, information items are removed or added through animated transitions. Each result item displays its title and constitutes a hyperlink leading to the actual online information source. It is possible to hover over a result item and view a preview of the description (see Figure 4). If the information item includes an image, it is included in the upper right corner of the detail overlay. The order of the result list is currently based on the publication date, however, supporting more sophisticated ranking mechanisms based on relevancy or popularity would be an interesting future direction. While result lists implicitly convey linearity in data ranges, a spatial layout of result items could be used for visualizing other relationships among information items [33].

3.4 Coordinated Interactivity

Individual VisGets allow interaction within one dimension and several linked VisGets provide multidimensional exploration through coordinated interactions: weighted brushing and query refinements.

3.4.1 Weighted Brushing

Hovering with the mouse-pointer over a visual element temporarily highlights related visual elements in all linked VisGets and the results list. This highlighting disappears immediately when the mouse-pointer moves beyond the edge of the visual element. For example, hovering over a tag in the tag VisGet highlights all related elements in the temporal bar chart and the geographic map (see Fig. 5). In our system the highlighting is done by changing the colour of visual elements from blue to pink and dimming unrelated items.

The degree of relatedness between visual elements in multiple VisGets usually differs, since each element can represent different quantities of collection items. Instead of having a binary type of linking and brushing, where weakly related elements are highlighted as much as strongly related elements, weighted brushing represents varying degrees of relatedness. The highlighting of linked visual elements is based on how much association there is between information items and the currently brushed element. For example, the currently brushed or activated element A in Figure 6 is highly related to the visual element B, as they both represent the same information items, possibly in different VisGets. The visual element C is weakly related to A, as it shares only one associated information item with A. The visual element D has no relation with A, and is therefore displayed with the default color.

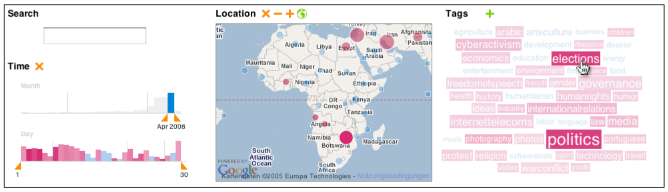


Fig. 5. Weighted brushing across three linked VisGets: hovering over 'elections' in the tag VisGet highlights visual elements that are related in all VisGets. While related items are displayed in pink, the degree of relatedness is visualized by their opacity.

