Aria: An Agent for Annotating and Retrieving Images

Aria is an interface agent designed to assist users by proactively looking for opportunities for image annotation and retrieval. While it doesn't completely automate the image annotation and retrieval process, Aria dramatically reduces user interface overhead, which can lead to better-annotated image libraries and fewer missed opportunities for image use.

Henry Lieberman
MIT Media Lab

Elizabeth Rosenzweig
Eastman Kodak

Push Singh
MIT Media Lab

George Eastman's original advertising slogan for Eastman Kodak was "You push the button, we do the rest." Eastman sought to convince consumers that the technology of photography, including Kodak's products and services, would act as their agent, recording their memories. Initially, photography was a highly technical art that didn't enjoy widespread adoption until users began to believe that its mechanical details wouldn't overwhelm them. We tell this story not because the goal of making technical innovations user-friendly is unique to Kodak, but because the idea that the user should express their wishes, then leave "the rest" of the process to an agent, human or otherwise, is a laudable goal for any technology.

Modern photography, especially digital photography, has come a long way, but the process of making and using photographs still requires more effort than it should. Organizing and retrieving images stored in a shoe box—or its digital equivalent—is so tedious that people avoid doing it, and many photographs are rarely seen again. Using software agents rather than human labor can help reduce some of this tedium. Currently, text labels can be used to annotate images and store them in a relational database to retrieve later using keywords. But users are unlikely to expend substantial effort to classify and categorize images in this way in the hopes of facilitating future retrieval. Testing shows that users simply won't expend that much effort. Furthermore, retrieval requires dealing with a search engine or other application that imposes additional overhead on the process, even if only in terms of starting and exiting the application to enter the keywords. Because of this overhead, opportunities to use images are often overlooked or ignored.

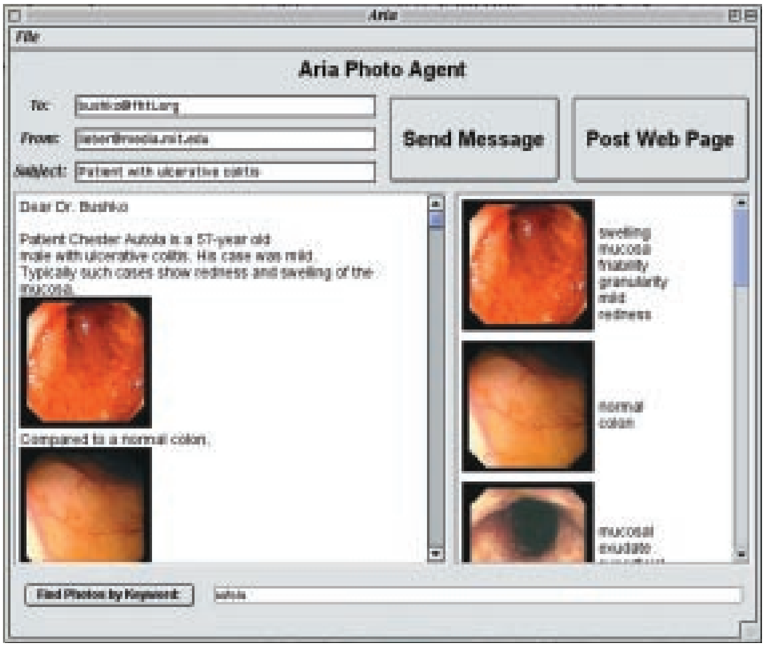
In the future, automated image analysis might be able to identify people, places, and things in a photograph and annotate the images. Although researchers have made considerable progress in this area,[1-3] we are still far from being able to rely on this kind of approach. Even if images can be roughly interpreted automatically, many salient features exist only in the user's mind. Indexing the image requires a way to communicate these features to the machine.

We've designed a user interface agent to facilitate—rather than fully automate—the textual annotation and retrieval process. The role of the agent lies not so much in automatically performing the annotation and retrieval but in detecting opportunities for performing these functions and alerting the user to those opportunities. The agent can also make it as easy as possible for the user to complete the operations when appropriate.

NO PICTURE IS AN ISLAND

Whether taken by consumers to record their family memories or by professionals in the course of a work day, pictures are usually part of a story that might

0018-9162/01/$10.00 © 2001 IEEE



To: bushko@fhtl.org

From: lieber@media.mit.edu

Subject: Patient with ulcerative colitis

Dear Dr. Bushko

Patient Chester Autola is a 57-year old male with ulcerative colitis. His case was mild. Typically such cases show redness and swelling of the mucosa.

