Marie-4: A High-Recall, Self-Improving Web Crawler That Finds Images Using Captions

Neil C. Rowe, *US Naval Postgraduate School*

Marie-4, a Web crawler and caption filter, searches the Web to find image captions and the associated image objects. It uses a broad set of criteria to yield higher recall than competing systems, which generally focus on high precision.

My students and I are building intelligent software agents—crawlers or spiders—to find information on the World Wide Web. Images, among the most valuable Web assets, make the Web an extensive virtual picture library. But finding images on the Web to match a query is quite difficult: Typically only a small fraction of page text describes associated images, and images are not captioned consistently. Content-based image retrieval systems that analyze the images themselves[1] are progressing, but the systems require considerable image-preprocessing time. Furthermore, surveys of users doing image retrieval show that users are more interested in the identification of objects and actions depicted by images than in the color, shape, and other visual properties that most content-based retrieval systems provide.[2] Because object and action information is more easily obtained from captions, caption-based retrieval appears to be the only hope for broadly useful image retrieval.[3]

Commercial tools such as AltaVista's Image Search search engine achieve respectable precision (the fraction of correct answers retrieved out of all answers retrieved) by indexing only "easy" pages, such as photograph libraries where images are one to a page and captions are easy to identify. Recall (the fraction of correct answers retrieved out of all correct answers) is equally or more important than precision, but users often do not realize how small it is for their queries. In experiments with 10 representative phrases, using pages retrieved by a traditional keyword-based Alta Vista search to calculate recall, we found Alta Vista's Image Search had a precision of 0.46 and recall of 0.10. Higher recall requires dealing with a large variety of page layout formats and styles of captioning.

Recent work has made important progress on general image indexing from the Web by intelligent information filtering of Web text.[4–6] By looking for the right clues, large amounts of Web page text can be excluded as captions for any given image, and the captions in the remaining text can be inferred. Clues can include caption candidate wording, HTML constructs around the candidate, distance from the associated image, image-file name words, and associated image properties. These clues reduce the amount of text to examine to find captions, and the reduced text can be indexed and used for keyword-based retrieval. But so far, the selection of these clues has been intuitive, and there has been no careful study of the relative values of clues.

This article reports on Marie-4 (see Figure 1), our latest in a series of caption-based image-retrieval systems.[7] Marie-4 uses a wide range of clues, broader than any system we know about, to locate image-caption pairs in HTML Web pages. It is in part an expert system where the knowledge used is not especially novel in itself, but the synergy of a variety of knowledge working together provides surprisingly good performance. Unlike some caption-based retrieval systems[3] and previous Marie systems, which require an image database with captions already extracted, Marie-4 is a Web crawler that autonomously searches the Web, locates captions using intelligent reasoning, and indexes them. It does not attempt full natural language processing and does not require the elaborate lexicon information of the earlier prototypes, so it is more flexible.

1094-7167/02/$17.00 © 2002 IEEE

The Web crawler and page scanner

Marie-4 uses a rule-based expert system designed to ensure high recall of captions. It fetches the HTML source code for a given page and scans it for image references. It also finds links to other pages (HREF, FRAME, AREA, and certain JavaScript constructs) and puts them in a queue; pages are subsequently considered in queue order to give a breadth-first search. To localize the search, Marie-4 only examines pages with the same last K words in their site name as the initial page, where K and the initial page are specified by the user. So if $K = 2$ and the starting site is nps.navy.mil, cs.nps.navy.mil and www.navy.mil would be considered but not www.army.mil. A site-URL hash table prevents revisiting the same page, and a page-content hash table prevents visiting a page with the same content.

Image references in HTML are both IMG constructs and HREF links to files with image extensions such as .gif and .jpeg. The page scanner searches for captions near each image reference. The types of captions considered are

1. The filename or words (with punctuation and low-information "stop words" removed) of the full path to the image file
2. Any ALT string associated with the image, which represents associated text
3. Clickable text that retrieves the image
4. Text delineated by HTML constructs for fonts, italics, boldface, centering, table cells and rows, and explicit captions (explicit captions are quite rare)
5. The title and nearest-above headings on the page (but not "meta" constructs because we found them often unreliable)
6. Unterminated or unbegun paragraph (P) constructs
7. Specific word patterns of image reference (for example, "Figure 5.1," "in the photo above," and "view at the right"), as found by partial parsing using a context-free grammar of image references and then checking consistency of reference direction (for example, "above" should refer to an image above the caption)