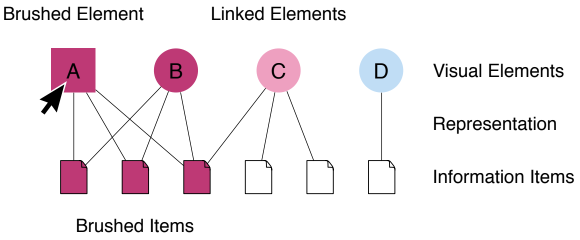


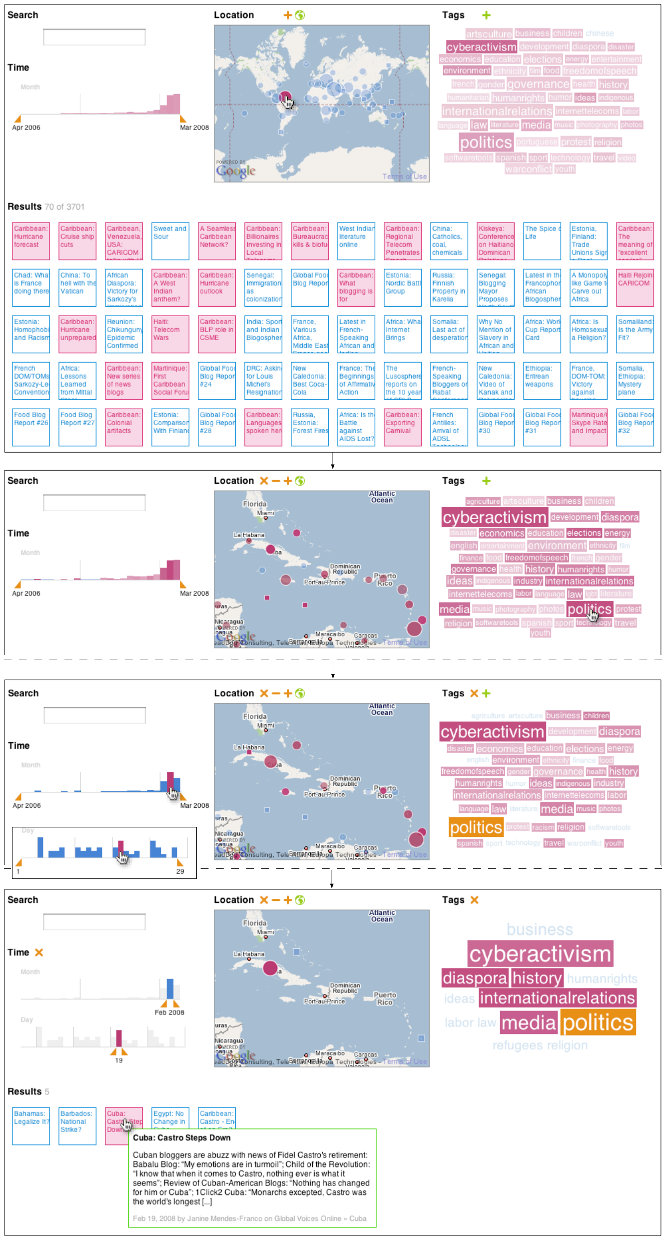
Fig. 6. Color and opacity of visual elements during weighted brushing depend on the intersection of information items represented by the brushed element (A) and the linked elements (B, C, D).

Information items can assume either one value or several values within a particular dimension. For example, an information resource usually has one publication date, while it can have multiple topics or tags associated with it. This has implications for the design of linking and brushing behavior within an activated VisGet. If the information seeker, for example, hovers over a tag, other tags in the same VisGet might be correlated and thus be highlighted. This provides the information seeker with visual cues indicating relations within the same dimension, in addition to highlights in other VisGets.

3.4.2 Query Refinements

An individual VisGet is used to constrain the query bounds of a particular information dimension through visual elements and controls that can be directly manipulated using the mouse pointer. Every refinement of a query triggers changes in the result list that are also reflected in all linked VisGets through animated transitions. Query parameters can be set in multiple dimensions, providing a technique for combining multidimensional parameters into a logical conjunction. Here are two examples of query refinement with VisGets.

Example 1 Consider an interest regarding politics on the Caribbean Islands around the middle of February 2008. Figure 7 depicts possible successive interactions with VisGets to follow this vague information need. A first step could be the selection of the circle close to the Caribbean on the geographic map, which highlights, in pink, the information items in the results list that are associated with the location of interest. In response to this action, the location VisGet zooms in to show the region of the Caribbean in more detail and the other VisGets update accordingly. Then one of the larger tags 'politics' could be selected from among the tags in the tag VisGet. Next, the temporal VisGet could be used to drill-down to the month of February. Brushing some of the days around the middle of the month of February highlights several items about Fidel Castro's retirement. Hovering over one of the result items shows the detail overlay of the news item “Castro Steps Down”.



Cuba: Castro Steps Down

Cuban bloggers are abuzz with news of Fidel Castro's retirement: Babalu Blog: “My emotions are in turmoil”; Child of the Revolution: “I know that when it comes to Castro, nothing ever is what it seems”; Review of Cuban-American Blogs: “Nothing has changed for him or Cuba”; 1Click2 Cuba: “Monarchs excepted, Castro was the world's longest [...]

Feb 19, 2008 by Janine Mendes-Franco on Global Voices Online » Cuba

Fig. 7. Example 1: Gradual refinement of query parameters along spatial, topical, and temporal dimension using the corresponding VisGets.