Compared to a normal colon.

swelling mucosal friability granularity mild redness

normal colon

mucosal exudate

Find Photos by Keyword: autola

Figure 1. Aria screen layout. At the bottom of the text editor, Aria displays a retrieval term taken from the text surrounding the editor's cursor. To the right of the text editor is a column of retrieved images that Aria dynamically updates. At the right of each image, the application displays a list of annotation keywords pertaining to each image.

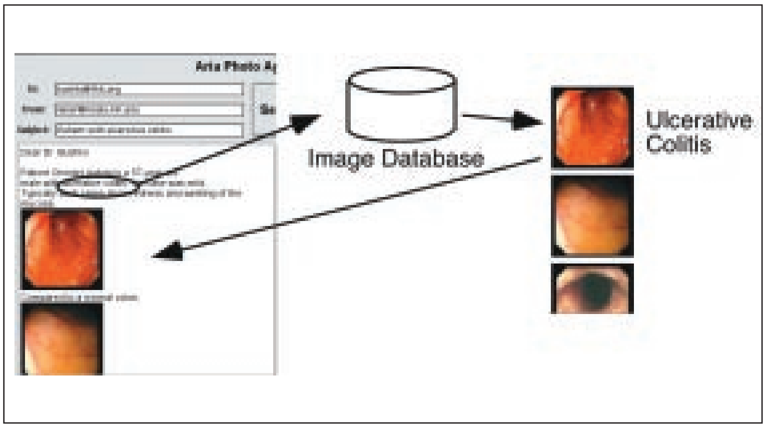


Figure 2. Image retrieval in Aria. The box below the text editor pane continuously displays keywords. Aria uses the keywords to query the image database and displays a ranked list of pictures in the column to the right in the pane.

appear in a written document, in an e-mail message, or on a Web page. Currently, software does not support any explicit connection between the applications in which materials relevant to the story might appear and the applications that store and organize the images pertaining to them. Thus, using photographs might involve several applications, while the task of integrating the story is left to the user.

Imagine a scenario in which a doctor needs to use images in a medical image library. Some of those images might be pictures or test results from individual patients, and some might come from medical reference sources such as research journals or textbooks. Searching, viewing, transmitting, and archiving the images would require the doctor to to do a lot of work using several different applications.

Our approach integrates image annotation, retrieval, and use into a single application, eliminating the confusing context-switch that using separate applications imposes. Much of what we call problem-solving intelligence is really the ability to identify what is relevant and important in a context and to make that knowledge available just in time.[4] An integrated application makes the appropriate context for relating text and images available and conveniently accessible.

For example, when editing e-mail messages, typing text descriptions often sets up a semantic context in which retrieving relevant pictures would be appropriate. Seeing the pictures sets up a context for which some textual descriptions might apply, which provides an opportunity for annotation.

ARIA: A PROTOTYPE AGENT

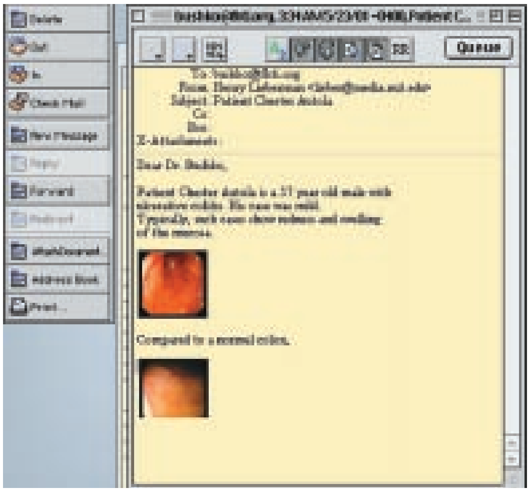
We built the annotation and retrieval integration agent—Aria—as a prototype application to test some of these ideas. The initial implementation consists of a standard Java Swing text editor coupled to a pane containing a custom-built image retrieval and annotation application. Figure 1 shows Aria's screen configuration.

Much like Remembrance[5] and Letizia,[6] both of which observe certain kinds of user behavior, Aria runs continuously and observes the user's typing. Aria analyzes the agent's input to extract keywords from the context surrounding the text cursor. We currently use a straightforward approach of common information-extraction heuristics[7] similar to those used by Web search engines to perform the text analysis, but we are experimenting with other methods. Aria continuously displays keywords in the neighborhood of the cursor in the box below the text editor pane. The application uses the extracted keywords to query the image database and displays a list of pictures ranked in order of relevance in the column just to the right in the text pane, as Figure 2 shows. Aria recomputes this list at every keystroke.