Brian Frew and I have previously used the first four categories,[4] but they are improved here. Sougata Mukhejea and Junghoo Cho have used the fifth,[5] and the sixth and seventh are new with our work here. We found that identifying these specific types of captions was considerably more precise and successful than just assigning a word weight that was a decreasing function of distance from the image reference.[6]

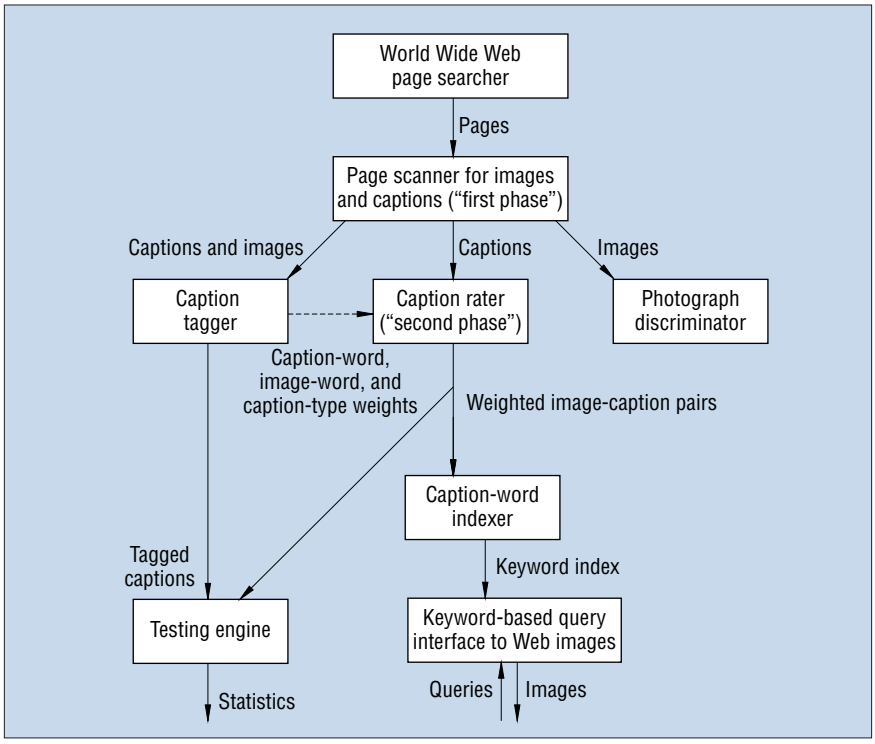

Figure 1: A block diagram of Marie-4, built in Java 2. A crawler searches the World Wide Web from a given starting page. It locates all images on each page and good candidate captions for them. This information is passed to a caption rater that assigns a likelihood to each image–caption pair on the basis of a weighted sum of factors, and (for selected images) to a caption tagger that lets a user confirm captions manually for training and testing. The indexer indexes the inferred caption words. The Web-based query interface uses them to answer queries in the form of keywords by providing images that match those keywords, sorted in order of decreasing likelihood of match. We also developed a discriminator for photographs from graphics, but experiments showed it did not help much for caption extraction, and we do not use it in the final system.

The fourth, sixth, and seventh types require the caption candidate to be within image-distance bounds. We set the bounds, on the basis of experiments, at within 800 characters of an image, provided the intervening characters do not contain a structural boundary, or within 1,500 characters if the candidate's construct surrounds the image reference. We determined that a structural boundary can usually be another image construct, the end of a table row, a horizontal line, a beginning or end of a paragraph when searching for weaker constructs, and an opposite of a sought construct (for example, if we encounter the end of italics when we are searching left for the start of italics). The inferred rules for scope handle a number of special cases and require a carefully designed rule-based system.

For an example, consider a Web page with this HTML source text:

```
<title>Sea Otters</title>
<h2>The California Sea Otter</h2>
<a href="images/otter.jpeg"><img
src="images/smallotter.gif" alt="Pair of sea otters"
    </a>
<center><I>Click on the above to see a larger
    picture.</I></center>
<hr><a href="home.html">Go to home page</a>.
```

This page has a small image "smallotter.gif" that when clicked retrieves a larger image "otter.jpeg." Both images have four caption candidates: a title of "Sea Otters," heading-font text of "The California Sea Otter," an "alt" string of "Pair of sea otters," and italicized centered text of "Click on the above to see a larger picture." In addition, the larger image has a filename caption candidate of "images otter jpeg" and the smaller has "images smallotter gif." "Go to home page" is not a candidate because it is separated from the image reference by a horizontal line ("hr").

