

## Content clouds as exploratory qualitative data analysis

**Abstract:** This article proposes the use of content clouds as a method of exploratory qualitative data analysis. Content clouds are a type of visualization that summarizes the contents of a document by depicting the words that appear most often in larger, darker type within the cloud. When utilized as a form of qualitative GIS, content clouds provide a powerful way to summarize and compare information from different places on a single issue. Two examples using different types of documents and differently-scaled environmental issues demonstrate the possible utility of this method in summarizing qualitative data and suggesting avenues for research.

**Keywords:** methodology, visualization, qualitative GIS, content clouds

### **Introduction**

Data visualization is a growing area of both research and application across many different fields. Within geography, GIS is where the bulk of visualization work takes place, traditionally with quantitative but increasingly also with qualitative data. In the online realm, visualization has become a way to quickly display and summarize content, whether identifying a structure to a text such as a Presidential speech or providing an automatic visual index to the contents of a web log (blog) or photo storage site.

Readers of blogs, LiveJournal, and other interactive or Web 2.0 content are familiar with the concept of tag clouds. The author of each entry in a blog creates their own keywords, or tags, that classify the entry. A bar along the side of the web page automatically generates a "tag cloud", where words are sized according to the frequency with which

they appear. This provides an at-a-glance understanding of the most common topics covered at that particular website, at least according to the author's classification of his or her writing.

However, this method of data visualization does not have to be limited to self-classification, nor to online material. *Content clouds* or word clouds are generated in the same manner as tag clouds but use input from a text other than blog tags (Figure 1). Any material that can be studied using content analysis can therefore be visualized through the generation of a content cloud. For data with a spatial component, mapping content clouds of the same source material for different places is a quick and easy way of demonstrating and exploring differences across space.

[insert Figure 1 here]

This paper proposes mapping content clouds as a method of exploratory qualitative data analysis, combining elements of qualitative GIS, content analysis, and visualization. Two examples demonstrate the potential of this method in terms of ease of use and utility of results. The content clouds clearly demonstrate the difference that place makes to an environmental issue, both within a region and across regions, and suggest avenues for further research.

The paper begins with a brief review of three bodies of methodological literature: qualitative GIS, content analysis within geography, and the generation and use of tag clouds or content clouds. This is followed by two examples of content clouds in use. The first is an analysis of the transcripts of eight public meetings held in the Chicago area concerning the purchase of a beltline railroad by a major North American railroad. The content clouds reveal substantial differences in terms of what speakers discussed at

meetings in different locations, demonstrating spatial variation in the issues of concern to the public. The second example is a comparison of the content of newspaper articles on green buildings from different regions of the U.S. It might not be surprising that the understanding and discourse of green buildings differs from place to place, but the ways in which it is discussed in different places have important implications for how people understand green buildings and urban sustainability. The final section explains the advantages and shortcomings of using content clouds in exploratory qualitative data analysis.

## **Literature review**

### *Qualitative GIS*

As many excellent reviews already exist of qualitative GIS (e.g., Kwan 2002, Knigge and Cope 2006, Pavlovskaya 2006), I will only briefly highlight the most relevant work, specifically the potential role of GIS in visualizing qualitative information, rather than enabling strictly quantitative analysis. In particular, I am answering Pavlovskaya's call for new ways to use GIS to develop social theory with regards to spatiality, including "visualizing nonquantifiable experiences, bringing spatiality in explanations through spatial representations, and using the ontological power of mapping" (p. 2015).

There have been a variety of different approaches to integrating qualitative data and GIS. As Elwood and Cope (2009) argue, "geographic phenomena, their relationships, and their meanings are produced and negotiated at many different moments in GIS development and application: in spatial data, in data structures, in spatial analysis techniques, in the meanings fostered or foreclosed in GIS-based maps and applications" (p. 2). For example, algorithms or techniques for portraying qualitative information such as "near"

or "close to" (Yao and Jiang 2005, Yao and Thill 2006) bring qualitative data into a GIS to enable analysis of how people talk about places without having to use geospatial coordinates. Closely related are the strategies of developing new software tools that allow methods of qualitative analysis such as open coding to be carried out within a GIS format (Kwan and Ding 2008) or actually bringing qualitative data into the data structures of a GIS (Jung 2009).