One scenario might look like this: A doctor starts typing in the editor, "Dear Dr. Bushko: Patient Chester Autola is a 51-year-old male with ulcerative colitis." As the doctor types, Aria continually scans the text surrounding the cursor and extracts the keywords "Chester," "Autola," "male," "ulcerative," and "colitis." The column to the right of the text editor displays a sequence of images, each possibly annotated with a set of keywords ranked in descending order of relevance to the text surrounding the cursor. Let's assume that an image illustrating the appearance of the patient's colon was previously annotated with some of these terms. That image would appear as the topmost image in the column without any explicit action on the part of the user other than typing the message. A single drag-and-drop action would insert the picture into the editor. If the desired image does not appear immediately, the user can scroll through the list until a suitable image appears or call up a dialog box to load other images. Even if the search requires using one of these alternatives, it still saves some interaction compared to a conventional approach.

When the doctor finishes making his report, he can press a single button to either send it as an e-mail message—shown in Figure 3—or post the message to a Web page, the two most common scenarios for using images.

We are experimenting with using other kinds of information—like temporal references—to aid image retrieval. If the user types "I examined the patient on 21 May 2001", the system compares that date to the dates time-stamped on every picture, and retrieves pictures having that date. We're planning to include a large vocabulary of time references, including relative dates ("nine months ago") and intervals ("about").



Dear Dr. Budkin,

Patient Charles Antola is a 57 year old male with ulcerative colitis. His case was mild.
Typically, such cases show redness and swelling of the mucosa.

Compared to a normal colon.

Figure 3. E-mail message produced by Aria. Pressing a single button sends the message, including the relevant images.

ANNOTATING IMAGES

But how do the annotations get there in the first place? Let's continue the earlier scenario. The doctor continues the letter by typing, "Typically, such cases show redness and swelling of the mucosa." He would like to include a reference picture that illustrates this condition to compare with the image from this patient. In addition to images from patient histories, the database also includes images from general medical information sources, such as hospital records and journal articles. Any images annotated with "redness," "swelling," and "mucosa," would have popped up immediately as the doctor typed the words. However, in this case, no picture in the database happens to be included in this annotation. So the doctor scrolls through the available images that do mention ulcerative colitis, sees one that illustrates the point he is trying to make, and drags it into his letter.

Aria automatically enters the keywords "redness," "swelling," and "mucosa," attaches them to the corresponding picture, and writes the annotations on the image database. The next time someone types those keywords, Aria will consider the picture to be a candidate for retrieval. This technique uses the text already existing in the message to annotate the images so that retrieval will be easier next time. As Figure 4 shows, Aria extracts keywords from the surrounding text and uses them to annotate the image.

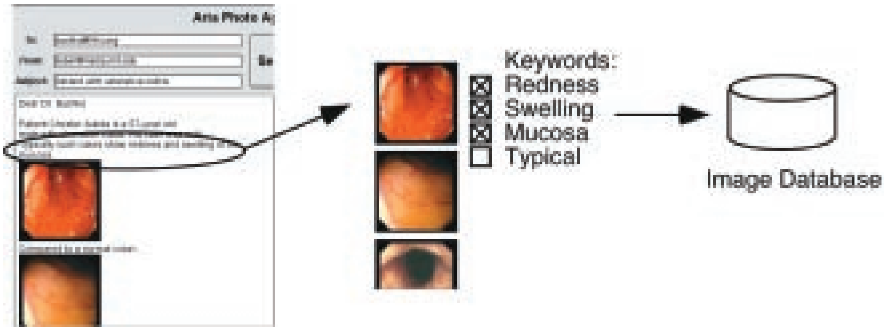


Figure 4. Image annotation. Aria uses keywords extracted from the surrounding text to annotate the image.

Repurposing of user input is an important aspect of agents in general, and it is a key to reducing user workload.

The user types the description of the picture in order to communicate to another person, not specifically to annotate the image. But once the computer has the input, why not take advantage of it? This repurposing of user input is an important aspect of agents in general, and it is a key to reducing user workload. Aria's guesses are, of course, imperfect, but the user always has the option of editing out incorrect guesses. When a user clicks on an image, for example, Aria displays a keyword editor so the user can select a set of appropriate keywords and avoid the irrelevant ones. If Aria misses an appropriate annotation, the user can manually drag words from the text editor to annotate an image, but the interaction is still more streamlined than conventional image annotation.