Example 2 It is also possible to begin with a conventional text-based search query, and then refine it using the VisGets. One could, for example, be interested in recent environmental topics mentioning storms (see Figure 8). At first ‘storm’ could be entered into the textual query box, after which all VisGets are updated visualizing only posts that contain the word ‘storm’. For example, the tag VisGet would show that posts containing ‘storms’ are predominantly about the environment. Selecting this tag would disambiguate the query from posts that do not use the word ‘storm’ in connection with the environment. After that, the query could be refined further along the temporal dimension with the time VisGet to select only the last three months. Few items remain and the location VisGet indicates that a storm may have recently caused discussion about a place southeast of Africa. Selecting the circle on the map narrows down the results to two items, both indicating a storm that hit Madagascar. Hovering over one of the items shows a graphic displaying the severity and extent of the storm. Note that during the flow of refinement the results list becomes shorter and more refined.

4 IMPLEMENTATION

We implemented a Web-based system that is aimed to support visual exploration of large information collections utilizing VisGets. The prototype performs online queries browsing approximately 3,000 information items in continually updated RSS feeds.

4.1 Architecture

The architecture of the visual information exploration system is divided into two parts: client-side and server-side (see Figure 9).

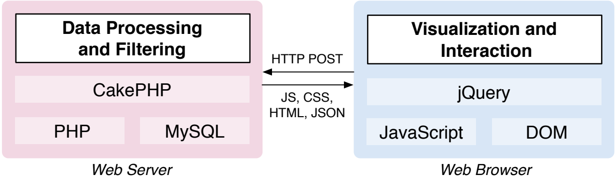


Fig. 9. Web-based architecture of the implemented system.

On the Web server, data is processed and filtered. We used CakePHP as a PHP programming framework, and MySQL as the relational database for storing processed data from RSS feeds.

Presentation and interaction are realized in the Web browser. We employ jQuery as a JavaScript framework that abstracts away the differences between multiple browsers. We chose jQuery because it enables both sophisticated query operations on the DOM (Document Object Model), as well as asynchronous communication with the server via the XMLHttpRequest object.

While the basic structure of the interface is transmitted as HTML, CSS, and JavaScript files, the actual processed and filtered data for the VisGets and result list is retrieved by the browser as JSON (JavaScript Object Notation), generated based on the current query parameters.

