Arbitrage Opportunities for Image Search: Changing Metaphors, Game Metadata and Gathering Common Sense

Ray Uzwyshyn, Ph.D. MLIS
Head, Digital and Learning Technologies
University of West Florida Libraries

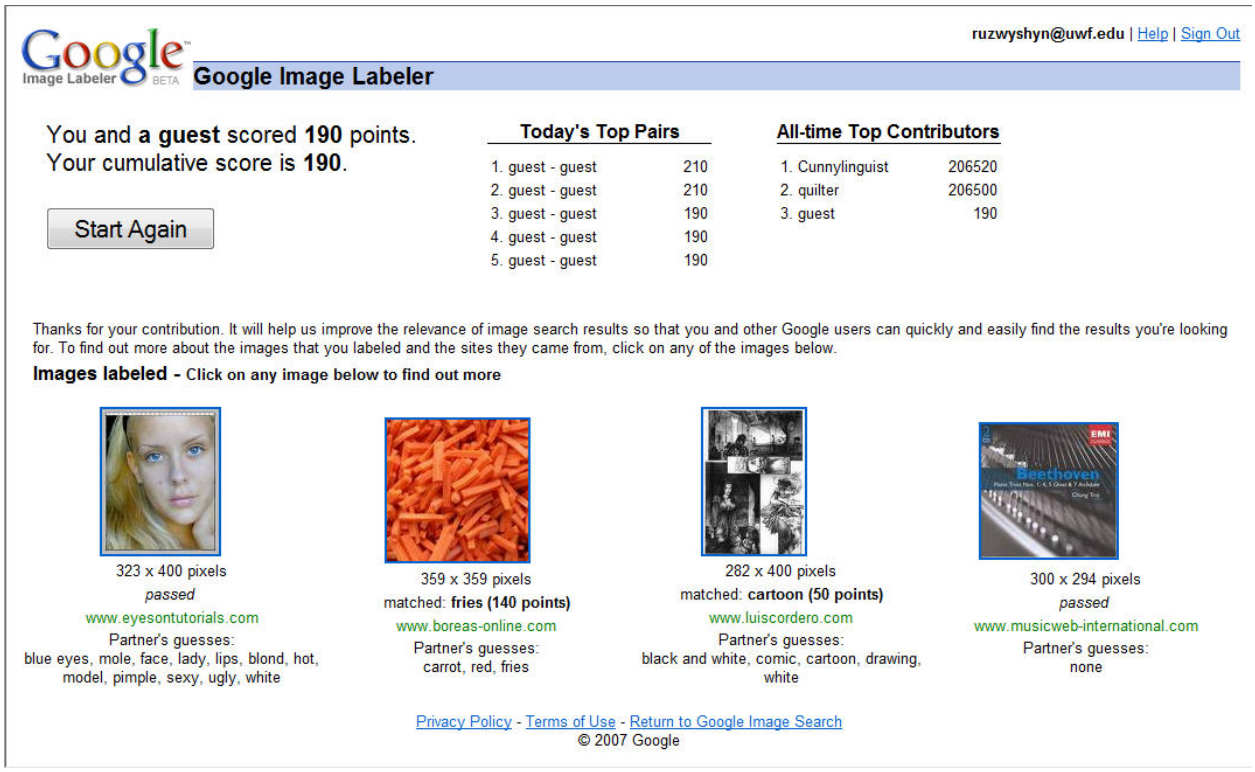
Cooliris Image and Media Browser (http://cooliris.com)

Recently, a new paradigm for image search metadata collection is emerging exemplified by the Human Computation School's application of agonistic principles to historical 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 towards wider innovation with an eye on fresh territory. Arbitraging differing methodologies opens new visual search possibilities. Affordances and differences between models present opportunities to leverage inefficiencies in one model with efficiencies of the other. This article capitalizes on inequities, prescriptively suggesting a synergistic path for combining new image retrieval metadata methodologies with new front end visual search directions for future application innovation.

I - Human Computation and Image Metadata

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

Perhaps a good place to begin this discussion is with Google's Image Search which claims to be the Web's most comprehensive image search. The historical computational challenge with regards to visual search and images has been one of relevancy, precision and textual matching with any larger group of images. How does one provide high quality metadata for images with relevancy, ranking and precision? Taking words such as 'dog', 'horse' or 'stock market' brings up a good representation of images, some relevant, others less so. Challenges become apparent as the level of keyword abstraction or ambiguity increases. Take for example, the abstractions 'bravery','intelligence' or 'courage' or cognates like ' intelligent dog' or 'courageous lion'. Traditionally, larger scale computational image search methodologies have 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 precision needed increases. A fresh approach to this historical metadata challenge is outlined in recent work by Luis von Ahn. Von Ahn proposes 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 provide reliable image metadata. Combining the gathered metadata with statistical methodologies opens a door to creating better databases of visual search image data.



.