Another option is a mixed-methods approach with GIS and ethnography (Pain et al. 2006, Hurley et al. 2008), triangulating more traditional qualitative analysis with mapped quantitative or qualitative data. This allows the shortcomings of one method to be balanced by the strengths of another and may lead to insights precisely *because* of the different findings that different methods generate (Nightingale 2003). Grounded visualization (Knigge and Cope 2006) and geo-ethnography (Matthews et al. 2005) work back and forth between ethnographic materials and mapping as part of the research design, informing both researchers and subjects about the topic at hand and making use of both groups' insights in the research process and the results.

Finally, visualization is one of the key contributions of GIS in general, and no less to qualitative GIS. The use of time-space maps (Kwan 2002, Kwan 2008), expressing lived spaces in cartographic form (Wickens Pearce 2008), and visualizing historical data in new and insightful ways (Gregory and Healey 2007) all take advantage of modern computing power to display information in innovative ways that may open lines of inquiry that could not have been explored using more traditional methods. This paper does the same, answering Pavolvskaya's call to use spatial representations to bring

spatiality into our explanations of geographic phenomena by visualizing content analysis as a form of qualitative GIS.

### *Geographical content analysis*

Content analysis as a method has been an important part of geographical research for many years, whether via quantitative measures of keyword counts and article lengths or qualitative analyses of discourses employed. Content analysis has been used to explore the discipline of geography itself in terms of the content and emphasis of journal articles (Wheeler 2001, Jackson et al. 2006). Similar studies of textbooks and secondary school standards reveal the extent to which primary and secondary geographic education are shaped by textbook content (Tani 2004, Standish 2008).

Most use of content analysis within geographical research has used newspaper articles as the data source. Because of the importance of print (and online) media in informing the public about environmental and political issues in particular, newspaper articles are an important source of information about how the average person perceives their surroundings (Wakefield and Elliott 2003, Dittmer 2005). In particular, the framing of an environmental issue is largely done through the media, often by writers who have not traveled to the places they are writing about and therefore rely on common tropes about these places (Larsen and Brock 2005). For political issues, the framing of stories is also key to how they are understood by the public, whether in terms of understanding a region (Dittmer 2005) or justifying a war (Falah et al. 2006). Media representations may also reinforce existing stereotypes and inhibit alternative understandings of social groups or controversies (Bauder 2005). However, while members of the public may be aware of

the biased or incomplete nature of news articles, they still often rely on them for their primary source of information (Wakefield and Elliott 2003.)

In other contexts, visual content analysis has been used to great effect (Rose 2007), including Lutz and Collins' (1993) analysis of photographs in *National Geographic* magazine and how they convey the non-Western world to a largely Western audience. Audiences may also receive powerful messages from popular culture, as with comic books meant to shape national identity (Dittmer 2007). Other written materials with widespread popular influence include Presidential speeches (O'Loughlin and Grant 1990, Jansson 2004); content analysis can show which places are represented in those speeches and how, as well as the implications for domestic and foreign policy. This paper explores a novel kind of content by using public meeting transcripts as well as newspaper articles, but it also extends content analysis to incorporate new techniques and methods of visualization, namely tag or content clouds.

#### *Tag clouds and content clouds*

The internet-based services known as Web 2.0 are based on interactive features that enable users to contribute to the content of a page and to navigate their own way through the World Wide Web (Cosh et al. 2008). One of these features is the use of tag clouds, sizing the tags or keywords on a page according to their frequency of use as a means of visualizing content (Viégas and Wattenberg 2008), an adaptation of earlier work that sized place names on a mental map according to how often they were mentioned by research subjects (Milgram and Jodelet 1976). Users can click on a tag in order to connect to entries or pages that match their particular interest. Tag clouds thus form a

means of navigation through a website as well as what has been termed a folksonomy or vernacular classification system (Morrison 2007).

Much of the published work on tag clouds has been about developing algorithms to make them more user-friendly or attractive (Halvey and Keane 2007, Cosh et al. 2008, Siefert 2008). Applications have also been developed to extend the tag cloud concept beyond blog or photo tags. These *concept clouds*, word clouds, or semantic clouds (Krygier and Wood 2009) are based on the same idea of font size being proportional to word usage. For example, one application enables the creation of content clouds based on movie scripts to elicit themes from the scripts (McKie 2007). Content clouds have become a popular means of analyzing political speeches because of their ease of generation through a website like ManyEyes or TagCrowd (see Figure 1).