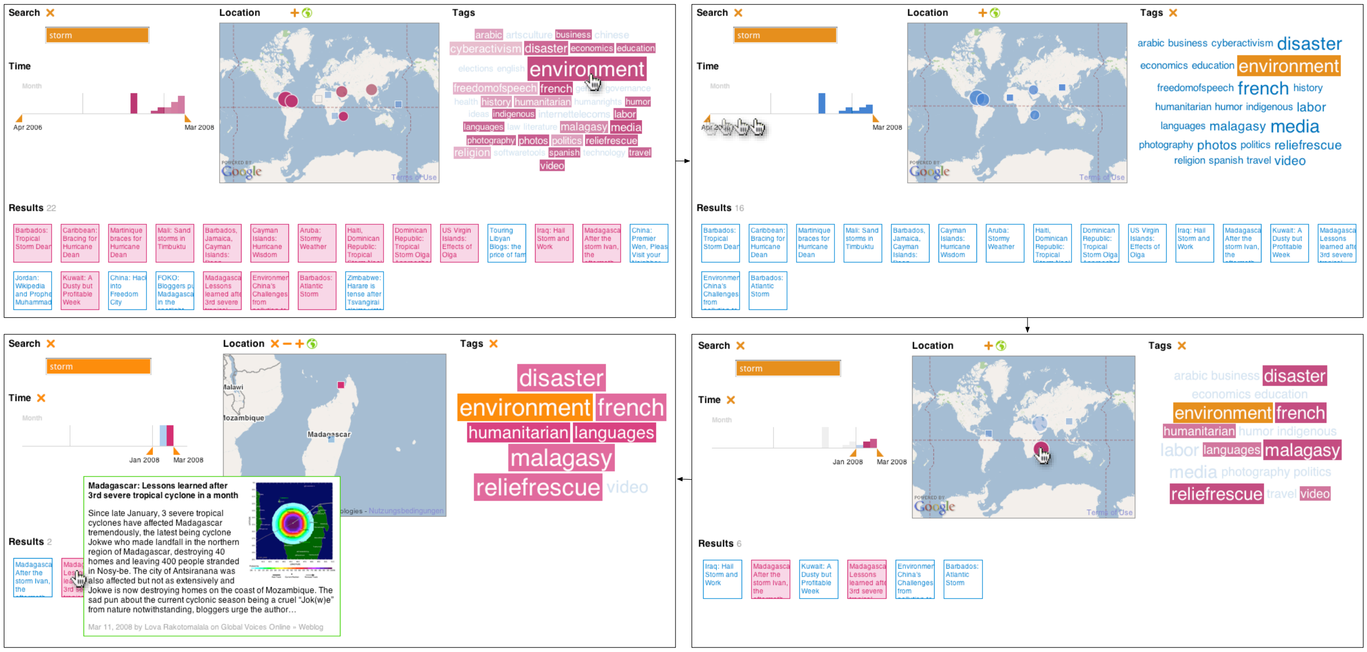
4.2 Data Extraction

Our prototype works with online, constantly updated data. Similarly to other search engines, the information items are extracted and stored in an online database, which is constantly updated from RSS feeds. While in principle VisGets could explore full Web information spaces, in practice further research is required in efficient Web-based data processing. In this direction, our delta queries (see Section 4.4) are a step towards minimizing data movement. The system assumes information items having title and description (text), date and time of publication (timestamp), geographic location (longitude and latitude), and tags (freeform keywords). If these characteristics are neither explicitly associated with an information item or cannot be derived from it, the respective entry is not displayed in the interface.

The extraction of title, description, and date information from RSS feeds is straightforward, and is done in our system using the MagpieRSS library. However, extracting location and tag information is more difficult, since this information is not always presented in a consistent format.

Tags can be represented in RSS feeds in different ways. The native approach in RSS is to put tags into `<category>` elements that are sub-elements of the `<item>` element. Tags can also be included as part of an extension to RSS, such as MediaRSS that is used by the photo sharing site Flickr. Furthermore, blog posts are often tagged using links to blog aggregation sites, such as Technorati, with the link’s `rel` attribute set to “tag”. In RSS feeds, these tag links appear in the description, and have to be parsed; in our system, we do this using regular expressions.

While the standard specifications for geospatial information embedded in RSS are unambiguous, the number of RSS feeds that include latitude and longitude on a per-item basis is limited. However, often the title and description includes geographic indications, such as city



Madagascar: Lessons learned after 3rd severe tropical cyclone in a month

Since late January, 3 severe tropical cyclones have affected Madagascar tremendously, the latest being cyclone Jokwe who made landfall in the northern region of Madagascar, destroying 40 homes and leaving 400 people stranded in Nosy-be. The city of Antsiranana was also affected but not as extensively and Jokwe is now destroying homes on the coast of Mozambique. The sad pun about the current cyclonic season being a cruel "Jok(w)e" from nature notwithstanding, bloggers urge the author…

Mar 11, 2008 by Lova Rakotomalala on Global Voices Online » Weblog

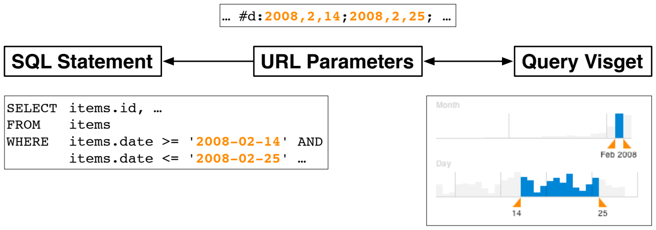
Fig. 8. Example 2: In combination with a conventional search query VisGets allow further refinement along conceptual dimensions.

or country names. To add geographic information to the RSS feeds, we used the GeoNames Web service, which (among many other gazetteer functions) can enhance regular RSS feeds into GeoRSS feeds [38]. The Web service takes the URL of a non-geospatial RSS feed as an input, and returns the same feed with (longitude, latitude) pairs added to items, if sufficient geographic indications are present.

The descriptions included in RSS feeds can vary greatly in terms of length and formatting. To provide a consistent display of information in the interface, we exclude any markup other than formatting with italics and bold. The length of the descriptions are also limited. If one or more images have been added to an item, the source URL of the first image is parsed using a regular expression, and later included in the interface.