Google Image Labeler Beta (http://images.google.com/imagelabeler/)

The covert harnessing of human processing cycles and common sense reasoning through overt gaming methodology is an interesting model which could also be further exploited to more difficult area challenges (i.e. polyphonic metadata for images, beginning to provide adequate metadata for video and film, accurate labeling sections of images). The wider idea is to leverage intrinsic human strengths with computer affordances and put these into efficient and natural synergy. The deeper insight is into medium specificity and object relations between human and computer. There is room for further work here with von Ahn's practical innovations most cogently displayed in his online Games with a Purpose Project (Gwap.com). In a sense, von Ahn's trajectory actualizes earlier more speculative endeavors in a Web 2.0 framework. To note, two earlier heterodox artificial intelligence researchers, Push Singh's and Christopher Mckinstry's earlier attempts to harness common sense reasoning, are also worth revisiting for further reflection and possibility (see bibliographies) .

II - 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, twenty thumbnail images are presented on a single page in a 4x5 (20 image/page) grid with links to larger images.

Google Image Search: Keyword "Kennedy"

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

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

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 to present other metaphors for display and navigation. For example, Cooliris (http://cooliris.com) takes image search's visual display into an interactive horizontal 3D wall methodology which can be scrolled or fastforwarded similar to a film reel and media controller (play, fast forward/reverse, etc.) set of controls.

Cooliris Image Wall Browser and Media Discovery Tool

The cinematic and interactive image wall methodology lends itself more humanly to searching and retrieving an image from a large number of images. Interestingly, the historical antecedents for the emerging Cooliris School of applications have been in place for a number of years but similar to von Ahn, the wider broadband web infrastructure and easier application of Web 2.0 have only recently made these practicable. To note, there is room for a recasting of historical interface possibility for wider dissemination which in the 90's and early millennia was only available in R&D environments. (See Bibliography, Card, Mackinley, Schneiderman and Rao).

III - Methodological Synthesis and Arbitrage

Looking back at the two examples outlined, clearly both offer better models for visual image search. The first presents new human computational metadata possibilities for harvesting common sense data for images through games. The second model improves front end interface metaphors. What is needed is a synthesis of paradigms. Interestingly, 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, these later rely heavily on traditional presentation and do not as yet utilize or attempt integration with new interface possibilities.

In the ever evolving human/computation relationship, the larger keyword here is 'human'. In harvesting these new vintages of metadata possibilities, it is increasingly important to be aware of placing new wine in old flasks. New container metaphors are available. A new synthesis, taking affordances into account, will provide a retrospective optic. This new foundation may also allow a reexamination of the present dominant text search metaphor (i.e. the long scrolling result list). A more robust point of departure is also needed for search applications investigating 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.

Bibliography and Links

Visual Image Search, Metadata and Common Sense Reasoning

Cyccorp. http://www.cyc.com/ Douglas Lenat Historical Common sense Website. Retrieved March 11, 2009.

Google Image Labeler Beta. http://images.google.com/imagelabeler/ Retrieved March 11, 2009.

Mckinstry, Chris http://en.wikipedia.org/wiki/Chris_McKinstry . Historical Mindpixel Links. Retrieved March 11, 2009.

Open Mind MIT Common Sense Database. http://commons.media.mit.edu/ Retrieved March 11th 2009.

Singh, Push. MIT Publications List. http://web.media.mit.edu/~push/#Publications Retrieved March 11, 2009

Singh, Pushpinder. *Em-One: An Architecture for Reflective Commonsense Thinking*. (Dissertation) Massachusetts Institute of Technology. Dept of Electrical Engineering and Computer Science. 2005.

Shirky, Clay. *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin, 2008.

von Ahn, Luis. Google Tech Talk on Human Computation. http://video.google.com/videoplay?docid=-8246463980976635143 Retrieved March 11, 2009.

von Ahn, Luis. *Human Computation* (Dissertation). School of Computer Science. Carnegie Mellon University. December 7, 2005.

von, Ahn, Luis. Games with a Purpose Site. http://www.Gwap.com. Retrieved March 11, 2009.

New and Historical Image Interface Search Possibilities

Ramana Rao. Website and Links to Publications. http://www.ramanarao.com/ Retrieved March 11, 2009.

Stuart Card, Jock Mackinlay, George Robertson. Historical Example of Perspective Walls. http://www.infovis-wiki.net/index.php/Perspective_Wall Retrieved March 11, 2009.

Card, Stuart et al. "The Perspective Wall: Detail and Context Smoothly integrated". Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New Orleans: 1991, pp. 173-176.

Cooliris Website and Application Download. http://www.cooliris.com/ Retrieved March 11, 2009.

Timewall. http://demo.labs.businessobjects.com/VizServer/demos/sample-MovieBoxOffice.html Historical Visualization Tool Demo. Retrieved March 18, 2009.

Shneiderman, Ben. Mackinlay, Jock and Card Stuart. *Readings in Information Visualization: Using Vision to Think*. New York: Morgan Kaufman, 1999.