There are some caveats to consider when using content clouds in analysis. For example, one user evaluation study found that content clouds were better for producing descriptive information than for relational concepts (Kuo et al. 2007). Unsurprisingly, words in larger font stand out more and may distract from other important tags or concepts that were not mentioned as frequently, thus biasing the results of an analysis (Halvey and Keane 2007). Additionally, either single words or pairs of words can be identified, but not both at once; in the second example below, this is evident in the content cloud generated from Albuquerque newspapers that includes "Mexico" as a commonly mentioned word but in reference to the state of New Mexico rather than the country. Nevertheless, concept clouds can offer a quick means of visual exploratory data analysis that is easy to interpret and display (Viégas and Wattenberg 2008).

## **Methods**

This paper presents a method of exploratory qualitative data analysis: generating content clouds and mapping them in a GIS. There are two case studies, each on a different topic and each using a different data source to generate the content clouds. This section explains the means by which the data were obtained and analyzed.

### *Creating content clouds*

The website TagCrowd.com, created by Daniel Steinbock of Stanford University, was used to generate the content clouds for this project. While there are numerous websites available for creating tag clouds or word clouds (such as tagcloud.com, Wordle, and Make Eyes), TagCrowd was chosen because it is free for non-commercial use, allows users to upload files rather than using only online content, and focuses on a simple visualization pattern rather than more complicated options. It also incorporates words with the same stem but different endings, such as "develop," "developer," and "development", into one entry, eliminating duplication of concepts and enabling a wider range of content to appear.

Users can either paste text into a box on the webpage or upload a file to be analyzed. Users can also choose how many words are displayed in a content cloud; the default number is fifty, which was used here. TagCrowd automatically ignores words such as "the" as part of what it calls a *stop list*; users can upload their own list of words to be ignored based on the content they are analyzing. These lists are included in the discussion of the two case studies below.

### *Example 1: Public Meeting Transcripts*

The first example uses a series of public meeting transcripts from the environmental review of the purchase of a beltline railroad around Chicago by a major North American railroad. In the fall of 2007, Canadian National (CN) filed an application with the Surface Transportation Board (STB) to purchase the Elgin, Joliet, and Eastern Railroad (EJ&E), which circumnavigates three-quarters of a circle around Chicago about thirty-five miles away from downtown. What used to be farmland on either side of the tracks has largely become suburban development, from some of the regions' wealthiest suburbs on the northwest side to what are currently the fastest-growing suburbs to the southwest to the deindustrialized area of northwest Indiana.

CN's purchase of the EJ&E is meant to alleviate the congestion affecting their trains by providing it with a bypass around the central city. Although an environmental review process is not usually required for a transaction such as this, the STB decided that the potential existed for environmental impacts along the route and therefore required an Environmental Impact Statement. A series of public meetings were held in August and September 2008 to get comments on the Draft EIS, nearly all of which were critical of the proposed increase in train traffic through their towns from around five trains a day to around thirty. The analysis that follows is based on transcripts from those public meetings, which are available on the STB website in PDF format.

One content cloud was generated for each public meeting. Before creating the content clouds, the transcripts had "public meeting" and the names of the moderator and the court reporter removed because of their frequency of use. Within TagCrowd.com, the following words were used on the stop list: Mr., Ms., evening, tonight, thank, and board, all of which were frequently used in reference to the meeting itself and the members of

the STB who were present, not to the subject at hand. The resulting content clouds were mapped according to the respective location of the public meetings and will be discussed in the next section.

*Example 2: Green buildings*

The building industry is being rapidly transformed by the Leadership in Energy and Environmental Design (LEED) standards of the U.S. Green Building Council (USGBC), a non-profit organization based in Washington, DC. Over two thousand green buildings have been certified by the USGBC across the U.S., in all fifty states (Cidell 2009). One of the reasons for LEED's success is that it sets up a flexible but standardized set of requirements for defining a "green" building. Nevertheless, implementation of those standards can vary from region to region (Cidell and Beata 2008) based on priorities for conserving water, energy, materials, or other factors. Furthermore, the multiple reasons for building green include environmental, social, and economic motivations that may or may not conflict with each other. For example, green buildings can be framed as saving money in the long term despite higher initial costs, or as protecting the natural environment, or as providing a better working environment for employees.