Several criteria prune candidate captions. Captions on images not retrievable from the Web (incorrect links or those removed sub-

sequently) are excluded by testing the links. We removed HTML and JavaScript syntax from the candidates, and eliminated subsequent null captions. Small images or those not reasonably square are more likely to be graphics and hence unlikely to have captions. We require that width and height be greater than 80 pixels and that the length-to-width ratio be less than 3. (Image file sizes are retrieved from the Web to estimate image sizes not specified on the Web page.) We eliminated images appearing more than once on a page and images appearing on three or more different pages from consideration because such images are almost always iconic and uncaptionable. We also eliminate duplicate captions, and only examined quoted constants within JavaScript code (because full analysis would require implementing a nondeterministic interpreter). We derived these criteria from experiments where we found thresholds that eliminated less than one percent of the correct candidates.

Caption assessment is necessarily subjective, but we decided that an acceptable caption should describe the image objects, their properties, or their relationships. Using a training set of 3,945 caption candidates from 14 representative sites with images (the first 300 candidates found, or all if the site has fewer than 300), we recognized 1,077 caption candidates for a precision of 0.273 at perfect recall. Each caption candidate was manually inspected to confirm it was a caption of its referenced image. Frew and I previously reported a precision of 0.014 for text queries for a standard browser that tried to find pages with images matching particular words.[4]

Recall is harder to estimate, but we got 0.97 in a manual inspection of 20 random Web pages (we defined recall as the fraction of the image-describing text on these pages that was found by our page scanner). The missed caption text was in paragraphs insufficiently related to corresponding images. Our program labeled 6.28 percent of the total characters in the training set as part of captions, thus reducing the data by a ratio of 16 to 1 while hurting caption recall by only three percent; 24.7 percent of the images had at least one proposed caption. 37.3 percent of the images were too small or thin, 5.7 percent were excluded because they appeared on three or more Web pages, 3.5 percent were excluded because they appeared twice on the same Web page, and the remaining 28.8 percent had no qualifying captions. Only approximately one percent of all descriptively captioned images were incorrectly excluded by these three criteria, so recall was 99 percent. As for precision, 69 percent of the images proposed in image–caption pairs had at least one caption. Execution time for the crawler and filter averaged approximately five seconds per page on a 500–MHz Pentium PC, but this varied widely per site.

Increasing crawler output precision

The caption-candidate filtering eliminates only the obvious noncaptions. For better precision and to rank caption candidates in answers to user queries, we assign likelihoods to candidates using a simple neural network with carefully chosen factors.

Modeling the effect of caption clues

We used the training set, with all captions tagged, to identify positive and negative clues for captions. Clues are the occurrence of specific words, caption attributes, or image attributes. The strength associated with clue i is the conditional probability that the clue occurs in a candidate when it does occur, estimated by $r_{ci}/(r_{ci} + r_{ui})$, where r_{ci} is the number of captions containing clue i and r_{ui} is the number of noncaptions containing clue i in the training set. A word's or attribute's absence from a caption can be a weak clue that we have a caption, but we did not find this generally helpful. Clue occurrence can be modeled as a binomial process; our approach says that a clue is statistically significant if it exceeds the binomial distribution prediction by more than one standard deviation in either direction, or

$$\left| r_{ci} - \left((n_c / n_u) r_{ui} \right) \right| > \sqrt{r_{ui} r_{ci} / \left(r_{ci} + r_{ui} \right)} ,$$

where n_c is the number of captions and is the number of noncaptions.

Nonlinear functions were applied to the factors so that their median value was about 0.5 and standard deviation was approximately 0.15. For a total caption rating from a set of clues, we use a linear model taking a weighted sum of the adjusted likelihoods of all clues. Linear models can be contrasted with Naïve-Bayes and association-rule methods; they are appropriate when clues are strongly correlated,[8] as are many caption word clues. Linear models are preferable to decision trees, because complex logical relationships are unlikely between clues, and preferable to case-based reasoning, because no small set of "ideal" captions exists.

Clues from specific words in the caption

We tabulated word counts and calculated the associated conditional probabilities for the training set. The expected value in the training set was 0.273, so we only used words deviating more than one standard deviation from this value in either direction. Some word clues found in the training set were valid for the Web in general (such as "gif," "center," and "photograph"). Others reflected unrepresentative phenomenon in a small Web sample (such as "child" and "destroyer") and needed to be diluted by more data. The word clues in a caption were totaled by

$$\exp\left(\left(\sum_{i=1}^{M} \left(q_i - \bar{q} \right) \right) \Big/ M \right) ,$$

where M is the number of word clues, q_i is the conditional probability for word i of the caption, and $\bar{q}$ is the fraction of captions in the training set, and we use exponentiation keep the result positive.