Because it is still a stumbling block for many beginning users, we thought it was important to automate the picture-loading process. Automatic image loading is also relatively easy to accomplish. When the user inserts flash card media into the computer, Aria immediately loads the images from the card without any further user intervention. This removes some flexibility in terms of storing the pictures, but it also removes a step that is annoying to most users.

Aria automatically polls for input every few seconds, which eliminates the need for a "load pictures" operation, a "save as" dialog box, figuring out where the file system should save the pictures, and remembering the names of the image files themselves. Because the user usually wants to see them immediately, Aria brings the most recently inserted pictures to the top of the retrieval window.

USER TESTING

At Kodak's Boston Software Development Center, we are currently conducting user studies of Aria based on in-depth interviews and observations. We feel that our preliminary results are indicative of what we would find in a larger study.

After seeing a brief demo of both Aria and a conventional image editing and cataloging application, the participants were asked to use both applications to compose an e-mail message to a friend including at least three images from a memory card full of photographs. All participants had used digital cameras and e-mail, but they were not computer professionals or programmers.

Initial session

During the first session, we told participants they could do whatever organizing activities they thought might help them find pictures in the future—creating annotations, folders, albums, and so forth—although they were not required to do anything but send the e-mail message. The participants returned two weeks later, and we asked them to write a letter to a different person about the same event. In addition to observing what they would choose to do, we also wanted to see whether Aria's annotations or the conventional albums or folders would be helpful in finding photos or remembering story details the participants might have forgotten after the two week hiatus.

Participants found Aria's automatic image loading feature especially helpful. They described the process of selecting pictures and e-mailing messages with Aria as being quick, fun, and easy. In particular, participants liked incorporating pictures into the text. They especially liked being able to view the pictures while writing their e-mail message without having to switch applications or modes. One subject observed that when sending pictures without any text, which he often did, just using an e-mail attachment might be faster.

Second session

When the participants used Aria in the second test, it automatically brought up appropriate pictures as they typed, making it easier to access their previously annotated images, which served as useful reminders both from the story to the pictures and from the pictures to the story. With Aria, using pictures for storytelling becomes an iterative process: A detail of the story brings up an appropriate picture, which then triggers more memories of the story in the user's mind, and so on.

In contrast to Aria, when participants used the conventional image application in the second test, simply browsing the contact sheet of thumbnail pictures was the only feature that actually aided image retrieval. Although several users initially bemoaned Aria's lack of folders, albums, or other grouping mechanisms, only one user actually created an album in the conventional image editing application. He called it "Story JA" (his initials), which was not likely to aid future retrieval. No participants in the test created folders in the file system, moved any of the files into existing folders, or renamed any of the files from their meaningless camera-supplied names (like P000007.jpg) during the test. Several indicated that they hadn't had the time to organize their home photo collections into folders or properly named files either.

USER FEEDBACK

Some study participants expressed concern that Aria might, in some cases, annotate or retrieve the wrong things. Some annotations that Aria proposed weren't correct, but having a few incorrect annotations didn't seem to be a problem, especially compared to the prospect of having little or no user-supplied annotation. Some participants used the option of editing Aria-supplied annotations to remove incorrect guesses. Although we didn't observe any egregious cases of mislabeling in the test, long-term use is necessary before we can assess the overall accuracy of annotations. Most negative comments on Aria concerned the limited features in our prototype, built using Java e-mail and image components, compared to applications such as Eudora or Photoshop that have features such as spelling checkers, a thesaurus, or image editing.

The participants provided some helpful suggestions, such as the need to maintain consistency between the annotations and text even when the text is subsequently edited. They also wanted to be able to go from a picture to a set of past e-mail messages that contained that picture. Despite these suggestions, results of a summary questionnaire showed that Aria scored decisively better overall than the conventional image editing and cataloging application.

One problem we hadn't expected in the testing is that it is actually difficult to get users to express frustration about bad software. It seems that people are so acclimated to the shortcomings of conventional software that they cease to question it or complain about it. For example, few of us complain about having to search through a hard disk file system with a standard file dialog box because we all do it so often. When Aria eliminated the file dialog box by loading pictures automatically, people complimented it but didn't criticize the other application for having required it in the first place.

At one point, the conventional image editing and cataloging application lost the text of an e-mail message a user was typing. This happened because the application requires the user to choose pictures before starting to type a message. If the user returns to the picture-selection screen, any previously typed text is lost without warning. We were shocked to watch, from behind the half-silvered mirror, as the user calmly said, "I guess I have to retype it." This user also did not criticize the conventional application on the evaluation questionnaire. Perhaps he expected computer software to be unreliable, so nothing seemed unusual.

