An Arbitrage Opportunity for Image Search and Retrieval

by Ray Uzwyshyn

Visual Representation, Search and Retrieval: Ways of Seeing

A new paradigm for image search metadata collection is emerging exemplified by the Human Computation School's application of gaming principles to information science search challenges. In parallel, a suite of Web 2.0 interface applications for visual search have recently appeared opening new interactive possibilities and visual metaphors for navigation. This article briefly introduces this paradigm shift and then looks critically toward wider innovation with an eye on fresh territory. Arbitraging differing methodologies opens new visual search possibilities, as affordances and differences between models present opportunities to leverage inefficiencies in one model with efficiencies of the other. This article capitalizes on such inequities, prescriptively suggesting a synergistic path for combining new image-retrieval metadata methodologies with new front-end visual search directions for future application innovation.

Human Computation and Image Metadata

Perhaps a good place to begin this discussion is with *Google's Image Search* [1], which claims to be the web's most comprehensive. The computational challenge with regard to visual search and images has been to improve relevancy, precision and the quality of textual matching when searching any large group of images. How does one provide high quality metadata for images that will optimize these parameters?

Searching on words such as *dog*, h*orse* or *stock market* usually brings up a good representation of images, some relevant, others less so. However,

FIGURE 1. Google Image Search (http://images.google.com)

challenges become more apparent as the level of keyword abstraction or ambiguity increases. Take, for example, the abstractions *bravery*, *intelligence* or c*ourage* or cognates like "intelligent dog" or "courageous lion."

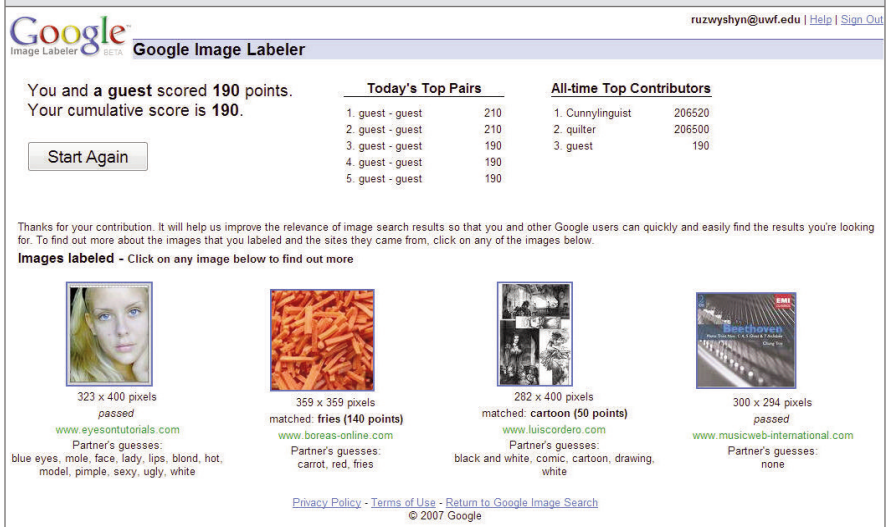
Larger scale computational image search methodologies have traditionally worked through algorithms that search and pair metadata (alt tags, keywords, file metatags, surrounding description) or, more commonly, text strings with various image file types. Because a computer has no common sense and cannot tell whether the surrounding description is appropriate, relevancy decreases as the precision needed increases.

A fresh approach to this metadata challenge is outlined in recent work by Luis von Ahn. Von Ahn proposes [2, 3] to capitalize on the efficiencies of human processing cycles through games to help solve traditionally intractable problems. By appropriating an online gaming methodology, two randomly paired participants are simultaneously and separately shown the same image and asked to propose matches. The recorded game play and results provide a new data-gathering mechanism to more accurately label or

Ray Uzwyshyn is head of digital and learning technologies at University of West Florida Libraries. He can be reached by email at ruzwyshyn<at>uwf.edu

NEXT PAGE >

FIGURE 2. *Google Image Labeler Beta* (http://images.google.com/imagelabeler/) [4]



provide reliable image metadata (Figure 2). Combining the gathered metadata with statistical methodologies opens a door to creating better databases of visual search image data.