Most people do not live or work in a green building, and therefore get most of their information about what green buildings are from other sources such as news media. Analyzing the content of newspaper articles about green buildings can therefore tell us what kind of information the average person is receiving about green buildings. For this portion of the study, newspaper articles were collected over a three-month period in early 2008 from thirteen major newspapers in metropolitan areas chosen across the U.S. to balance regional location and population size (and on the free availability of newspaper

archives). The newspapers are listed in Table I. Articles including the terms "green", "building" and "LEED" were downloaded and combined into a single text file and stripped of all headers (such as newspaper name, byline, etc.). The terms "green building", "LEED", and "USGBC" were on the stop list for TagCrowd analysis. For display purposes, only some of the content clouds are mapped here.

[insert Table I here]

### **Content cloud case studies**

This section demonstrates two main ways that content clouds can be used for exploratory data analysis, based on the two examples described above. The first involves looking at the general pattern of what matters most in a particular location, including which geographic scales and specific issues appear. The second is to take a word common to all or most of the content clouds and compare it across locations to see how its frequency of use varies. Both of these approaches were applied to the two case studies.

#### ***Example 1: Public Meeting Transcripts***

Figure 2 shows the content clouds mapped across the region by the location of the public meeting (sized as explained below). The general pattern varies across the sites, suggesting that the same issue—a significant increase in freight rail traffic along an existing, lightly-used rail line—is perceived very differently from place to place. This has important implications for government regulators seeking to broker a compromise between CN and local communities. It also demonstrates that a single decision has environmental impacts that vary from place to place depending on their existing physical and social geographies.

[insert Figure 2 here]

In Mundelein to the north, the most common terms are fairly generic: "CN" and "EJ" (from EJ&E), "impact,"<sup>1</sup> "line," "traffic," and "train," with "Lake Zurich" visible as a city name split in two. In the northwest, "Barrington" obviously matters the most to people in Barrington (to be fair, this area includes several municipalities based on that name, such as Lake Barrington or Barrington Hills). The subtle difference between "community" here and "communities" in the first transcript may be of interest, indicating the extent to which speakers are talking about only the municipality they are from or are trying to draw comparisons across suburban borders. To the west, the meeting held in Bartlett obviously had attendees from that city as well as West Chicago, while the same terms of "CN," "EJ," and "trains" were present. Here, "safety" was as prominent as "traffic." Also to the west, Aurora speakers had more diffuse concerns; in particular, "traffic" appears as frequently as "train" because of a major arterial having an at-grade crossing with the EJ&E tracks.

On the southwest side of the region, speakers in Joliet made comments similar to those in Mundelein, including more mention of a neighboring town (Plainfield) than the larger city where the meeting itself was located. This might suggest that these public meetings were not located close enough to the most affected populations, or that residents from other communities are simply more vocal. On the south side, in Matteson, "Chicago" was mentioned quite frequently, along with "area" and "region", suggesting a more widespread outlook beyond a single municipality. "Mitigation" was also mentioned more often, suggesting speakers at this meeting were more interested in reaching a compromise

---

<sup>1</sup> Because "statement" does not also show up in most of the content clouds, we can be confident that "impact" is not only being used as part of "environmental impact statement".

with CN than in other parts of the region. To the southeast in Gary, the only meeting in Indiana, the state name predominated, along with "town", the equivalent of village or city in Illinois, again suggesting a bigger picture approach. Finally, the content from Chicago is obviously different because here train traffic would be removed, not added. "Benefit" stands out along with "community", "impact", "rail", and "region."

The second way in which multiple content clouds are useful is to take the same word and trace it across different locations. There are expected terms like "rail" and "train" that are always high in frequency, but others vary quite a bit. "Safety" is quite large in Bartlett, of medium importance in Mundelein, Barrington, and Aurora, occasionally mentioned in Matteson and Gary, and not on the top-fifty list for Joliet or Chicago. This frequency of mention might depend on the location of the EJ&E tracks within a city, for example, whether they run through the downtown as in Barrington or West Chicago (near Bartlett) or are on the municipal edge as in Matteson or Aurora.

