

Designing an Interactive Visualization for Exploring *Madison Commons*: **Final Report**

Joe Kohlmann & Mitch Schwartz

CS 638, Spring 2012 · 14 May 2012

Description We want to glean design principles from each of our Project 1 writeups to build an interactive visualization for displaying a set of news articles. We're particularly interested in using metadata to guide the content exploration process through this visualization. We'll be using the *Madison Commons* dataset, because the articles already have metadata tags based on topic and location. Our goal is to produce a compelling visualization for the article archive using web technologies.

Goals

- Determine the best visualization design for an article exploration.
- Integrate elements from Joe's metadata filters, custom views, sunburst and from Mitch's display network display.
- Consider any additional features beyond our work in Project 1.
- Explore implementation options for interactive visualization with web technologies.

Outcomes

- A visualization that facilitates exploration of the *Madison Commons* article archive.
- A project writeup that justifies our design decisions.

The Problem

Joe and Mitch tackled similar problems for their first projects. Both focused on providing interactive tools for exploring a set of news/magazine articles. Joe built from elastic lists. Mitch built from network displays. But both design concepts shared foundations in metadata and interactivity.

Joe was unsatisfied with elastic lists, so he designed a more interactive visualization that explicitly incorporated a system of metadata filters. His design allowed users to manipulate the visualization based on the most relevant metadata. Mitch's design used metadata tags as the basis for creating a network display. Each article served as a node and links were determined by shared tags. Mitch also incorporated semantic interaction (Endert, Fiaux, and North, 2012) to allow users to manipulate the importance of different tags/links.

For this second project, Joe and Mitch returned to their shared problem: how best to visualize a set of articles in a way that uses metadata to guide recommendations and user exploration.

The (Meta)Data

Because metadata was central to both their designs, Joe and Mitch decided to use a set of *Madison Commons* articles because these articles have already been assigned content-related tags. When published, each *Madison Commons* article is tagged according to its topic area(s) and geographically-relevant location(s). There are 21 available topic tags (e.g., city, diversity, education, environment, food, local business, transportation, youth). And there are 134 available location tags. The location tags are structured in two tiers: region and neighborhood. There are nine regions (East, Far West, Isthmus, Near East, Near West, North, South Central, Southwest, West). Each region then contains various neighborhoods. The Isthmus region, for example, contains seven neighborhoods (e.g., Capitol, Marquette, Tenney-Lapham, UW Campus).

None of the tags are mutually exclusive, and any article may be tagged with multiple topic tags and/or multiple location tags. Other metadata associated with each article include author, date of publication, and total number of views. But the primary focus for this implementation is the topic and location tags. While other metadata may be relevant, we feel both the underlying goals of *Madison Commons* and user interest revolve around the topic and location tags more than other metadata.

Madison Commons relaunched in March 2011 with a new reporting structure. We used this relaunch as the starting point for our visualization dataset. Our dataset includes every article published to the site between March 23, 2011 and April 23, 2012—142 articles in total. *Madison Commons* is built with Drupal, so we used the Node Export and Taxonomy CSV import/export modules to acquire the relevant content for our dataset.

The Design Process

Joe liked the premise of Mitch's network display, so very early into the design process we settled on displaying articles as nodes. After this decision, the design considerations of greatest import were how to determine links and how to display those links.

The process of determining links built from the forms of metadata we knew were available. It seemed obvious to form connections among articles based on which metadata they share. But we were also keenly aware that a traditional node-link display would not make for an ideal visualization. Given the number of articles and the number of tags, we knew

the edges of node-link diagram would create an extraordinarily cluttered display that would be difficult for users to interpret.

The design problem, then, was how to display connections without using edges. Joe and Mitch each offered what would become important sources of inspiration. Thinking of the game *World of Goo*, Mitch suggested physics-based clustering. Articles that share tags would be attracted to one another, resulting in clusters of related articles. This would eliminate the need for edges, using proximity instead to encode connections. Physics-based clustering solved many of the scalability problems of a node-link display and, as such, became a core component of the final design.

