Information Visualization and Keyword Searching

Matt Conner and Melissa Browne

Our paper presents new research on the intersection between new virtual landscapes of knowledge and the learning behavior of the Millennial generation; in a sense the paper is about the very problem of representing information.

Visual Literacy

We begin with a short history of the concept of "visual literacy" and its relevance to libraries. The concept has similarities with "information literacy," another large term in librarianship. Information literacy has been defined to embrace issues of critical thinking, evaluation, creation, and ethical judgment of material (ACRL, 2000). Visual literacy lacks an official accepted definition such as this. Yet, one reasonable place to start is the International Visual Literacy Association whose co-founder, Jack Debes, is credited with inventing the concept of "visual literacy" in 1969. The Association's website defines visual literacy as:

> … a group of vision-competencies a human being can develop by seeing and at the same time having and integrating other sensory experiences. The development of these competencies is fundamental to normal human learning. When developed, they enable a visually literate person to discriminate and interpret the visible actions, objects, symbols, natural or man-made, that he encounters in his environment. Through the creative use of these competencies, he is able to communicate with others. Through the appreciative use of these competencies, he is able to comprehend and enjoy the masterworks of visual communication.[1]

Conner *(Instruction/Reference Librarian)* and
Browne *(Instruction/Reference Librarian)*
University of California, Davis [Davis, CA]

The similarity to the ACRL definition is evident in conceiving of literacy as a process of consumption and creation of ideas. As Anderson (2008) observes, "Visual literacy is about both interpreting images and producing images. We should no longer consider people to be media literate if they can consume but not produce media" (p. 64).

Visual literacy also assumes that, among the senses, vision has a special status in the way in which it is bound with cognition. There is a good deal of tantalizing though speculative research on how this connection is based on anatomical structures of the eye and brain.[2] In sum, the work suggests that the mind operates by unconscious associations between abstract concepts and visual structures so that vision, rather than a form of static input, is an aspect of the thought process. Perceptual psychologist Richard Gregory has described the "'inner logic' of perception in visual problem-solving."[3] To see is to think. Visual literacy so defined and understood has proven most useful in its application to Cultural Studies, an interdisciplinary field that heavily influences English departments, and Art History with its study of images.

Yet, it is not clear just how it applies to librarianship, specifically information literacy instruction. We suggest that perhaps what is relevant is not visual literacy, but instead a new and emerging subfield called "information visualization." This field arises out of a combination of computer science, psychology, and learning theory. (It dates to a particular publication a National Science Foundation report DeFanti, T.A., and M.D. Brown. "Visualization in Scientific Computing." *Computer Graphics 21*(6) (1987).) Researchers seek to represent large amounts of data through visual forms that assist rapid comprehension and insight (Card, MacKinlay, & Shneiderman, 1999). With this guiding principle and a specialized set of tools, the field points towards librarianship and its emphasis on information technology. A summary statement shows how information visualization intersects with the larger goal to foster critical thinking in library instruction and education generally. "Information visualization can help make us smart. Of course, leverage works both ways. It can also make us stupid by misadvised mappings and unworkable user interfaces just as 'chart junk' graphics makes information harder to comprehend."[4]

Library Instruction and Millennial Search Behavior

Information visualization allows us to address a central but unacknowledged problem in library instruction. The databases that have transformed the field are very good at generating output from input—so much so that users sometimes feel swamped with information in one of the common scenarios of the "information age." But while libraries devote great attention and expense to acquiring databases and manipulating their advanced search features, little or no time is given to the generation of input—that is the keywords —that produce results in the first place. Studies indicate that the resulting lacuna in teaching is a serious omission. Students who search the net successfully with few terms or long natural language search strings struggle with navigating scholarly research tools, often making no use of subject headings and controlled vocabularies. "[P]atrons rarely utilize correct and complete subject terms and they retrieve zero results in almost half of their searches." (Antell & Jie, 2008, p. 68) This phenomenon has been widely observed, as stated by Antell and Jie (2008):

> American libraries coined the phrase "search fatigue" to describe the "feeling of frustration and dissatisfaction" that users suffer when they spend hours looking in databases for information that they know ought to be there, but that they cannot find. According to the author, Jeffrey Beall, 'The chief cause of search fatigue is a reliance on keyword searching as opposed to controlled vocabulary searching" (p. 68).

