Introduction
by Diane Neal, Guest Editor of Special Section

Visual Representation, Search and Retrieval: Ways of Seeing

While most of us still read print media, the widespread existence of visual information in our culture is undeniable. Today’s visually oriented society demands further development in the area of visual representation, search and retrieval. Websites such as Flickr [1] and YouTube [2] contain millions of visual documents, but it is not always easy to find the exact video or photograph you “just know” exists on one of these sites. Increasingly, major websites such as cnn.com and msn.com provide video and still images to supplement or replace text-based articles, although there is no useful way to search for them. Google Earth’s [3] ability to overlay user-supplied geographic information, photographs and other data onto maps is revolutionary, but finding the appropriate files to display what we want to see can be tricky. The newspaper USA Today [4] is known for its striking visual representations of quantitative data, but it is difficult to find them again after publication.

The increase in the volume of digital visual information is certain to continue. Digital still and video cameras are the norm, and photograph storage services such as Flickr and Shutterfly [5] continue to grow at amazingly fast rates. Increasingly, people are watching television shows and films through on-demand websites such as Netflix [6], Joost [7] and Hulu [8]. Worldwide, cultural institutions are undertaking digitization projects to allow people to view their collections online; the Library of Congress Photos on Flickr project [9, 10] is only one of many. Over 4 million people use Fitday [11], a website that allows dieters to track progress toward their weight loss goals visually. The trend toward online visual information is expected to continue: recent research has demonstrated that Millennials, or people born approximately between the years of 1982 and 2001, prefer to learn visually, and 93% of American teenagers use the Internet. Given the wealth of visual information in existence and the assured continuance of this trend, it is essential for information science professionals to develop better ways to organize, store, access and retrieve it. This need fuels the motivation behind this special section about visual representation, search and retrieval, sponsored by ASIS&T’s Special Interest Group/Visualization, Images and Sound (SIG/VIS).

Defining the Concerns

When you look at a visual document, such as a photograph, a film or a graphical depiction of quantitative data, what happens? Think about whether any of these apply to you:

- You do not necessarily know how to describe all aspects of it in words.
- You understand the gist of the document rather quickly, but deeper interpretation sometimes takes additional time.
- You have emotional reactions to it.
- You notice things about it that others do not notice.

For example, consider Leonardo da Vinci’s famous painting the Mona Lisa (Figure 1).

Diane Neal was an assistant professor in the School of Library and Information Sciences at North Carolina Central University at the time she created this special section. She will join the Faculty of Information and Media Studies at the University of Western Ontario on July 1, 2009. She can be contacted at the following email address: diane_neal<at>sbcglobal.net.
You can explain in words that the painting features a woman with straight dark hair, dark eyes and dark clothing. But how do you define the look on her face? How is her gaze leading her thoughts? Where is she, exactly? How does she make you feel? Is the background filled with mountains, trees or something else? Would a friend viewing it with you online or standing next to you in a museum answer these questions with the same responses? Deliberations such as these are central to this special section on visual search and retrieval.

There are actually two separate issues within the topic. First, describing and searching for visual materials themselves, such as photographs, films or graphical depictions of quantitative data, is difficult for several reasons. Words cannot be automatically extracted from visual documents to be used as search terms. If we want to search visual materials using words, we must assign the terms manually, which is a time-consuming and subjective process. Additionally, since visual materials are not words, and vice versa, we can be assured that something gets lost in the translation between the intellectual content of a visual document and the words we use to describe it. If words are not always the optimal method in which to search for and retrieve visual documents, then what other methods of representation and retrieval are available to us? Researchers and practitioners in the area of visual information representation and retrieval are actively seeking to answer this difficult question.

The second issue relates to visual displays of quantitative information. According to research on human vision and cognitive processing, humans process visual information, such as pictures, much faster than text-based information. Think about how quickly you can scan a page of thumbnail images on sites such as Google Images or Flickr to determine whether the one you want to see is displayed; compare that to how long it takes you to look at a list of search results in a text-based search engine. This principle is demonstrated in data displays as well. Visual displays of quantitative information allow us to process deceptively large amounts of data very quickly. Consider how difficult it would be to display – and understand – the data in the following two graphics if they were represented textually in a tabular spreadsheet or other non-graphical format:

1. A local interactive weather radar map from Weather.com [12] for Austin, Texas, for the afternoon of April 17, 2009, is illustrated in Figure 2. Weather visualizations, which are common both online and on television news broadcasts, allow us to quickly determine whether we should pack an umbrella, a jacket or a bottle of sunscreen as we head out for the day. The amount of data in this graphic would be staggering in a text-based, tabular format.