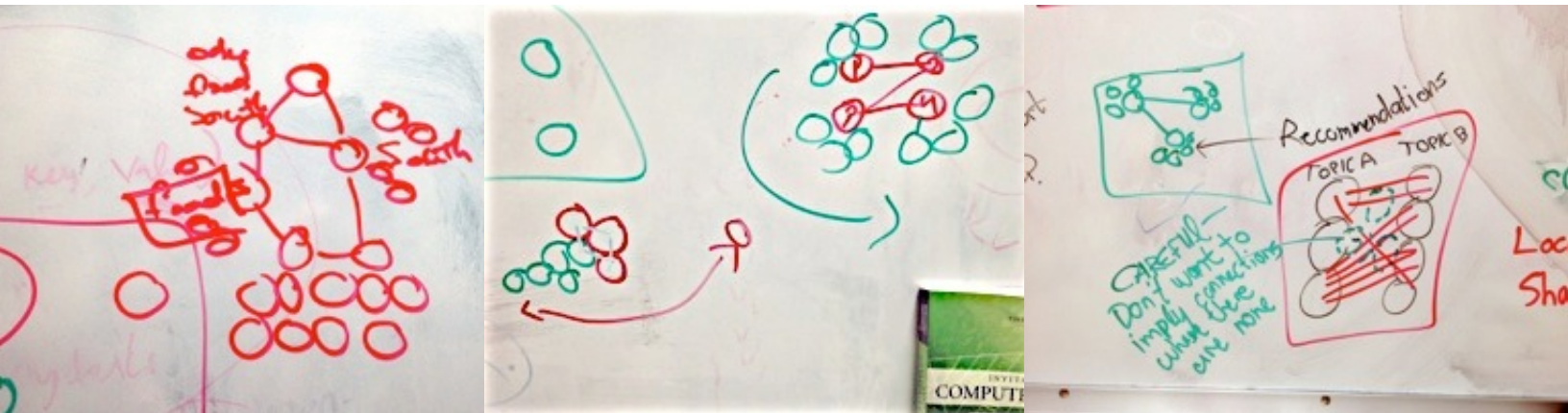


Figure 1

Whiteboard drawings depicting our early ideas for node groupings by topic and location. As seen in the right photo, we didn't want to overwhelm users.

How-

ever, physics-based clustering did not address metadata filtering. While clustering allows users to see which articles are related, clustering alone conveys nothing about *how* the articles are related. We still needed a design mechanism for encoding which forms of metadata a cluster shares. In this area, Joe's suggestion of a Voronoi-based design proved especially fruitful. Joe suggested that clusters could be colored in a manner similar to a Voronoi diagram. The result would be colored clusters/cells, with the color of the cluster/cell encoding relevant metadata information.

Although the Voronoi design did not apply perfectly to our goals, the suggestion provided two important design results:

We settled on color as one way to encode the relationships among articles. But the specifics of our color design required much further consideration. Our final design choices regarding color are explained in more detail in the following section.

We began to think about the use of a map to encode location tags. Because Voronoi diagrams resemble maps, the Voronoi design prompted us to think about using the pre-existing map of Madison regions/neighborhoods in order to encode location tags. In this way, the two different kinds of tags would be differentially encoded. Topic tags would be encoded using color. Location tags would be encoded using position, overlaid on a map of Madison.

The inspiration outlined above led us to the core components of our implementation: Articles would appear as nodes. Articles would be mapped to corresponding geographic locations, using (absolute) position to encode geographic tags. Further, within each map region, physics-based clustering would use proximity (relative position) to encode shared topic tags. Color would be used to further encode shared topic tags. Color would be a redundant encoding within a region, reinforcing the clustered articles. But color would also connect clusters across region, encoding connections among nodes even if they do not have close proximity.