Destemming words first is important for word clues because related forms often occur in natural languages, such as "picture," "pictures," "pictured," "picturing," and "picturedly" in English. We developed a destemmer using Martin Porter's algorithm[9] but enhanced it to cover important cases it missed, such as "ier," "edly," "ity," and "tionism" endings, and the necessary irregular forms (422 words and 1002 intermediate forms) that it did not enumerate. We improved it using a Unix spelling utility dictionary of 28,806 common English words, mapping them first through the Wordnet thesaurus system to eliminate approximately 4,000 words, and then manually inspecting 50 separate classes of endings to eliminate another 4,000 words. This provided 19,549 words, which we then supplemented with 674 technical words from computer science papers and words from the training set that were incorrectly destemmed (261, mostly proper nouns ending in s). The final lexicon was 20,223 words.

Other text clues

The caption type is a good clue, both for captions and noncaptions. Table 1 shows the statistics on the training set. No types are cer-

tain to be captions; even "alt" strings can just be a useless word such as "photo."

The image-file path words also furnish clues; we found 67 negative clues (for example, "button") and 10 positive clues (for example, "media"). Powerful clues are the occurrence of the same word in both the caption and image file name, such as "Hermann" and "Hall" in the image name "http://www.nps.navy.mil/hermann_hall.gif" and caption "View of Hermann Hall." The image format is a good clue, because 53.6 percent of JPEG images were valid in the training set and 16.2 percent of GIF images. Other useful clues are digits in the image filename (images important enough to be captioned are often numbered), sentence length, and the caption's distance from its associated image. We also explored several formulations of a "template fit" clue that measured how common that kind of caption and its placement (above or below the image) were for other pages on its site, but they were not sufficiently reliable.

Deciding whether an image is a photograph or a graphic

Knowing whether an image is a photograph or a graphic is helpful.[4,6] A sample of our training set showed that 95 percent of the photographs had captions, whereas 10 percent of the nonphotographs had captions. Both can be stored in similar image formats, so some content analysis is necessary to confirm photographs. We followed Marie-3's linear model, but using these factors

- The image's size, measured by the diagonal's length
- The number of color bins having at least one associated pixel, for 256 bins evenly distributed in intensity-hue-saturation space
- The count in the color bin having the most associated pixels
- The average "saturation"
- The average color variation between neighboring pixels as measured in intensity-hue-saturation space
- The average brightness variation between neighboring pixels

Again, we applied nonlinear functions to the factors to adjust their scales. Figure 2 shows precision versus recall for discriminating photographs, for the six factors and their weighted average on the 648 photographs and 309 nonphotographs in the training set (excluding those that had become unavailable). Results were disappointing. We optimized to find the best weightings of the factors. The fifth and sixth factors are clearly negative influences, but we could get no improvement by assigning them negative weights. The "Best weighted average" represents the best weighted sum with the first four factors, for which we got 93.4-percent precision at 50-percent recall on the training-set images, insufficiently better

Table 1: Statistics showing the likelihood of a caption given candidate type in the training set.

Caption candidate type	Number that were captions	Number in training set	Probability	Significant?
i (italics)	2	5	0.40	No
b (boldface)	24	67	0.36	No
em (emphasis)	0	1	0.00	No
strong	1	15	0.07	No
big	1	4	0.25	No
font	45	120	0.37	Yes
center	4	63	0.06	Yes
td (table datum)	90	193	0.47	Yes
tr (table row)	141	352	0.40	Yes
caption	0	0	—	—
object	0	0	—	—
h1 (heading font 1)	5	15	0.33	No
h2	63	129	0.49	Yes
h3	2	39	0.05	Yes
h4	0	2	0.00	No
h5	0	7	0.00	No
h6	0	1	0.00	No
title	320	936	0.34	Yes
alt (substitute text)	119	481	0.25	No
a (dynamic link)	97	149	0.65	Yes
filename (of image)	42	1143	0.04	Yes
wording	21	45	0.47	Yes