4.3 Query Parameter Conversion

Query parameters limit the currently selected set of information items. In the implemented system there are four types of query parameters: dates, location, tags, and search terms of a conventional text field. We will neglect the latter, and focus in the following on parameters that are set and changed by VisGets.



```
… #d:2008,2,14;2008,2,25; …
```

```
SELECT  items.id, …
FROM    items
WHERE   items.date >= '2008-02-14' AND
        items.date <= '2008-02-25' …
```

Fig. 10. Query parameter conversion, here with temporal dimension.

In a visual information exploration system (see Figure 10), query parameters are constantly transformed between three forms: visual query parameters (as part of VisGets), parameters encoded for URLs and HTTP requests, and query parameters (as part of SQL query statements). The challenge is to represent query parameters in these different forms, and apply transformations between them without losing any accuracy or semantics along the way.

Encoding query parameters for URLs or HTTP requests is essential in a distributed environment, where the information exploration system has some logic running on the server and some within the Web browser. In the implemented prototype, every query state is encoded as a URL. This allows the information seeker to bookmark the current parameters, and send these to friends or colleagues. Temporal, spatial, and topical parameters therefore have to be encoded into a form that both the server and client can interpret and generate. Because JavaScript code running in the Web browser can only change the local part of the address, query parameters in the implemented system are put behind the hash sign ('#'). This allows the client-side code to change the query parameters in the URL, once the information seeker has modified them through a VisGet.

Visual query parameters in VisGets can be slider positions, map bounds, and tag selections. The selections of one or multiple tags as filters or query parameters is set using the class parameter of the respective `<span>` element that encompasses the tag. The geographic bounds of a map can be retrieved and set using the functions provided in the Google Maps API. The temporal query parameters are represented by the position of triangle-shaped `<div>` elements, which are used as sliders.

The query parameters are used to filter the information items stored in the database. Therefore the URL-encoded parameters are converted into `WHERE` constraints in SQL statements.

4.4 Delta Queries

Responsiveness is a vital feature in interactive information visualization. In the Web context, this requirement places practical constraints on the number of transactions between client and server, and the data volumes exchanged, because of round-trip latency and network bandwidth consumption. As a result, one of our key design considerations was avoiding unnecessary data movement, especially for redundant data.

To provide fast and comprehensible updates, we developed *delta queries*. After a visual query parameter has been modified, the corresponding VisGet determines which items are to be removed, which items can remain, and what kind of overlap exists between the current selection s_i and the new selection s_{i+1} of information items. The resulting change in the information set is referred to as the delta.

We have identified four types of overlap, which are depicted with sliders and Euler diagrams in Figure 11. By determining the overlap (if any), it is possible to reduce the number of items requested from the Web server, reducing network latency and thus the time to update the interface. Furthermore, it is possible to provide animated transitions to indicate the removal or addition of information items in the result list, as well as the changes happening within the VisGets.

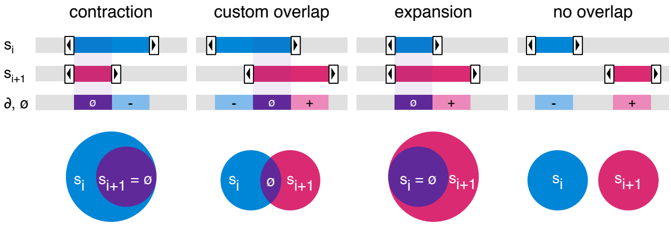


Fig. 11. Types of overlap; s: selection, δ: difference, ø: overlap.

Implementing delta queries requires careful tracking of information items. Once the VisGet determines the ids of the information items that have to be removed, these items are removed from all VisGets and the result list. Depending on the VisGet, animated transitions may be used to convey the interface update. The ids of the retained information items are sent within the HTTP request to the Web server, indicating that these items can be excluded from the returned JSON data. In the case of a narrowed selection (contraction), the query to the Web server can be omitted altogether.