"Freight" is of the most importance to the north in Mundelein, less so around the circle in Barrington, Bartlett, and Aurora, and not in the top fifty words for the meetings on the southern half of the region. The majority of firms currently using the EJ&E for freight traffic are on the south and southeast sides, so the change in use would not be as severe for these areas. "School" follows the same pattern, with much more frequent mention in the north, northwest, and west suburbs. This might be due to the more-developed nature of these areas, with populations and school districts that sprawl across both sides of the tracks and therefore have a greater chance of transport to and from school being hindered by a freight train. "Economic" only makes the list in Matteson and Chicago, which

connects back to the more regional nature of comments in the former and the concept of benefits in the latter.

In order to triangulate the content clouds against another method of analysis, comments made at the public meetings were classified according to whether they referred to local (e.g., an individual railroad crossing), municipal, regional, state, national, or international concerns. Figure 2 shows the meeting locations sized by the number of average references per commenter to a regional issue (not necessarily by using the word “region”). The map confirms that speakers at the meetings in the south and southeast suburbs and Chicago, which tend to be more in favor of the transaction because of the positive impact, focused more on the region as a whole than speakers at other meetings, as shown in the exploratory analysis with the content clouds.

### ***Example 2: Green building articles***

As explained above, newspaper articles are an important source of information for citizens on environmental issues—sometimes the only source. With that in mind, it is important to know how an emerging environmental issue like green buildings is being portrayed in the media. The acceptance of a new way of building is likely to depend significantly on whether or not the average person thinks it is a good idea, both in terms of cost and environmental impact. The question is, how does information portrayed in the media about green buildings vary from region to region?

Content clouds were generated from thirteen major metropolitan newspapers; a selection is displayed in Figure 3 (Portland, OR, Grand Rapids, MI, Atlanta, GA, Birmingham, AL, and Albuquerque, NM), along with the average score of all LEED-certified buildings within each metropolitan area (classified as Certified, Silver, Gold, and Platinum and

worth one to four points respectively). The content clouds show that people in different regions are likely to come to a very different understanding of green buildings based on newspaper content alone. As with the first example, we begin by looking at the general pattern of terms used. "Energy" was one of the most common across all locations, along with "design", "city", and "environmental"<sup>2</sup>. The five examples show a diversity of approaches to reporting on green buildings. In Grand Rapids, there was more of an emphasis on the methods employed in construction, including "floor", "forest", "friendly", "light", "materials", and "recycled". Atlanta tended to highlight specific projects, including those at Emory and Georgia Tech Universities. It appears that most of Birmingham's green projects are related to schools, since that was the second most frequently mentioned word; connections among multiple scales and locations were readily apparent here, including "federal", "Alabama", "Atlanta", "city", and "state". Finally, Albuquerque and Portland, had the most emphasis on the institutions of green building, with "certification", "city", "council", "program", and "standards" appearing more frequently here than elsewhere.

[insert Figure 3 here]

The inconsistency of use of different terms across the country was interesting as well. For example, "water" was very important in Albuquerque and Atlanta; less so in Boston and Portland, OR; only occasionally mentioned in Salt Lake City, Chicago, Grand Rapids, Syracuse, Pittsburgh, New York, and Washington, DC; and not in the top fifty words in Madison or Birmingham. One-third of the cities mentioned "efficiency" more

---

<sup>2</sup> As with "environmental impact statement", the fact that "leadership" did not show up as frequently as the other three components of the LEED acronym suggests that the frequent mention of "energy," "environmental," and "design" is not simply writing out the acronym.

often than "sustainable," suggesting a focus on economy rather than environment in defining the value of a green building. "Certification" was very important in Albuquerque, Portland, and Madison; somewhat important in Salt Lake City, Chicago, Grand Rapids, Birmingham, Atlanta, Washington, DC, Pittsburgh, and Syracuse; and barely mentioned in Boston and New York. As an indicator of the importance of cost, "million" scored high in Grand Rapids and Birmingham; moderately in Albuquerque, Salt Lake City, Madison, Atlanta, Washington, DC, Pittsburgh, New York, and Syracuse; and not very important in Portland and Chicago.

Other categories that differed across space beyond the mention of specific places, projects, or firms included the methods that were employed in construction (e.g., "recycled" or "solar"); the types of projects (e.g., "school", "house", "office"); or the actors involved in green building (e.g., "mayor", "builders", "occupants"). While there is less of a spatial pattern apparent here than in the previous case study, this is not as comprehensive an example; a more focused comparison of, for example, multiple cities in the Pacific Northwest or the Southwest might find clearer regional differences.