The covert harnessing of human processing cycles and common sense reasoning through overt gaming methodology is an interesting model that could be further exploited to meet more difficult challenges such as providing polyphonic metadata for images, adequate metadata for video and film or accurate labeling for sections of images. The wider idea is to leverage intrinsic human strengths with computer affordances and bring these into efficient and natural synergy. The deeper innovation is into the "medium specificity" [5] or the unique possibilities opened by computational media in synergy with natural human strengths and inclinations. The insight here involves "object relations" [6] between human and computer – or the dynamic and evolving process of augmenting cognition in the wider ecology between human and computer. By reexamining this relationship, new solutions to present-day computational challenges are enabled. There is room for further work here with von Ahn's practical innovations most cogently displayed in his online *Games with a Purpose Project* [7]. Von Ahn's trajectory actualizes earlier more speculative endeavors within a Web 2.0 framework. Of note is the work of two earlier heterodox artificial intelligence researchers, Push Singh [8, 9] (see the Open Mind Common Sense Site [10, 11]) and Christopher Mckinstry [12]. Their attempts to harness common sense reasoning are worth revisiting for further reflection and possibility. Other pioneering efforts include those by Douglas Lenat [13] and Marvin Minsky [14].

New Visual Search Interface Metaphors

Traditionally, visual image search on the web has been presented through an interface and photographic contact sheet metaphor. For example, in a Google image result set, 20 thumbnail images are presented on a single page in a 4x5 (20 image/page) grid with links to larger images (Figure 3).

FIGURE 3. *Google Image Search*: Keyword "Kennedy"

The visual metaphor used for presentation is the photographic contact sheet. By clicking through a numbered list, one clicks through contact sheet pages. Clearly, for the result set of 20,300,000 pages produced by keyword "Kennedy" (Figure 4), this presentation is hugely inefficient for humans, yet it is the dominant interface metaphor in practice for image-search navigation.

FIGURE 4. Pages 1-16 of 20,300,000 Pages for Keyword "Kennedy"

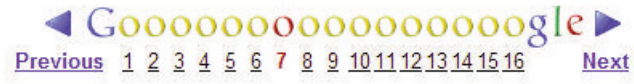


FIGURE 5. *Cooliris*: Image Wall Browser and Media Discovery Tool (www.cooliris.com/)

Recently, various online applications have emerged to challenge this method and metaphor with new, more interactive and agile visual navigation possibilities enabled through AJAX (Asynchronous JavaScript and XML) and FLASH (.swf) based technologies that present other metaphors for display and navigation. For example, *Cooliris* [15] takes image search's visual display into an interactive horizontal 3D wall methodology that can be scrolled or fast-forwarded with commands like those available on a film reel or media controller, such as "play," "fast forward" or "rewind."

The cinematic and interactive image wall methodology lends itself to searching and retrieving an image from a large number of images more humanly. Interestingly, the antecedents for the emerging Cooliris School of applications have been in place for a number of years. Similar to von Ahn, the wider broadband web infrastructure and improved application environment of Web 2.0 have only recently made these ideas practicable. There is room for a recasting of historical interface possibilities that in the 90s and early millennia were only available in R&D environments for wider dissemination. See, for instance, work by Card, Mackinlay, Schneiderman, Rao and others [16, 17, 18, 19].

Methodological Synthesis and Arbitrage: Common Sense Metadata and New Interface Metaphors

Looking back at the two examples outlined (human computation and improved interface metaphors), clearly both offer better models for visual image search. The first presents new metadata possibilities by harvesting common sense data for images through games. The second model improves front-end interface metaphors. What is needed is an arbitrage and synthesis of paradigms. Because of their overwhelming attention to the front end, applications exemplified by *Cooliris* and this new metaphor/interface school lack strong attention to metadata application or, as yet, integration of innovative metadata methodologies to improve search/retrieval. These applications simply overlay other search engines' metadata or map antecedent methods. Similarly, while the *Google Image Labeler* and the Human Computation School provide new avenues for better metadata collection, they rely heavily on traditional presentation and do not as yet utilize or attempt integration with new interface possibilities. Capitalizing on inefficiencies in both models through a leveraging of the innovation in each results in a new synthesis. Further opportunity may also be opened through a similar methodological arbitrage of prevailing disparate paradigms.