5 VISGETS IN USE

As part of our design process, in order to direct our development of VisGets, we conducted an exploratory observation of VisGets in use. We observed 10 people (4 female, 6 male; age range 19 to 37) who self-estimated their Internet experience as between 5 and 15 years and their Internet usage as between 2 and 8 hours per day. On a 20 inch desktop computer screen, VisGets in a Web browser window accessed online data that consisted of approximately 3000 articles from several RSS feeds from the Global Voices site [12].

Participants were first given a tutorial explaining how the system worked and were provided with free exploration time to become familiar with the VisGets. Each participant was asked to perform two sets of tasks. The first set of tasks consisted of nine focused questions about current events and had specific countries or regions, topics, or dates as answers. Each question of a given dimension had clues from the other two dimensions. For example, one question was: “Which region declared its independence in February 2008?” Here the answer was spatial and the clues were topical and temporal. When the answer was topical the clues were spatial and temporal, and for temporal answers the clues were spatial and topical. The second set of tasks were more open-ended. Participants were asked to imagine that they were a newspaper journalist, a health inspector from the WHO, or a human rights investigator with Amnesty International. The tasks were to plan their travel to uncover interesting news stories, track the global health situation, or monitor human rights issues. After tasks were completed, participants were asked to fill out a post-session questionnaire and a semi-structured interview was conducted.

5.1 Discussion

The post-session questionnaire contained five-level Likert scale questions that asked participants’ opinions on the interface and the discovery of information, gaining of overview and access to detail information. The results are shown in Figure 12.

All VisGets were rated favourably and similarly by participants. Through the observations and interview the temporal VisGet was noted as particularly helpful for discovering information. During the study we noticed that participants used the location VisGet less for visual filtering, than either the temporal or the topical VisGets. Spatial cues for countries were used by some participants as textual queries. In the interview, five out of ten participants chose the temporal VisGet as their preference for filtering through the information items, and four participants described the location VisGet as unnecessary. When asked where such a visual search would be helpful four participants mentioned news and current events, three noted possible value for academic research, and three saw it as applicable to conventional search.

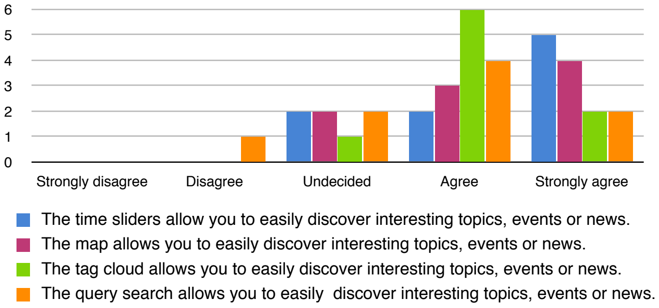


Fig. 12. Participant opinions about usefulness of each VisGet in discovering information.

It is possible that the degree to which participants favoured the temporal VisGet may be, at least in part, due to the relative difficulty of expressing a time range in a few of words of text that would be suitable for a search query. In contrast, even though map bounds or ranges are also difficult to express as a text query, spatial concepts like country and city names are sufficiently functional as textual queries. Similarly tags themselves can be readily used as queries.

6 CONCLUSION

In this work we have introduced VisGets, coordinated information visualization query widgets that aim to make it possible for information seekers to orient themselves within Web-based information spaces and to incrementally build complex filtering queries. Our general design considerations led us to define VisGets as widgets that:

- Provide a visualization of an information dimension that can be used to specify an information query.
- Allow query specification through dynamic manipulation of the range of these information dimensions.
- Propagate range adjustments to other associated VisGets, providing immediate visual evidence of query refinement.
- Offer responsive results lists that immediately and incrementally change according to the manipulation of the VisGets.
- Reflect actions in a given VisGets in all VisGets through highlights of current choices and parameter range selections.

We have presented temporal, spatial, and topical VisGets, within a fluid and responsive Web-based implementation. An exploratory study indicates positive reactions towards visual information exploration with VisGets.

The main contribution of this work is the application of dynamic queries to Web-based information seeking using multiple coordinated VisGets. We have developed weighted brushing as a novel interaction technique that displays different degrees of relatedness between visual elements in coordinated views, and delta queries as a novel mechanism to convey query changes and reduce the transmission load during query refinements.