Comparing the content clouds to the layer of the map showing the average greenness of buildings in the metropolitan areas in question demonstrates the contributions of qualitative GIS in this case. The choropleth map shows cities classified by the average score of their green buildings; a darker color indicates more Gold or Platinum buildings, while a lighter color indicates more Certified or Silver. Grand Rapids and Birmingham have relatively light-green buildings, and their content clouds show more emphasis on cost ("millions"), "light", and "school" and "students". Darker-green Albuquerque, Atlanta, and Portland have in common an emphasis on "energy", "programs",

“sustainability”, and “water”. Where cost predominates as a concern (which would include school projects), green buildings appear to settle for meeting the minimum requirements, whereas in metropolitan regions where sustainability and conservation are more valued (as portrayed in the newspaper) and more likely to be part of official policies or programs, green buildings generally achieve higher standards.

### **Content clouds as a methodology**

The above examples demonstrate some of the ways that content clouds can be used in exploratory qualitative data analysis. At the most basic level, both the transportation and green building examples clearly demonstrate that place matters: the same issue or topic is discussed and understood in different ways across the country and even across a region. Further information is obviously needed to both give a context to the different clouds and explain their results, but many possible avenues for research can be suggested by a quick look.

For example, one possible next step with the green building content clouds might be to search through the newspaper articles for a word such as "water" or "efficiency" and understand the different contexts in which it was used, or to incorporate the political slant of the newspapers in question to see if this might explain why they framed their articles the way they did. Additional spatial analysis could compare the number of credits earned by building projects for specific elements like water conservation or recycled materials to see if there is a match with the content clouds; for example, do green buildings in Albuquerque include more water conservation measures than in other regions, or does the local paper report on them more because of the need to conserve water in the region?

Further analysis with the CN/EJ&E data might include mapping demographic information such as income or race beneath the content clouds. The frequency of mention of scales other than the regional could be compared to the content clouds as well. Additional content clouds could be generated and compared from other data sources such as Environmental Impact Statements, news articles, or interview transcripts. Activists might be interested to see the overall impressions generated by the content clouds in order to suggest different arguments to put forward in the future.

Beyond analysis, content clouds can be useful for presenting information to an audience; for example, public meetings could include a summary of previous comments placed on a map to indicate which topics had been brought up where. Keeping in mind the caveats about reading content clouds (Halvey and Keane 2007), they can usefully summarize information and allow a quick visual comparison across sites. Using content clouds can also be advantageous because they present information in writers' or speakers' own words, if one word at a time. For example, the STB summarized the topics of comments they received using their own categories such as "hazardous materials" or "traffic congestion" in their response to the draft EIS. This relies on STB staffers' judgment to correctly classify comments rather than directly presenting speakers' own words.

There are some cautions to keep in mind when using this methodology. First, the content clouds are all independently sized. In other words, "energy" might have been the most commonly used word in newspaper coverage of green buildings, but the number of times it was mentioned varies from city to city based on the total number of articles and/or their length. Additionally, all of the usage counts are relational; for example, "water" might be mentioned an equal number of times in Atlanta and Boston, but there might be many

more words that are mentioned more frequently in Boston, thus pushing "water" down the list.

There are also difficulties involved in analyzing one word at a time, as discussed above with regard to place names such as New Mexico. The website for TagCrowd mentions that future capabilities will include counting two-word phrases and single words at the same time, which would help with this problem. Additionally, while TagCrowd combines words with like roots and different endings, it obviously misses a few (note "rail" and "railroad" or "certification" and "certified"). Trial and error can lead to more useful stop lists; for example, "train" and "rail" could be pulled out of the CN/EJ&E content clouds to allow more words with less frequent usage to appear. Similarly, words that do not contribute to the topic at hand such as "according" or "being" could be added to the stop list to bring out more content-relevant information.

Nevertheless, content clouds have the potential to be a useful form of exploratory qualitative data analysis and to contribute to qualitative GIS. The examples presented here suggest that not only can content clouds demonstrate in a broad way that place matters to the study of environmental and many other issues, but they can offer a starting point for exploring precisely how place matters. This includes tracing a single topic from location to location as well as understanding *how* an issue is described in general terms in a particular place. Using content cloud visualization within qualitative GIS creates a spatial representation of an environmental (or other) issue that makes space and place matter, and in an way that is easy to generate and read.