2. A sample business intelligence (BI) dashboard from InformationBuilders.com [13] is displayed in Figure 3. BI involves using organizational data to guide business decisions. Visually oriented intelligence dashboards such as this one allow businesspeople to analyze key performance indicators (KPIs), sales statistics and other BI information at a glance.
Representation

Every visual document is a representation or a surrogate of an actual object. A photograph is a surrogate of the particular point in time and place that is captured in the photograph. A painting is a surrogate of a scene, whether it existed physically or in the artist’s mind. A data visualization represents a dataset.

Visual documents, which are surrogates by nature, require their own representation in a search and retrieval system. Without effective and appropriate representation, search and retrieval, attempts will likely be unsuccessful. The method of desired representation depends on a variety of factors. For example, are the users experts on the collection or is access geared toward casual browsers? What formats are included in the collection is also important; films, photographs, works of art, data visualizations, gaming environments and so on call for different representation approaches. These classification and description issues are important to consider in any document collection, but the subjective nature and the lack of native metadata in visual documents compound the concerns.

Concept-based image retrieval, or the use of human-assigned words to describe, search for and retrieve images, is the most prevalent method in library practice as well as in library and information science education and research. A variety of methods have been implemented to achieve this approach. For example, controlled vocabularies that list the terms that can be assigned to a document, such as the Library of Congress Subject Headings (LCSH) [14] and the Art and Architecture Thesaurus (AAT) [15], are used in many libraries’ collections. At the other end of the spectrum, folksonomies present on social websites such as Flickr and YouTube allow users to contribute their own keywords with no restrictions placed on their choices.

As discussed above, while words are useful for describing certain aspects of a visual document, words cannot capture some essences of them, because meaning is lost in the translation. The concept-based image retrieval approach, which focuses on semantics, has not yet been successful in utilizing pictures to describe pictures. This area is definitely in need of research.

Computer scientists develop algorithms that allow images to describe other images, although these products are mostly limited to creating relationships between the physical aspects of an image, such as colors, lines and shapes, and patterns present in the picture. The technique is known as content-based image retrieval. Several approaches exist, but most commonly, users can identify one picture in the search engine after completing a traditional textual search and then indicate that they want to find “more like this,” using a technique called query by example (QBE). A commercial example is available at like.com [16], which allows users to shop for clothing based on similar features (satin black dresses, as opposed to cotton blue pants, for example). QBE is useful in this context, but does not necessarily meet the needs of all users’ information-seeking contexts.

Representation present in data visualizations calls for separate consideration. Data visualization techniques, as developed by icons in the field such as Edward Tufte [17] and Ben Shneiderman [18], are certainly a welcome relief to analyzing raw forms of textual tabular data. The applications of these basic techniques are being extended and refined to solve other information problems creatively. For instance, visualization can be used to represent relationships between words and meanings or word-based searches. Visual Thesaurus [19] helps us find related words through spatial associations, such as the example for the word dog in Figure 4.

In the case of visual search engines, the surrogate is a copy of the web page itself. In this example, the user can scroll through snapshots of the actual...
websites, rather than read a list of text-based surrogates for the pages, as in traditional search engines such as Google. Like traditional search engines, a search for the rock band U2 in the visual search engine searchme.com [20] displays the most popular or relevant websites first, with U2.com, U2’s last.fm page and U2’s Myspace Music page as the top results (Figure 5). The links at the top of the searchme.com page allow users to limit their searches by formats such as video, images or music by subtopics such as Christianity, tickets, forums.

However, future work is needed to ensure that data surrogates within the visualization itself make intuitive sense to the viewer. Standard methods of visualization, such as line graphs, histograms and scatter plots, all contain abstract representations of the data, which may call intuitive conveyance of information into question. Graphics such as “A closer look at fertility data over the years,” from MSNBC.com [21] (Figure 6), provide examples of how visualizations can sometimes contain too much information. The fact that a solid green line represents fertility rates of women aged 30-34, a dotted green line indicates twin birth rate per 1,000 women and so on, might cause cognitive overload for many viewers.

The Glass Engine, which allows users to serendipitously access and listen to music by the composer Philip Glass [22] (Figure 7), presents a different type of visual representation problem. The tiny blue vertical lines each represent a work, and the lines comprise large sliding bars that the user can individually manipulate to select a work. The small white squares outlined in black indicate the amount of joy, sorrow, intensity or density present in the currently selected work. Are these representations intuitive to most users?

Issues such as these still present translation issues in some contexts of data visualization and point the way for a new generation of interface design development.
The Future

Given the increase in the existence of digital visual information, the indicators pointing toward a continued trend in this direction and the relative newness of the area, additional research and development is greatly needed from all areas of information science. Information behavior researchers and usability professionals might consider gaining a better understanding of how people want to find and process visual information in various contexts. Metadata experts can develop new approaches to describing and organizing visual documents. Information technology professionals who design BI reporting tools and other visual displays of quantitative information could explore new ways of conveying that information. Web designers and search engine programmers can evaluate the few existing visual search engines and consider how they could be further improved.