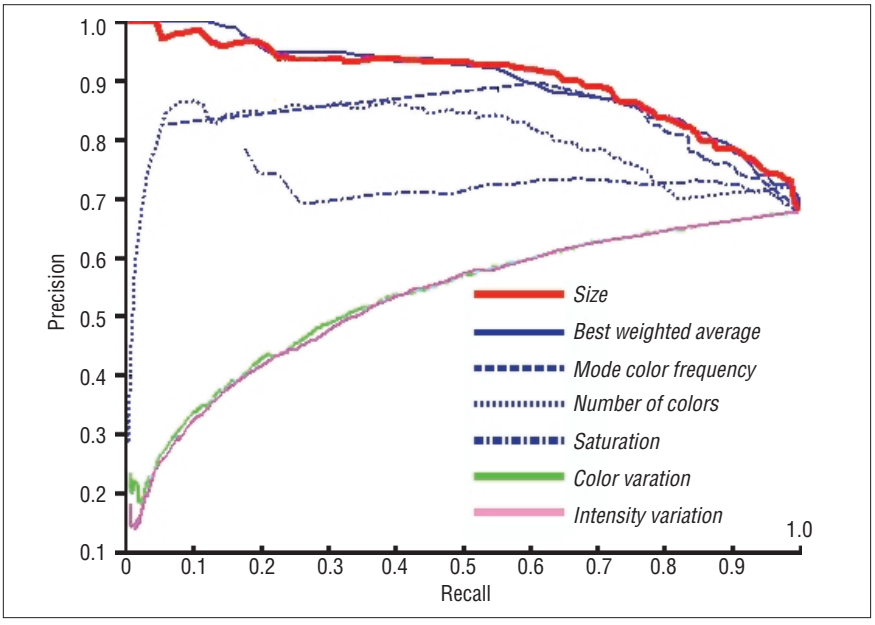

Figure 2: Precision versus recall on the training set for photograph discrimination using six factors.

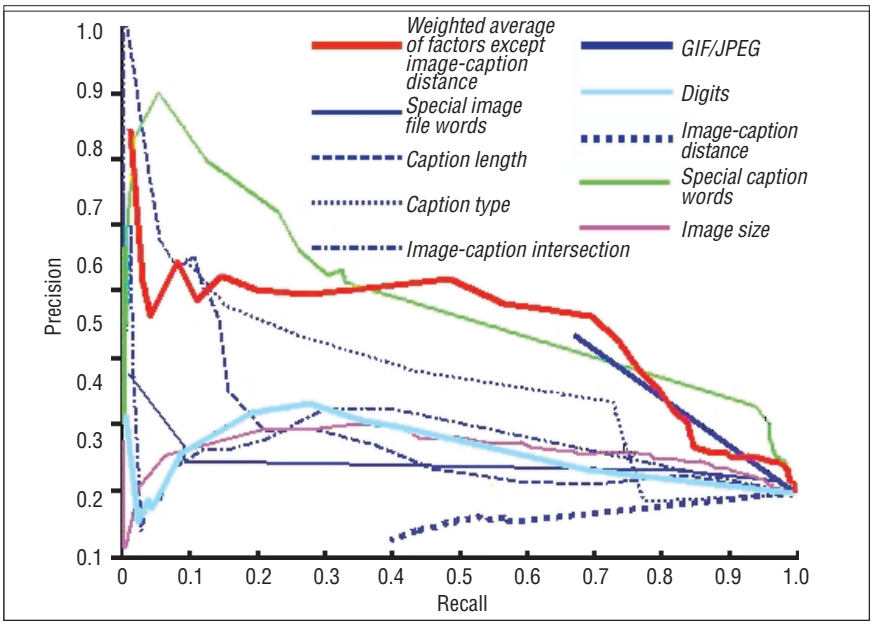

Figure 3. Precision versus recall on the training set for caption discrimination using nine factors.

than using the size factor alone. In addition, size is easier to compute than the other image properties because it can be extracted from the image-file header without image processing. So we used only the size factor in subsequent assessment of caption candidates.

Putting all the caption clues together

Finally, we implemented our linear model (a simple neural network), added the caption clue strengths with the image size clue, and rated each candidate's likelihood of being a caption. Eight of the nine factors are helpful (see Figure 3), with the exception of the distance between the caption and the image. (Recall in Figure 3 is for only the results of the caption candidate rater and should be multiplied by 0.97 to get the total recall.) We obtained weightings for the eight factors by both least-squares linear regression and steepest-ascent optimization on the training set. The latter weightings were better, but only 2 percent better than weights of 0.1 except for 0.3 for the caption-word factor. This unimpressive improvement suggests a danger of overtraining and argues against using a more complex neural network.