As a type of qualitative GIS, mapping content clouds does not alter data structures or spatial analysis techniques (Jung 2009). It does, however, introduce a new kind of spatial

data in the form of concise content analyses based at a particular location. Combining these clouds with other kinds of data, as in Figures 2 and 3, allows the researcher to literally situate the knowledge produced by the content clouds within a broader context of spatial data. This method therefore offers the potential of combining content analysis, visualization, and qualitative GIS, bringing “non-cartographic spatial knowledge into conventional GIS” (Elwood and Cope 2009, p. 1) and offering another means of carrying out qualitative analyses within a geographic information system.

## Works Cited

- Bauder H** 2005 Landscape and scale in media representations: the construction of offshore farm labour in Ontario, Canada *cultural geographies* 12 41-58.
- Cidell J and Beata A** 2009 Spatial variation among green building certification categories: does place matter? *Landscape and Urban Planning* 91 142-151.
- Cidell J** 2009 Building green: the geography of LEED-certified buildings and professionals *Professional Geographer* 61 1-16.
- Cosh K Burns R and Daniel T** 2008 Content clouds: classifying content in Web 2.0 *Library Review* 57 722-729.
- Dittmer J** 2005 NATO, the EU and Central Europe: differing symbolic shapes in newspaper accounts of enlargement *Geopolitics* 10 76-98.
- Dittmer J** 2007 "America is safe while its boys and girls believe in its creeds!": Captain America and American identity prior to World War 2 *Environment and Planning D* 25 401-423.
- Elwood S and Cope M** 2009 Introduction: Qualitative GIS: forging mixed methods through representations, analytical innovations, and conceptual engagements in **Cope M and Elwood S** Qualitative GIS: A Mixed Methods Approach London Sage 1-12.
- Falah G-W Flint C and Mamadouh V** 2006 Just war and extraterritoriality: the popular geopolitics of the United States' war on Iraq as reflected in newspapers of the Arab world *Annals of the Association of American Geographers* 96 142-164.
- Gregory I and Healey R** 2007 Historical GIS: structuring, mapping and analysing geographies of the past *Progress in Human Geography* 31 638-653

- Halvey M and Keane M** 2007 An assessment of tag presentation techniques. *Proceedings of the 16th International World Wide Web Conference Banff Canada* 1313-1314.
- Hurley P Halfacre A Levine N and Burke M** 2008 Finding a "disappearing" nontimber forest resource: using grounded visualization to explore urbanization impacts on sweetgrass basketmaking in Greater Mount Pleasant, South Carolina *The Professional Geographer* 60 556-578.
- Jackson A Harris R Hepple L Hoare A Johnston R Jones K and Plummer P** 2006 Geography's changing lexicon: measuring disciplinary change in Anglophone human geography through journal content analysis *Geoforum* 37 447-454.
- Jansson D** 2004 American hegemony and the irony of C. Vann Woodward's "The irony of southern history" *Southeastern Geographer* 44 90-114.
- Jung J-K** 2009 Computer-aided qualitative GIS: a software-level integration of qualitative research and GIS in **Cope M and Elwood S** Qualitative GIS: A Mixed Methods Approach London Sage 115-135.
- Knigge L and Cope M** 2006 Grounded visualization: integrating the analysis of qualitative and quantitative data through grounded theory and visualization *Environment and Planning A* 38 2021-2037.
- Krygier J and Wood D** 2009 Ce n'est pas le monde in **Dodge M Kitchin R and Perkins C** eds Rethinking Maps: New Frontiers in Cartographic Theory London Routledge 189-219.

**Kuo B Hentrich T Good B and Wilkinson M** 2007 Tag clouds for summarizing web search results *Proceedings of the 16th International World Wide Web Conference* New York ACM Press 1203-1204.

**Kwan M-P** 2002 Feminist visualization: re-envisioning GIS as a method in feminist geographic research *Annals of the Association of American Geographers* 92 645-661.

**Kwan M-P** 2008 From oral histories to visual narratives: re-presenting the post-September 11 experiences of the Muslim women in the United States *Social and Cultural Geography* 9 652-669.