The Implementation

Given Joe's affinity for web technologies, we early on decided to build a working prototype using JavaScript. This led us to consider and use two primary JavaScript libraries:

- **D3.js**: "Data Driven Documents" framework focusing on dataset to DOM element binding.
- **Polymaps**: Tiled interactive maps framework.

Both these libraries are very focused on SVG manipulation, and sharing a common author of course means they play well together. Overall the two are very robust libraries that reduced the number of novel challenges taken on. Additional tools such as CloudMade for map image tiles from OpenStreetMap made it easy to add a level of polish to the mapping component. SVG animation performance is slow on mobile devices as of this writing, but the final implementation performs well in a good browser on modern Intel hardware.

The consistent challenge for this project was to "massage" the source data into something workable. Using an offline dataset from *Madison Commons* eliminated a big dependency on a non-local data source. Once we acquired all *Madison Commons* articles as XML, Joe used extensive XSLT transformations to manipulate this and other XML files into the final format delivered to the web application. These transformations involved element cross-referencing, as the articles only included indices for location and topic tags, and content regrouping for optimal enumeration. For instance, the final file containing article data is hierarchically organized by location and contains sets of all tags used in a location.

The final project somewhat inadvertently uses both an XML and a JSON dataset, each for different reasons, though in the latter case it was to eliminate the overhead of parsing and converting XML to JSON in-browser. This experience suggests that future projects may solely rely on JSON, though XML and XSLT are still a powerful combination.

As the project required Joe to synthesizing his newfound knowledge of XSLT, XML, SVG, D3 and Polymaps, characteristics of the final web application are dictated by these tools. For instance, after considering possible layouts for articles within a location on the map, Joe used D3's high-performance force-directed graph implementation to create bundles of

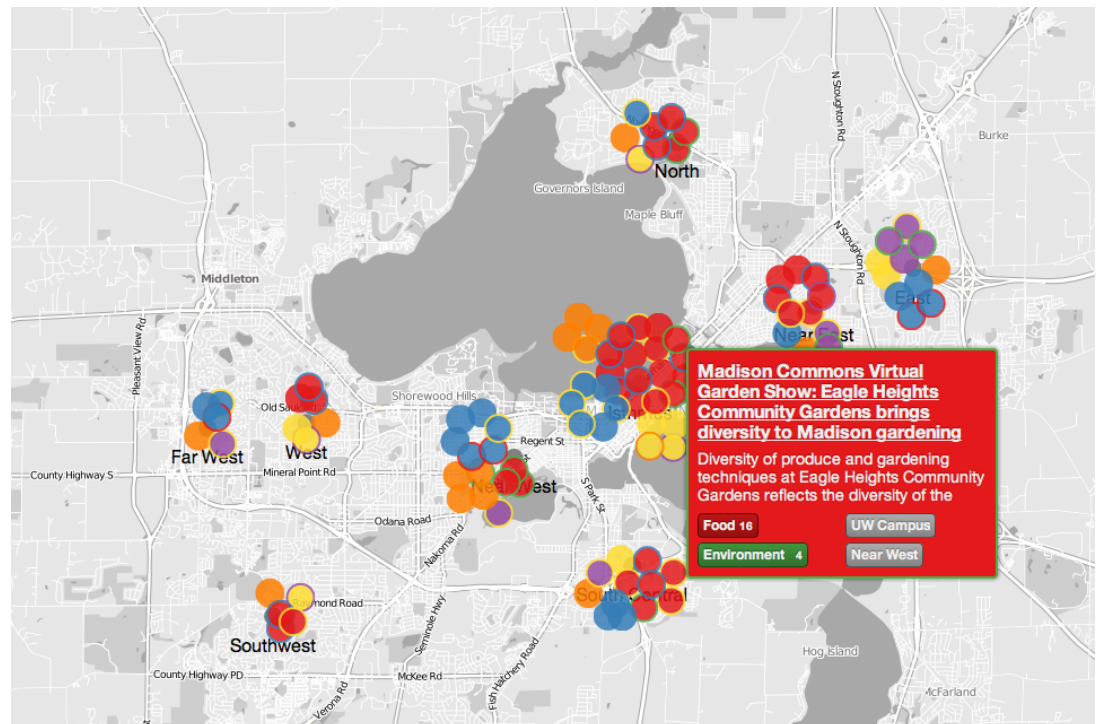
articles grouped by their most popular tag. Joe's sure this layout would have been much different (i.e. poorer) if it weren't for D3. That said, the combination of geographical position and topic affinities achieved in the final implementation comes very close to achieving the vision for the map component we originally discussed.