Improvement at the beginning of the search process with the generation of search statements promises to have vast ramifications for the students, the quality of their work, and the effectiveness of libraries in the learning process.

Study Purpose and Hypothesis

In a new effort to apply information visualization techniques to facilitate generating keywords in the context of library instruction, this paper reports on preliminary results from a study comparing the impact of three information visualization-based tools on the search practices of undergraduate subjects. We hypothesize that the use of information visualization techniques improves students' abilities to conceptualize topics and generate terms for online academic research.

Materials and Methods

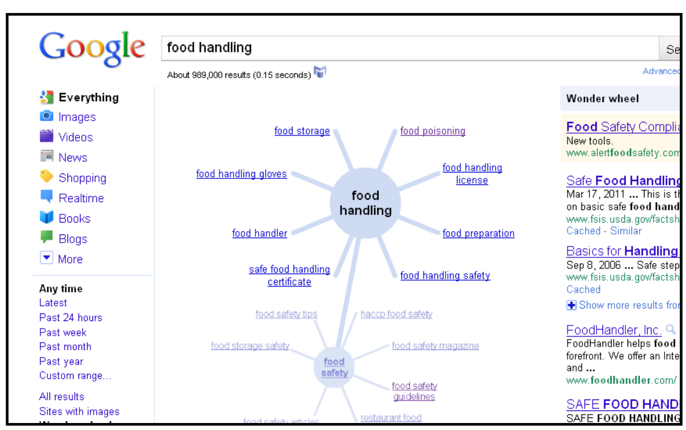
The study has been approved by the campus Institutional Review Board (IRB) and will gather data on 60 undergraduate subjects at the University of California, Davis, randomly assigned to one of four test conditions. The preliminary results reported here were collected from 19 students.

1. Control: Subjects conduct searches in Academic Search Complete according to their usual method without instruction.

2. Keyword Matrix: Subjects use a matrix to generate a range of search terms. Given a topic, students supply at least two terms that are more general and at least two that are more specific. With the topic conceptualized in this fashion, the subject will (hopefully) proceed to utilize the terms to search Academic Search Complete in a systematic way, moving dialectically between broader and narrower terms to arrive at results.

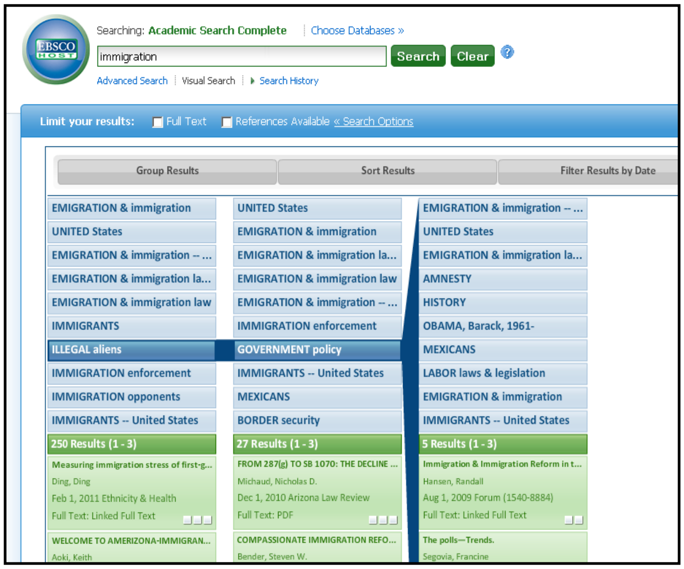
General:	Ethics Scientific research	Human-animal relationships
Topic: animal rights		
Specific:	Lab animals Animal experiments	PETA Cosmetic testing

3. Google Wonder Wheel: Google has an option to view search results in a concept map-like display, whereby a query generates suggested terms arranged radially around it. Clicking on a suggested term repeats the process, so the chosen term becomes the center of a new circle with its own related concepts. The search proceeds with one circle linked to another in this manner. Subjects write down terms that appeal to them and then use them to search Academic Search Complete.



4. Visual Search: Academic Search Complete has its own visual search tool that represents suggested subject

terms in rows and columns. Clicking on one of the search terms generates a list of sources.