Visual immersion is an exciting pathway to follow in the future of visual information. Members of the research team Prometheus (of which this author is a member) have found that a therapeutic, immersive video game may decrease the symptoms of Attention Deficit/Hyperactivity Disorder (AD/HD) in children without the use of medication. Researchers Pattie Maes and Pranav Mistry of the MIT Media Lab are developing a “sixth sense” tool that would allow us to interact with our environment and information that enhances it in seamless, unprecedented manners. As we continue to develop our existing methods of visual information – and plunge into the untested waters of immersive visual environments – we must not forget to evaluate the human risks and benefits of every approach and design accordingly.

In This Section

The articles in this special section provide a spectrum of perspectives on the problem of visual search and retrieval. Practitioners, researchers and visionaries allow us to ponder current implementations and future directions and inspire us to consider how we might advance the area in our own professional contexts.

“Information Visualization Services in a Library? – A Public Health Case Study” by Barrie Hayes, Andrés Villaveces and Hong Yi, all of the University of North Carolina at Chapel Hill (UNC-CH), presents a real-life solution to an information visualization need and demonstrates true collaboration in action. Dr. Villaveces, a researcher at UNC-CH’s Gillings School of Global Public Health, wanted to visually see the relationship between injury occurrence and interventions. He worked with Dr. Yi, a programmer with the Renaissance Computing Institute (RENCI), a collaborative North Carolina-based organization that promotes the fusion of technology, research and visualization tools, to create the necessary software. UNC-CH’s Health Sciences Library partnered with RENCI to acquire a large display wall for visualization applications, which is housed in the library.

Courtney Michael, Mayo Todorovic and Chris Beer describe the “Visualizing Television Archives” project in the Media Library and Archives at WGBH, Boston’s Public Broadcasting Service television station. In their efforts to make their vast multimedia archive available online via innovative visual access techniques to both the general public and academic researchers, they found very different user needs. The general audience enjoys browsing using thumbnails of the archived materials and utilizing a targeted search tool. The researchers desire deep access to the detailed metadata linked to each document. The article describes the development undertaken to give researchers the access they desire to the cataloging data as well as visualization tools, such as a results bar, facets, a mosaic and a relationship map. Their effort demonstrates a rare application of user needs analysis to innovative technology implementation.

In “Surveillance: Personal Edition,” Jodi Schneider and Nathan Yau show us how we can track personal information in useful ways using visualization tools, either online or on our desktops. For example, we can enter data about our exercise and diet habits, our moods or how much we drive our cars and view the trends in our behavioral patterns via graphs, charts and newer forms of visualization. It is also possible to view group-based trends in this manner, since many companies store and track our personal data for us. While this method of personal surveillance definitely has its advantages, we must be careful with what and how we disclose our personal information online in order to maintain our self-defined security and privacy boundaries.

Richard Anderson and Brian O’Connor present original research addressing the representation issues inherent in describing film in
“Reconstructing Bellour: Automating the Semiotic Analysis of Film.” In an effort to recreate Raymond Bellour’s frame-based structural analysis of Alfred Hitchcock’s film *The Birds* using digital technology, the authors analyze the color values in frames extracted from the Bodega Bay sequence of the film and the semiotic or semantic meaning of the frames. They conclude that a “separate, complementary” relationship exists between the physical structure and the semantic meaning. This research leads us toward the necessary but unrealized combining of content-based retrieval with concept-based retrieval to describe, search for and retrieve visual documents using other visual documents as surrogates and descriptions.

Ray Uzwysyn provides a look toward the future of image searching as well as visual search engine design in “An Arbitrage Opportunity for Image Search and Retrieval.” In the Google Image Labeler, user-assigned semantic descriptions for images are collected and implemented via a game-oriented format using human processing theories. He also discusses efforts to move the image search paradigm past the “photographic contact sheet” of thumbnails retrieved via a targeted text-based search, such as Cooliris, a search engine that displays results in a 3D “film reel” format. He believes that humans and machines can leverage or “arbitrage” from each other’s strengths to produce a synergy that will move the field forward.

**Acknowledgements**

Former SIG/VIS chair Diane Neal would like to thank SIG/VIS chair Chris Landbeck for his support of this publication.

**Resources Mentioned in the Article**

1. Flickr.com: http://www.flickr.com
2. YouTube: www.youtube.com
4. USA Today: www.usatoday.com
5. Shutterfly: www.shutterfly.com
7. Joost: www.joost.com
8. Hulu: www.hulu.com
11. Fitday: www.fitday.com
12. Weather.com: www.weather.com
13. Information Builders: www.informationbuilders.com
16. Like.com: www.like.com
20. SearchMe: www.searchme.com
21. MSNBC.com: www.msnbc.com
22. Glass Engine: www.philipglass.com/
Resources for Further Reading


Glass Engine. www.philipglass.com


Pew Internet & American Life Project. www.pewinternet.org