Surprisingly, image size did not have much effect. Assessing caption candidates (not captionable images) means that other factors matter more. We excluded the smallest and most asymmetric images with the page scanner, something not done for the Figure 2 experiments. The "special image words" factor appears unhelpful, but this is misleading: Only a few image file names had special word clues, but when they occurred, there was a clear advantage to exploiting them. On the other hand, the distance between the caption and image is clearly unhelpful because of the many "filename" and "alt" candidates at distance 0 that are not captions; this questions the reliance on it in the work of Stan Sclaroff and his colleagues.[6]

Testing the caption rater

To test the caption rater we modified the crawler to more randomly sample Web pages. This is harder than it might seem; an estimated three billion Web pages exist, and the Web's loose organization precludes any easy way to choose a random page. So we started with 10 representative pages (not necessarily those with many images as with the training set) and performed a random search to retrieve 600 pages starting from each of them. For a more depth-first search, we used only two random links on each page (not necessarily links to its site) to find new pages, and we selected one random caption–image pair for each page. This encouraged exploration because the search starting with a metallurgy journal site spent most of its time on country music sites and the search starting with a fashion site spent most of its time on sports-news sites. But this did bias search toward sites with many links to them, which raises ethical questions as it does with the popularity-weighting Google search engine.

Testing found 2,024 caption–image pairs for 1,876 images. The number found per Web page varied from 0.17 to 16.71 over the 10 runs. Captionable images were fewer than small graphical icons; captions themselves are not routine even on captionable images. The author tagged the caption–image pairs as to whether they were captions. The fraction of captioned images per site varied widely, from 0.020 (www.nytimes.com) to 0.260 (www.amazon.com) to 0.464 (www.arabfund.org) to 0.843 (www.charteralaska.net) to 0.857 (www.kepnerfamily.com). The proposed clues were helpful for this test set; other clues found were whether the page name ended in "/" (negative) and whether the site name ended in ".mil" (positive). Confirmed clue words in image-file names were "logo," "icon," "adobe," and "service" (negative), and "people," "library," and "photo" (positive). Confirmed word clues in captions with their caption probabilities were "update" (0.000), "thumbnail" (0.000), "download" (0.029), "customer" (0.038), "week" (0.780), "forward" (0.875), and "photographer's" (0.960).

To test whether our caption rater could learn from experience using statistics on caption words, image-file words, and caption types, we ran further tests. These tests had a second tagged set of 2,148 caption candidates on 1,577 images obtained from the crawler by the random search starting on 16 additional sites. We rated the captions using four sets of probabilities obtained from statistics. Version 1 used no statistics but did use the image-clue words from Marie-3. Version 2 used statistics from just the training set. Version 3 used statistics from both the training set and the first (2024-pair) test set, with 75 percent of the filename and title candidates eliminated to provide a better balance among caption types. Version 4 used artificially tagged data.

Figure 4 shows the results for precision versus recall, demonstrating the clear advantage of more knowledge, except for greater random fluctuations at low values of recall (with small sample sizes). A steady increase in precision occurred as recall decreased, and we observed no significant differences in the curve shape on any major subsets of the test data.

Ultimately the system should tag obvious captions itself to provide further statistics. The dashed line in Figure 4 illustrates this on 40,239 candidates from the crawler, a super-

set from which we derived the training and test sets. We rated these candidates using statistics of the training and first test sets, assumed the top 10 percent were captions to derive a new set of statistics, then reran the second test set with the guidance of the new statistics. (The top 10 percent gave 80-percent precision on the second test set, so the new statistics should be roughly 80-percent correct.) Although performance was not as good as for the manual-tag statistics, this approach can be improved with smarter tagging.

The query interface

The words of all proposed captions found by the page scanner are indexed. The index is used by keyword-lookup Java servlets that run on our Web site (http://triton.cs.nps.navy.mil:8080/rowe/navmulib.html). Users enter keywords for the images they seek and specify how many answers they want. The servlet destems the keywords, uses its index of destemmed words to find images matching at least one keyword, ranks the matches, and displays the best matches of images and captions. The user can click on links to go to the source Web pages. Figure 5 shows an example output for a database of all images on Web pages at our school.

Table 2 shows some statistics on a more ambitious project analyzing and indexing the 667,573 images we found on all 574,887 publicly accessible U.S. Navy (or navy.mil) sites using our single 500-MHz PC. The servlets takes approximately 15 seconds to load on a Unix server machine and 10 to 90 seconds to answer a typical three-word query. A companion servlet built with many of these principles indexes all Navy audio and video clips.