In the current realization, our VisGets are limited to fairly simple visual representations and query parameters. This is partly due to the constrained capabilities of current Web browsers and also a result of rather simple data extraction. For example, in the current implementation spatial constraints are defined by the map’s rectangular bounds. A more powerful VisGet could allow advanced spatial queries, described with arbitrary, potentially disconnected polygons or by geopolitical boundaries such as continents and countries. Developing more sophisticated data modelling, visualization, and interaction design to support more types of selections, different granularities, and data uncertainties is an exciting future research direction.

Furthermore, the amount of information items extracted from RSS feeds is limited by the capabilities of the database and the processing power of a single Web server. Providing access to a number of information items comparable to the Web’s proportions would require a substantively more powerful computer system made up of many individual systems crawling, indexing, and analysing the Web.

Because VisGets include, extend, and combine existing visualization techniques, a large set of InfoVis techniques may be considered to be utilized for the exploration of different types information spaces. While the design and realization of VisGets can benefit from previous experience, visualizations have to be adapted to allow for visual query formulation. To learn more about the possible role of interactive visualizations in information seeking, we will consider further information spaces and formats beyond RSS as we explore new types of VisGets.

ACKNOWLEDGEMENTS

Many thanks to Marcel Götze, Mark Hancock, Uta Hinrichs, Petra Isenberg, Matthew Tobiasz, and Amy Voida for valuable assistance and advice. We are also very grateful to the referees for their constructive feedback and suggestions. Thanks to Global Voices Online for publishing their content under a Creative Commons license. Funding for this research was provided by NSERC, iCore, and NECTAR.

REFERENCES

[1] S. Ahern, M. Naaman, R. Nair, and J. H.-I. Yang. World Explorer: Visualizing aggregate data from unstructured text in geo-referenced collections. In *JCDL '07: Proc. of the ACM/IEEE Joint Conf. on Digital Libraries*, pages 1–10. ACM Press, 2007.

[2] C. Ahlberg and B. Shneiderman. Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In *CHI '94: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 313–317. ACM Press, 1994.

[3] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer, 2006.

[4] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: A nucleus for a web of open data. In *Proc. of the Int. Semantic Web Conf.*, 2007.

[5] M. Q. W. Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *AVI '00: Proc. of the Working Conf. on Advanced Visual Interfaces*, pages 110–119. ACM Press, 2000.

[6] M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.

[7] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):28–37, 2001.

[8] G. Cai. Geovibe: A visual interface for geographical information in digital libraries. *IEEE/ACM Joint Conf. on Digital Libraries (JCDL), Workshop on Visual Interfaces to Digital Libraries.*, 2001.

[9] E. H. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions and the Web. In *CHI '01: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 490–497. ACM Press, 2001.

[10] M. Christel. Accessing news video libraries through dynamic information extraction, summarization, and visualization. *Visual Interfaces to Digital Libraries*, pages 98–115, 2002.

[11] J. English, M. Hearst, R. Sinha, K. Swearingen, and K.-P. Yee. Flexible search and navigation using faceted metadata. In *SIGIR '02: Proc. of the ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 11–15, August 2002.

[12] Global Voices Online. `http://www.globalvoicesonline.org/` (Retrieved 2008-05-14).

[13] E. Golub and B. Shneiderman. Dynamic query visualisations on world wide web clients: A DHTML solution for maps and scattergrams. *International Journal of Web Engineering and Technology*, 1(1):63–78, 2003.

[14] J. Harris and S. Kamvar. We Feel Fine. `http://www.wefeelfine.org/` (Retrieved 2008-05-14), 2006.

[15] M. A. Hearst. Tilebars: Visualization of term distribution information in full text information access. In *CHI '95: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 59–66. ACM Press/Addison-Wesley, 1995.

[16] H. Hofmann and M. Theus. Selection sequences in MANET. *Computational Statistics*, 13(1):77–88, 1998.

[17] D. F. Huynh, D. R. Karger, and R. C. Miller. Exhibit: Lightweight structured data publishing. In *WWW '07: Proc. of the Int. World Wide Web Conf.*, pages 737–746. ACM Press, 2007.