One "limitation" that worked out very well is the fill and stroke coloring scheme used in the final implementation. Joe and Mitch decided to use a six-color scheme to allow the analyst to quickly identify five of the "tent-pole" topics *Madison Commons* covers. We felt that six colors would be a number large enough to capture the core topic areas, but small enough for people to remember the color-topic encodings without requiring constant reference to a legend. (Of course, the tag colors are also reiterated in the popup displays.)

The six colors were selected using Colorbrewer 2.0. Specific color encodings are not meant to convey meaning, but were still given design consideration. Green was assigned to the environment tag due to the common association between the two. Orange was assigned to transportation due to Google Maps' use of orange for coloring streets. Yellow was assigned to the 'other', catch-all category due to its neutral nature. We wanted to draw attention most to the core topics rather than to the 'other' category. The city topic relates to city government, so we did not want to use red or blue for this tag, since these two colors have their own political associations. This left purple for the city tag. And blue generally has a negative association with food (people will eat less when plates/bowls are blue, for example), so red was assigned to food. As the remaining color, blue was assigned to education.

Figure 2

The final project implementation. A grayscale map of Madison contains about 100 colored circles, each representing an article. Articles are "anchored" to a specific geographical region ("North", "West", etc.) and receive a fill color based on the most popular tag in the region. Stroke colors are based on the second most popular tag in the region. Hovering over an article displays a popover with more info.



The Future

We regard this project as a compelling example of what's possible with attractive, geocentric presentations of content with rich metadata. The relatively simple elements of the design combine to form a fun, interactive browsing interface for *Madison Commons*. There's a mile-long list of what could be added to this core design next, so we present a few ideas:

- **Live data implementation**, presumably integrated with Drupal. This would allow the project to find a permanent home among *Madison Commons*'s growing content database.
- **Static physics implementation**. As experimented with in this project, there are several ways to do much of the data massaging and processing on a server rather than inside the analyst's web browser. D3's force-directed graph implementation is incredibly fun to play with, but calculating layout positions ahead of time would improve rendering performance in some environments.
- **Smarter article groupings**. What's there works nicely, but there are other metadata fields, such as author, publish date, even sub-location (i.e. locations within the top-level ones shown on the map) that could be put to better use. Some of these may or may not have a geographical component.
- **More direct manipulation tools**. At one point Joe and Mitch discussed a user interface that involved dragging articles into the center of the map to build a custom subset the analyst could experiment with. Examples of "experimentation" include changing the grouping criteria for the subset, allowing the analyst to discover new connections that geography or tagging had previously obscured.
- **Analyst personalization tools**. From the simple option to display the analyst's location on the map, to queries into the analyst's article reading history, to generation and visualization of reading recommendations, there are many more ways to personalize this display.

The challenge of working with real data has resulted in a good visualization design experience for both of us. Myriad different techniques, frameworks and design thought went into this design, and while we set very ambitious goals for something neither of us have done before, Joe and Mitch frankly think the final product is pretty darn awesome.

References

- Endert, A., Fiaux, P., & North, C. (2012). Semantic interaction for visual text analytics. Proceedings from CHI 2012: Conference on Human Factors in Computing Systems. Austin, TX. Retrieved from: http://people.cs.vt.edu/aendert/Alex_Endert/Research_files/Endert_CHI2012_Semantic_Interaction.pdf