Student volunteers from a variety of majors met individually with one or the other of the co-investigators, and responded to a series of preliminary questions to determine the scope of their prior research experience. One difficulty was eliminating the bias of prior subject knowledge. To address this, students were divided into broad categories of humanities, sciences, and social sciences, each designated with a set of representative topics. Subjects were given topics from an area other than his or her own. While not foolproof, this method corrected in some degree for knowledge accrued in a student's major.

All participants completed a pretest search on an unfamiliar topic in Academic Search Complete to allow the investigators to observe their normal search behaviors. Control subjects then performed one additional search on a second topic. Students in the three experimental conditions (Keyword Matrix, Wonder Wheel, Visual Search) received brief instruction on the appropriate information visualization method, and were then asked to apply it to generate keywords and identify search terms for two additional searches in Academic Search Complete. All students were told to search until they identified 2-3 article citations that they would consider relevant if they were writing a paper on that topic.

Data Collection & Analysis

Students wore headphones with a microphone, and narrated their thoughts as they searched. Adobe Captivate 3 software was used to record each subject's audio and screen captures of their online searches. Each subject was also asked to complete a written summary, comparing the experimental technique to their standard practice. Data from the sessions were transcribed and entered into an Excel spreadsheet, and the quantitative data will be analyzed with assistance from a statistician. Notes from each session were also uploaded to Saturate, http://www.saturateapp.com, a free, collaborative tool for qualitative data analysis, and a coding scheme was applied to identify repeated themes in the data. Of particular interest were those that indicated an effort to organize the material.

To analyze the impact of visual search methods, we identified two major quantitative markers: search time and numbers of searches. Search time was recorded as the time spent using Academic Search Complete to identify relevant citations on a given topic. Data was also taken on the amount of time utilizing the Keyword Matrix (Condition 2) or Google's Wonder Wheel (Condition 3) to generate terms before proceeding to search in Academic Search Complete. We assumed that the length of time searching was correlated inversely to the efficiency of the search: A shorter search indicates more efficiency. But assessing the value of the numbers of searches performed for a given trial was more complex.

Results/Discussion

At this writing, we have gathered data on 19 subjects, with results from 4-5 subjects in each condition. For both length of time and number of searches attempted for the various conditions, we conducted an analysis of variance with a t-test to account for the small number of subjects in our samples. Specifically we compared the mean times and numbers of searches and tested to see if there were statistically significant differences among our test conditions. With a 95 percent confidence interval, our analysis indicates that there is no difference among the test conditions. Stem and leaf plot analysis of our frequency codes has not been subjected to statistical analysis, but inspection does not indicate obvious patterns. In particular, there are no obvious differences correlated to our conditions, especially not with those codes indicating more systematic searching behavior for the visual analysis conditions.

Our data thus far indicate some directions for the analysis. First, interpreting our data directly, one might conclude that information visualization in the forms that we have tested has no effect on search effectiveness. Or, secondly, one might suppose that the testing assumptions have been thrown into question: for example, perhaps search efficiency does not correlate with less time spent searching or a more extensive arsenal of search terms. The qualitative evidence and observations we have gathered thus far tend to point towards this second possibility. A strong trend we have noted across our study conditions is that rather than relying on keywords as the literature suggests, the students seem driven by their search results, sometimes losing sight of their original topic. The influence of the internet, therefore, may not take the form of using disconnected keywords but of following links. Subject behavior is not unlike that of search engine spiders which crawl about the web haphazardly from one link to another. Thus, a characterization of internet surfing exclusively in terms of random keywords is simplistic and so are the conclusions that result about how the internet influences research behavior in libraries. A better model for student searching appears to be a blend of link-crawling combined with brainstorming and other nascent paper-composing behaviors.

Our initial results also indicate an enthusiasm for tools that help students generate terms e.g., Google's Wonder Wheel and EBSCO's Visual Search. We hypothesize that they appeal precisely because they are integrated with the search process. This criteria should favor the Visual Search, which is closely integrated with database search results. Preliminary reactions to the Keyword Matrix have been more ambivalent. The value of the Keyword Matrix may lie in providing students with a framework of concepts that works in the background providing guidance indirectly, even if the subjects do not use its terms directly. But studying this effect will be difficult.