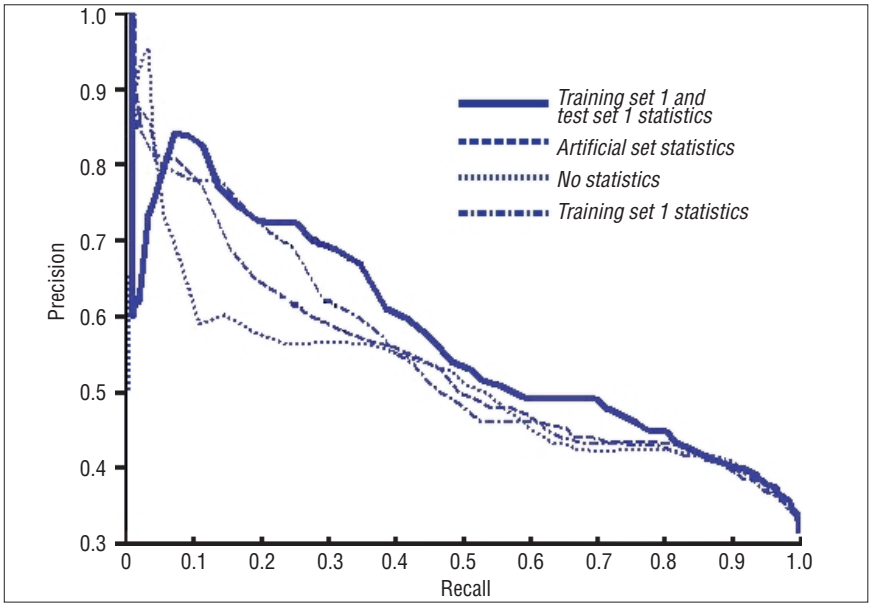

Figure 4: Precision versus recall for four versions of the caption rater, illustrating its learning from experience.

Match ranking exploits additional factors besides the overall caption likelihood. Following the recommendations of Gerand Salton and Chris Buckley for short queries and short independent documents such as captions,[10] we should add weights for all matching keywords, and an inverse document-frequency factor should ensure a higher weight on rarer keywords. In addition, for a random sample of 363 true captions from our full training set, the probability of a keyword being depicted (corresponding to some image area) in the image decreased steadily from 0.87 for three-word captions to 0.24 for 90-word captions. (Length has the opposite effect for keyword-match rating that it has for ascertaining caption likelihood.) A word's probability of being depicted in the image also steadily decreased as a function of relative position in the caption, from 0.68 for words in the first 10 percent of the caption to 0.15 for words in the last 10 percent. This is because long captions tend to include background material toward the end, more so than other kinds of text. So we did least-squares fitting for these factors from the sample.

For the overall weight, we use predominantly a Naive-Bayes approach (because the factors are close to independent) where we sum the products of the factors for each keyword (because we expect the keywords to be correlated). We added minor factors for

Table 2: Statistics on the building of the Marie-4 servlet that indexes all images at navy.mil sites.

Number of items found	Size of result (Mbytes)	Computation real time (minutes)	Description
6,002,295	1,468.4	approximately 13,000	Initial page scan (in which 574,887 Web pages were examined)
2,198,549	582.4	860	Checking for the existence of image files, retrieving the size of those not described on the Web page, excluding captions on too-small images, and removing images with too many references
2,198,549	897.1	130	Rating of caption candidates
211,398	462.0	197	Indexing of caption candidates (by root words)
211,398	3.8	—	Main-memory hash table for the servlet to the secondary-storage index
85,124	5.5	—	Text of all distinct Web-page links for captions
667,573	67.4	—	Text of all distinct image-file links for captions
2,193,792	124.0	—	Text of all distinct captions

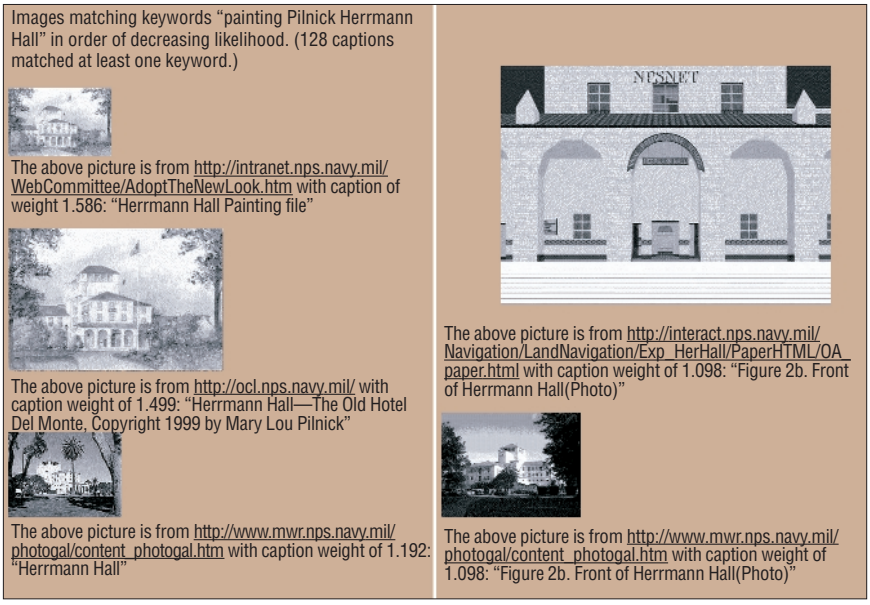

Figure 5. Example use of the query interface, showing the best five candidates found for the query "painting Pilnick Herrmann Hall."