In the ever-evolving human/computation relationship, the larger keyword is *human*. In harvesting these new vintages of metadata possibilities, it is increasingly important to beware of placing new wine in old wineskins. New container metaphors are available. A new synthesis, taking affordances into account, will provide a better lens through which to look back at both schools of applications. This new foundation may also allow a reexamination of the present dominant text search metaphor – the long scrolling result list. A more robust point of departure is also needed for search applications investigating the more uncharted territory of digital film or video. Beginning to integrate these newer paradigms will provide a better window for visual image search. Opportunities outlined present fertile territory for the future of media-based information retrieval. ■

RESOURCES on next page

TOP OF ARTICLE
< PREVIOUS PAGE
NEXT PAGE >

Resources

Visual Image Search, Metadata and Common Sense Reasoning

[1] *Google Image Search*: http://images.google.com.

[2] von Ahn, Luis. (July 26, 2006). *Google Tech Talks: Human computation*. Retrieved March 11, 2009, from http://video.google.com/videoplay?docid=-8246463980976635143.

[3] von Ahn, Luis. (December 7, 2005). *Human Computation*. (Doctoral dissertation. Carnegie Mellon University.) Retrieved April 27, 2009, from http://reports-archive.adm.cs.cmu.edu/anon/2005/CMU-CS-05-193.pdf

[4] *Google Image Labeler Beta*: http://images.google.com/imagelabeler/.

[5] This term – *medium specificity* – is appropriated from art history/media theory and used in a digital framework here. It may not be as well known in information science contexts. For further explanation, see http://en.wikipedia.org/wiki/Medium_specificity and http://csmt.uchicago.edu/glossary2004/specificity.htm

[6] The term – *object relations* – as used here is from psychoanalytic theory. As with the term *medium specificity*, it may not be as well known in information science contexts. For further context see http://en.wikipedia.org/wiki/Object_relations and www.objectrelations.org/introduction.htm.

[7] *Games with Purpose*: www.gwap.com.

[8] Singh, P. (2006). *MIT publications list*. Retrieved March 11, 2009, from web.media.mit.edu/~push/#Publications.

[9] Singh, P. (June 2005). *Em-One: An architecture for reflective commonsense thinking*. (Doctoral dissertation. Massachusetts Institute of Technology). Retrieved April 21, 2009, from http://web.media.mit.edu/~push/#Publications.

[10] *Open Mind MIT Common Sense Database*: http://openmind.media.mit.edu/

[11] *Open Mind Initiative*: http://openmind.org

[12] Chris Mckinstry. (April 15, 2009). *Wikipedia*. Retrieved April 21, 2009, from http://en.wikipedia.org/wiki/Chris_McKinstry. [Includes Mindpixel links.]

[13] *Cyccorp*: www.cyc.com. [Douglas Lenat's pioneering common sense website.]

[14] *Marvin Minsky*: web.media.mit.edu/~minsky/

Image Interface Search Possibilities

[15] *Cooliris: Website and application download*: www.cooliris.com/

[16] *Ramama Rao's information flow*: www.ramanarao.com/. [Website with links to Rao's publications]

[17] Robertson, G, Mackinlay, J.D., & Card, S.K. (1991, June). The perspective wall: Detail and context smoothly integrated. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New Orleans* (pp. 173-176). New York: ACM. Retrieved March 11, 2009, from www.infovis-wiki.net/index.php/Perspective_Wall. [Shows an example of perspective walls]

[18] Marc Schmidt, C. Expressing information [PowerPoint presentation]. Retrieved April 23, 2009, from www.christianmarcschmidt.com/NYU2007/components/071023_presentation.pdf. [Includes information on the Timewall [Slide 40] and other pioneering visualization methodologies]

[19] Shneiderman, B., Mackinlay, J., & Card, S. *Readings in information visualization: Using vision to think*. New York: Morgan Kaufman, 1999.