While information visualization tools appear to impact students' initial choice of terms, so far we have noted that students tend to modify their subsequent searches based on content they encounter in their database results, personal identification with a topic, and questions that surface as their search progresses, reflecting the complex, iterative and serendipitous nature of information research. Our work thus far suggests that strategic use of information visualization theory by teachers may positively influence student attitudes toward online searching. The technology itself, by making conceptual relationships explicit, may also encourage the characteristics of expert learning (Bransford, Brown, & Cocking, 1999, pp. 31-33). A detailed analysis of our full data set will offer additional insights into student search behaviors.

Endnotes

1. (Avgerinou) (Anderson, 2008, p. 64) Avgerinou enlarges on the points of convergence among Visual Learning Theorists.

   A visual language exists.

   - Visual language parallels verbal language.
   - VL is a cognitive ability but also draws on the affective domain.
   - The terms 'ability,' 'skill,' and 'competency' have been invariably and interchangeably used to describe VL.
   - The VL skills have been specified as (a) to read/decode/interpret visual statements, and (b) to write/encode/create visual statements.
   - The VL skills are (a) learnable, (b) teachable,(c)capable of development and improvement.The VL skills are not isolated from other sensory skills.
   - Visual communication, visual thinking, and visual learning are inextricably linked to VL.
   - VL has accepted and incorporated theoretical contributions from other disciplines.
   - VL's main focus is intentional communication in an instructional context.

   (Avgerinou, 2009, p. 29)

2 (Barry, 1997, pp. 10-39) (Burmark, 2008)

3 (Barry, 1997, p. 8) Bamford describes a "visual syntax" of: scale, dimension, motion, boldness, arrangement, framing, depth, dimension, colour, light, shadow, flow of movement." (Bamford & Adobe Systems, 2003, p. 3)

4 (Card, et al., 1999, p. 34) Card expounds further on the rationale for information visualization in terms of anatomical structure: "Even at the visual cortex, perception appears to rely on spatially distributed parallel construction processes in a topography that corresponds to the real physical world. The central conjecture behind the approach to text visualization described here is that the same spatial perceptual mechanisms that operate on the real world will respond to a synthetic one, if analogous cues are present and suitably integrated. The bottleneck in the human processing and understanding of information in large amounts of text can be overcome if the text is spatialized in a manner that takes advantage of common powers of perception." (Card, et al., 1999, p. 443) See also (Chen, 2004, p. 23) (Wright, 1999)

Works Cited

ACRL, (2000, January 18). Information literacy competency standards for higher education. Retrieved from http://www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm#ildef

Anderson, B. (2008). Information literacies. *Behavioral and Social Sciences Librarian, 26*(2), 63.

Antell, K., & Jie, H. (2008). Subject searching success transaction logs, patron perceptions, and implications for library instruction. *Reference & User Services Quarterly 48*(1), 68-76.

Avgerinou, M. D. "What is visual literacy?" Retrieved from http://www.ivla.org/org_what_vis_lit.htm#definition

Avgerinou, M. D. (2009). Reviewing visual literacy in the "bain d' images" era. *Tech Trends, 53*(2), 28-34.

Bamford, A., & Adobe Systems. (2003). The visual literacy white paper. Uxbridge, England: Adobe Systems.

Barry, A. M. S. (1997). *Visual intelligence: Perception, image, and manipulation in visual communication*. Albany: State University of New York Press.

Bransford, J., Brown, A., & Cocking, R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington D.C.: National Academy Press.

Burmark, L. (2008). Visual literacy: What you get is what you see. In N. Frey & D. Fisher (Eds.), *Teaching visual literacy: Using comic books, graphic novels, anime, cartoons and more to develop comprehension and thinking skills*. Thousand Oaks, CA: Corwin Press.

Card, S. K., MacKinlay, J. D., & Shneiderman, B. (1999). *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers.

Chen, C. (2004). *Information visualization: Beyond the horizon*. London: Springer.

Wright, W. (1999). Research report: Information animation applications in the capital markets. In S. K. Card, J. D. MacKinlay & B. Shneiderman (Eds.), *Readings in information visualization: Using vision to think*. San Francisco: Morgan Kaufmann Publishers, Inc.