Among related work, FotoFile[8] offers some annotation-retrieval integration, but doesn't come close to full integration into common applications like e-mail. This probably represents the current state of the art in consumer-oriented image annotation and retrieval systems. This system incorporates some automatic image analysis to propose annotations, but it does not do any observational learning. Watson[9] is an image-retrieval system that uses an agent to observe user action, but it doesn't do annotation.

Our future work will center on taking advantage of more opportunities to use context to determine appropriate situations for image annotation, image library browsing, and retrieval. Perhaps in the future, GPS systems in cameras could even report the location where the picture is taken. We're often asked how our approach will scale to large image collections. We have some initial ideas that need to be worked out, but we are investigating ways of automatically annotating groups of images. For example, if one picture is about a wedding, there's a good chance that subsequent pictures taken within a three-hour span and close to the same location will describe the same event.

Future work will center on using context to determine appropriate situations for image annotation, image library browsing, and retrieval.

Keywords could relate to ontologies and knowledge bases, such as WordNet,[10] to do inheritance or simple inference searches. Aria's retrieval technology treats sets of images as an unstructured database, but perhaps a better method would be to look at sets of pictures as linked networks.

In the long run, we are interested in capturing and using common-sense knowledge about typical picture-taking situations. Though automated full-image understanding remains out of reach, image-based retrieval continues to improve. Image retrieval systems based on computable image properties, such as color histograms or textures, appear to be achieving some success. Future work might integrate one of these systems into our agent to automatically propagate user-annotated keyword candidates to similar images. ✸

Acknowledgments

The authors acknowledge Charles Judice and Mark Wood at Kodak for their support of this work, Mona Patel and Kristen Joffre for the user evaluation, and Brad Rhodes for his contributions to the code. We also acknowledge the support of the News in the Future and Digital Life Consortiums at the MIT Media Lab.

References

1. J. Ashley et al., "The Query by Image Content (QBIC) System," *Proc. 1995 ACM SIGMOD Int'l Conf. Management of Data,* ACM Press, New York, 1995, p. 475.
2. S-F. Chang, "Content-Based Indexing and Retrieval of Visual Information," *IEEE Signal Processing,* July 1997, pp.45-48.
3. A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Tools for Content-Based Manipulation of Image Databases," *IEEE Multimedia,* Summer 1994, pp. 73-75.
4. H. Lieberman and T. Selker, "Out of Context: Computer Systems that Adapt to and Learn from Context," *IBM Systems J.*, vol. 39, no. 3, 2000, pp. 617-631.
5. B. Rhodes and T. Starner, "The Remembrance Agent: A

Continuously Running Automated Information Retrieval System," *Proc. 1st Int'l Conf. on the Practical Application of Intelligent Agents and Multiagent Technology (PAAM 96)*, The Practical Applications Company, London, UK, 1996, pp. 487-495.

6. H. Lieberman, "Autonomous Interface Agents," *ACM Conf. Human-Computer Interfaces*, ACM Press, New York, Mar. 1997, pp. 67-74.
7. G. Salton, *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison Wesley, New York, 1989.
8. A. Kuchinsky, "FotoFile: A Consumer Multimedia Organization and Retrieval System," *ACM Conf. Human-Computer Interfaces*, ACM Press, New York, 1999, pp. 496-503.
9. J. Budzik and K.J. Hammond, "User Interactions with Everyday Applications as Context for Just-in-Time Information Access," *ACM Conf. Intelligent User Interfaces (IUI 2000)*, ACM Press, New York, Jan. 2000, pp. 44-51.
10. Y.A. Aslandogan et al., "Using Semantic Contents and WordNet in Image Retrieval," *Proc. 20th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, ACM Press, New York, 1997, pp. 286-295.

***Henry Lieberman** is a research scientist at the MIT Media Laboratory. His research interests include building software agents that learn from interacting with the user. He received an habilitation degree (PhD equivalent) in computer science from the University of Paris VI-Pierre et Marie Curie. Contact him at lieber@media.mit.edu.*

***Elizabeth Rosenzweig** is a research scientist at the Eastman Kodak Company. Her research interests include human-computer interaction, intelligent agents, and digital imaging. She received an MS from MIT and is president of the Usability Professionals Association. Contact her at erosenz@kodak.com.*

***Push Singh** is a doctoral candidate in media arts and sciences at the MIT Media Laboratory. His research interests are in the representation of common sense knowledge. He directs the Open Mind project, which aims to collect an open-source common-sense knowledge base via the Web. Contact him at push@ media.mit.edu.*