capitalization matching and keyword adjacency in the caption with small-scale factors, as with many current Web search engines. So, the weight on caption-image pair i is

$$\left(c_i\left((0.176*\ln(k))+0.968\right)\sum_{j=1}^{m}\left[\ln\left(N/n_j\right)*\left(\left(2.33*p_j\right)+2.717\right)\right]\right)$$
$$+\left(0.1*m_i\right)+\left(0.1*a_i\right)+\Sigma\left(0.05*b_i\right)$$

where c_i is the likelihood the caption describes the image, k is the number of nontrivia in the caption, j is the index number of a keyword, m is the number of keywords, N is the number of captions, n_j is the number of captions containing keyword j after destemming, p_j is the fraction of the distance through the caption that keyword j first appears, m_i is the number of capitalized keywords that exactly match capitalized caption words, a_i is the number of keywords that appear adjacently in the caption, and b_i is the number of keywords that appear separated by a single word in the caption.

To test the formula, we generated 32 three-keyword queries by choosing 150 random caption candidates and picking three representative keywords from each of those that were true captions. In 22 of the 32 cases, the above formula gave better answers than a control formula using only the caption-likelihood and document-frequency factors. In nine cases the answers were the same; in one case they were worse.

Web diversity requires automated tools to find useful information. But this diversity means the tools must have some intelligence to cope with all the different formats they find. We have shown that the seemingly wide diversity of image formats on the Web can be substantially indexed with our tool. Careful tests on 8,140 caption candidates for 4,585 representative images have confirmed the factors we use and how they are combined. But this comprehensive approach requires using a spectrum of methods, not just one, and learning from experience must play an important role. ■

ACKNOWLEDGMENTS

Jorge Alves, Sharif Calfee, Mark Evangelista, Vanessa Ong, Steve Simmons, and Nickolaos Tsardas built prototypes of several software modules.

REFERENCES

1. A. Yoshitaka and T. Ichikawa, "A Survey on Content-Based Retrieval for Multimedia Databases," *IEEE Trans. Knowledge and Data Eng.*, vol. 11, no. 1, Jan./Feb. 1999, pp. 81–93.
2. C. Jorgensen, "Attributes of Images in Describing Tasks," *Information Processing and Management*, vol. 34, nos. 2–3, 1998, pp. 161–174.
3. R.K. Srihari, "Use of Captions and Other Collateral Text in Understanding Photographs," *Artificial Intelligence Rev.*, vol. 8, nos. 5–6, 1995, pp. 409–430.
4. N.C. Rowe and B. Frew, "Automatic Caption Localization for Photographs on World Wide Web Pages," *Information Processing and Management*, vol. 34, no. 1, 1998, pp. 95–107.
5. S. Mukherjea and J. Cho, "Automatically Determining Semantics for World Wide Web Multimedia Information Retrieval," *J. Visual Languages and Computing*, vol. 10, no. 6, Dec. 1999, pp. 585–606.
6. S. Sclaroff et al., "Unifying Textual and Visual Cues for Content-Based Image Retrieval on the World Wide Web," *Computer Vision and Image Understanding*, vol. 75, nos. 1–2, July/Aug. 1999, pp. 86–98.
7. N.C. Rowe, "Precise and Efficient Retrieval of Captioned Images: The Marie Project," *Library Trends*, vol. 48, no. 2, Fall 1999, pp. 475–495.
8. I. Witten and E. Frank, *Data Mining: Practical Machine Learning with Java Implementations*, Morgan Kaufmann, San Francisco, 2000.
9. M.F. Porter, "An Algorithm for Suffix Stripping," *Program*, vol. 14, no. 3, July 1980, pp. 130–137.
10. G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol. 24, 1988, pp. 513–523.

The Author

Neil C. Rowe is a professor and the associate chair of computer science at the US Naval Postgraduate School. His main research interest is intelligent access to multimedia databases; he has also done work on image processing, robotic path planning, and intelligent tutoring systems. He has SB, SM, and EE degrees from the Massachusetts Institute of Technology and a PhD in computer science from Stanford University. Contact him at Code CS/Rp, 833 Dyer Rd., Naval Postgraduate School, Monterey, CA 93943; ncrowe@nps.navy.mil.