[18] S. Jones. VQuery: A graphical user interface for boolean query specification and dynamic result preview. *UIST '99: Proc. of the Symposium on User Interface Software and Technology*, pages 143–151, November 1999.

[19] K.-S. Kim. Effects of emotion control and task on web searching behavior. *Information Processing & Management*, 44(1):373–385, January 2008.

[20] P. Klein, F. Muller, H. Reiterer, and M. Eibl. Visual information retrieval with the SuperTable + scatterplot. In *Proc. of the 6th Int. Conf. on Information Visualisation*, pages 70–75, 2002.

[21] B. Y.-L. Kuo, T. Hentrich, B. M. Good, and M. D. Wilkinson. Tag clouds for summarizing web search results. In *WWW '07: Proc. of the Int. World Wide Web Conf.*, pages 1203–1204. ACM Press, 2007.

[22] G. Marchionini. Exploratory search: From finding to understanding. *Comm. of the ACM*, 49(4):41–46, 2006.

[23] L. T. Nowell, R. K. France, D. Hix, L. S. Heath, and E. A. Fox. Visualizing search results: Some alternatives to query-document similarity. In *SIGIR '96: Proc of the Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, pages 67–75. ACM Press, 1996.

[24] D. J. Peuquet. *Representations of Space and Time*. Guilford Press, 2002.

[25] C. Plaisant, B. Shneiderman, K. Doan, and T. Bruns. Interface and data architecture for query preview in networked information systems. *ACM Trans. on Information Systems*, 17(3):320–341, 1999.

[26] A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen. Getting our head in the clouds: Toward evaluation studies of tagclouds. In *CHI '07: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pages 995–998. ACM Press, 2007.

[27] R. M. Rohrer and E. Swing. Web-based information visualization. *IEEE Computer Graphics and Applications*, 17(4):52–59, 1997.

[28] A. Scharl and K. Tochtermann, editors. *The Geospatial Web – How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. Springer, 2007.

[29] B. Shneiderman. Dynamic queries for visual information seeking. *IEEE Software*, 11(6):70–77, 1994.

[30] A. Slingsby, J. Dykes, J. Wood, and K. Clarke. Interactive tag maps and tag clouds for the multiscale exploration of large spatio-temporal datasets. In *IV '07: Proc. of the Int. Conf. on Information Visualization*, pages 497–504, 2007.

[31] R. Spence and L. Tweedie. The attribute explorer: Information synthesis via exploration. *Interacting with Computers*, 11(2):137–146, 1998.

[32] A. Spoerri. Infocrystal: A visual tool for information retrieval. In *VIS '93: Proc. of the Conf. on Visualization*, pages 150–157, 1993.

[33] A. Spoerri. Visual mashup of text and media search results. In *IV '07: Proc. of the Int. Conf. on Information Visualization*, pages 216–221. IEEE Computer Society, 2007.

[34] M. Theus. Interactive data visualization using mondrian. *Journal of Statistical Software*, 7(11):2002, 2002.

[35] F. Viégas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Many eyes: A site for visualization at internet scale. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1121–1128, Nov/Dec 2007.

[36] M. Wattenberg. Baby names, visualization, and social data analysis. In *INFOVIS '05: IEEE Symposium on Information Visualization*, pages 1–7. IEEE Computer Society, October 2005.

[37] R. W. White, B. Kules, S. M. Drucker, and m. schraefel. Supporting exploratory search – introduction. *Comm. of the ACM*, 49(4):36–39, 2006.

[38] M. Wick and T. Becker. Enhancing RSS feeds with extracted geospatial information for further processing and visualization. In A. Scharl and K. Tochtermann, editors, *The Geospatial Web – How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*, pages 105–116. Springer, 2007.

[39] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1129–1136, Nov-Dec 2007.

[40] J. Wood, K. Brodlie, and H. Wright. Visualization over the world wide web and its application to environmental data. In *VIS '96: Proc. of the 7th Conf. on Visualization*, page 81, Los Alamitos, CA, USA, 1996. IEEE Computer Society.

[41] J. Wood, J. Dykes, A. Slingsby, and K. Clarke. Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Trans. on Visualization and Computer Graphics*, 13(6):1176–1183, November 2007.