**Kwan M-P and Ding G** 2008 Geo-narrative: extending geographic information systems for narrative analysis in qualitative and mixed-method research *Professional Geographer* 60 443-465.

**Larsen S and Brock T** 2005 Great Basin imagery in newspaper coverage of Yucca Mountain *Geographical Review* 95 517-536.

**Lutz C and Collins J** 1993 Reading National Geographic Chicago University of Chicago Press.

**Matthews S Detwiler J and Burton L** 2005 Geo-ethnography: coupling geographic information analysis techniques with ethnographic methods in urban research *Cartographica* 40 75-90.

**McKie S** 2007 Scriptcloud.com: content clouds for screenplays *Second International Workshop on Semantic Media Adaptation and Personalization* 221-224.

**Milgram S and Jodelet D** 1976 Psychological maps of Paris in **Proshansky W and Rivlin L** eds Environmental Psychology New York Holt, Rinehart, and Winston 104-124.

**Morrison P** 2007 Why are they tagging, and why do we want them to? *Bulletin of the American Society for Information Science and Technology* 34 12-15.

**Nightingale A** 2003 A feminist in the forest: situated knowledges and mixing methods in natural resource management *ACME: An International E-Journal for Critical Geographies* 2 77-90.

**O'Loughlin J and Grant R** 1990 The political geography of presidential speeches, 1946-87. *Annals of the Association of American Geographers* 80 504-530.

**Pain R MacFarlane R Turner K and Gill S** 2006 "When, where, if and but": qualifying GIS and the effect of streetlighting on crime and fear *Environment and Planning A* 38 2055-2074.

**Pavlovskaya M** 2006 Theorizing with GIS: a tool for critical geographies? *Environment and Planning A* 38 2003-2020.

**Rose G** 2007 Visual Methodologies: An Introduction to the Interpretation of Visual Materials 2nd edition London Sage.

**Seifert C Kump B Kienreich W Granitzer G and Granitzer M** 2008 On the beauty and usability of tag clouds *Proceedings of the 12<sup>th</sup> International Conference on Information Visualisation* 17-25.

**Standish A** 2008 Changing perspectives in high school world geography: 1950-2005 *Journal of Geography* 107 121-130.

- Tani S** 2004 Curriculum reform and primary geography in Finland: a gap between theory and practice? *International Research in Geographical and Environmental Education* 13 6-20.
- Viégas F and Wattenberg M** 2008 Tag clouds and the case for vernacular visualization *Interactions* July + August 49-52.
- Wakefield S and Elliott S** 2003 Constructing the news: the role of local newspapers in environmental risk communication *Professional Geographer* 55 216-226.
- Wheeler J** 2001 Assessing the role of spatial analysis in urban geography in the 1960s *Urban Geography* 22 59-558.
- Wickens Pearce M** 2008 Framing the days: place and narrative in cartography *Cartography and Geographic Information Science* 35 17-32.
- Yao X and Jiang B** 2005 Visualization of qualitative locations in geographic information systems *Cartography and Geographic Information Science* 32 219-229.
- Yao X and Thill J-C** 2004 Spatial queries with qualitative locations in spatial information systems *Computers, Environment and Urban Systems* 30 485-502.

<i>Metropolitan area</i>	<i>Newspaper</i>
Albuquerque, NM	<u>Albuquerque Journal</u>
Atlanta, GA	<u>Atlanta Journal and Constitution</u>
Birmingham, AL	<u>Birmingham News</u>
Boston, MA	<u>Boston Globe</u>
Chicago, IL	<u>Chicago Tribune</u>
Grand Rapids, MI	<u>Grand Rapids Press</u>
Madison, WI	<u>Capital Times</u>
New York, NY	<u>New York Times</u>
Pittsburgh, PA	<u>Pittsburgh Post-Gazette</u>
Portland, OR	<u>The Oregonian</u>
Salt Lake City, UT	<u>Salt Lake Tribune</u>

Table I. Newspapers from which articles on green buildings were obtained for content cloud analysis.

## List of Figures

Figure 1. Sample content cloud of Barack Obama's Inaugural Address, generated by the author using TagCrowd.com.

Figure 2. Content clouds from CN/EJ&E public meeting transcripts, generated by the author using TagCrowd.com.

Figure 3. Sample content clouds from green building newspaper articles, generated by the author using TagCrowd.com.