Cyborg Metadata: Humans and machines working together to manage information – Part 2: Images

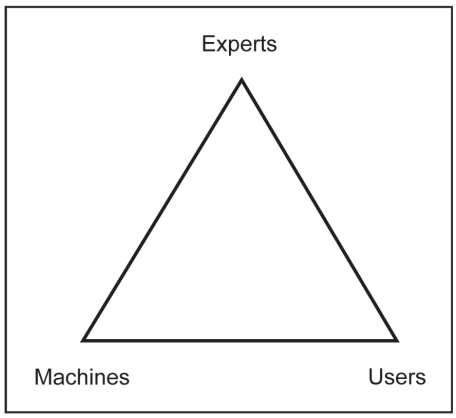
Matt Moore*

This article discusses different emerging techniques for managing digital images. Various metadata schemes and thesauri are outlined. Research into user-generated metadata from sites such Flickr are discussed alongside novel game-based techniques such as the ESP game and Peekaboom. Finally, software-driven content-based image retrieval techniques such as the Imense search tool are outlined.

INTRODUCTION

The first article in this series proposed that metadata management needs to shift from being the sole responsibility of experts to a more distributed activity carried out by a combination of experts, machines and users.[1] Each circumstance will require a particular mix of these three roles, but few situations will be the preserve of just one. This article expands that proposal from the domain of text to that of images.

FIGURE 1 The Cyborg metadata triangle



The proliferation of electronic text was a 1990s phenomenon. The combination of easy-to-use word processing software, email and HTML-based webpages simplified the creation and distribution of the word, especially in its typeset form. Images were a different story. We had to wait until the 2000s for cheap, effective digital cameras and scanners to become available for the mass market. As these digital cameras proliferated, so distribution mechanisms such as the Flickr photo sharing site (http://www.flickr.com) grew in popularity. The results have been dramatic. Commercial stock photography companies such as Corbis (http://www.corbisimages.com) and Getty Images (http://

* Matt Moore is a director of Innotecture, an occasional lecturer at the University of Technology Sydney and chair of the New South Wales Knowledge Management Forum. He has spent over a decade working in knowledge and information management, learning and development and internal communications with organisations such as PwC, IBM, Oracle and the Australian Securities and Investment Commission (ASIC).

Many thanks to David Riecks, CEO, Controlled Vocabulary for generously sharing his experience with image metadata standards and tools.

All websites referred to in this article were viewed on 1 February 2011.

[1] Moore M, "Cyborg Metadata: Humans and Machines Working Together to Manage Information – Part 1: Text" (2010) 24 OLC 131.

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

www.gettyimages.com) hold hundreds of millions of images. However the images held on consumer sites such as ImageShack (http://www.imageshack.us), Facebook (http://www.facebook.com), Photobucket (http://www.photobucket.com) and Flickr number in the tens of billions.[2] What can be done to manage this deluge of images?

EXPERTS

The commercial world of images has many stakeholders (see Table 1) with many concerns. For these groups, image metadata is not an academic topic. The digital image distribution industry is worth billions of dollars annually. Images that cannot be found or tracked represent lost revenues. With this diversity of goals, it is therefore unsurprising that different groups have worked to produce digital image metadata standards that suit their own needs. Sometimes these efforts overlap. Sometimes they conflict. Sometimes they completely ignore each other.

TABLE 1 A selection of stakeholders in the digital image economy

Entity	Gaols	Concerns
Commercial photographer	Receiving payment and recognition for their work	Are my images easy for potential buyers to find? Are my images correctly identified as mine throughout their lifecycle? Is the assignment of metadata an acceptable overhead on my time?
Technology providers (eg Canon, Adobe)	Profitable uptake of technology	Can our customers use our products more effectively and easily than those of our competitors? Are our products affordable?
Stock image libraries (eg Getty Images, Corbis)	Profitable acquisition and sale of media assets	How do we acquire images that will be valuable? How do we provide our customers with the most appropriate images? How do we do the above cost effectively?
Image users (eg newspapers)	Effective use of images	How do we find the images we need?
Public sector image collectors (eg Library of Congress)	Cost effective use of public assets	How do we make our collections available to our citizens in a cost effective manner?

Some of the digital image metadata systems that have been developed are described below.

- Technical metadata includes data related to the camera used to take the image and any subsequent tools used in processing. This is largely the responsibility of the camera manufacturers and the Camera and Imaging Products Association (CIPA).
- Descriptive metadata related to the content and context of the image. The key standards here have been developed by the International Press Telecommunications Council (IPTC). See Appendix 1 for a list of IPTC Core Photo Metadata fields. A list of agreed extension fields such as "Person shown in image" and "model release status" is also available from the IPTC website.[3]
- In-house metadata schemas and controlled vocabularies have been developed by the large, commercial image collections such as Corbis and Getty Images.[4]

[2] Pickerell J, "Exactly How Many Images Are Available Online?" (21 June 2010), http://www.rising.blackstar.com/exactly-how-many-images-are-available-online.html.

[3] International Press Telecommunications Council, *IPTC Photo Metadata Standards* (July 2010), http://www.iptc.cms.apa.at/cms/site/index.html?channel=CH0099. See also Saunders S, "IPTC Controlled Vocabulary" (Paper presented at the 4th Photo Metadata Conference 2010, Dublin, Ireland, 9 June 2010), http://www.phmdc.org/phmdc2010/PhMdC2010--SSaunders-ImageCV.pdf; Wikipedia, http://www.en.wikipedia.org/wiki/IPTC_Information_Interchange_Model.

[4] LaBonte A and Forster M, "Searching for Images" (Paper presented at the 4th Photo Metadata Conference 2010, Dublin,

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

- Various thesauri to describe image metadata have been developed by those with large image collections. These include the United States Library of Congress Thesaurus for Graphic Materials (TGM)[5] and the Australian Pictoral Thesaurus (APT).[6]

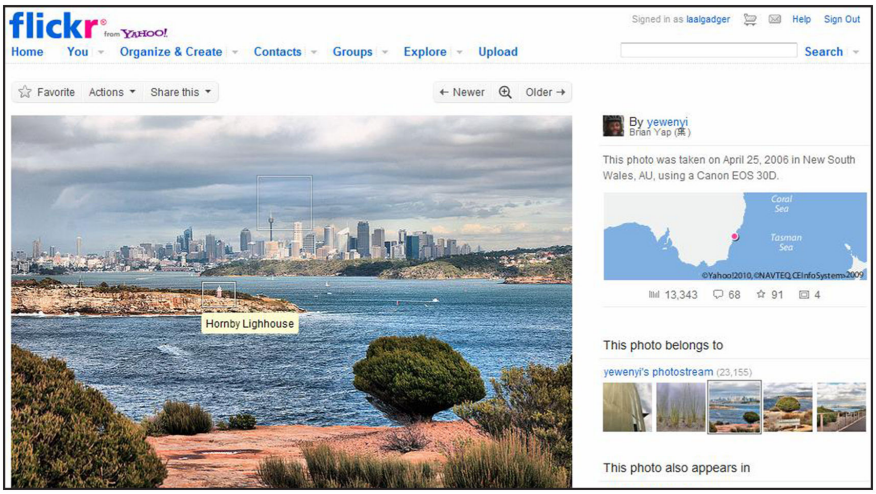
The differences in focus between groups is obvious when the standards and thesauri are compared. The IPTC Photo Metadata Standards has numerous fields that detail the ownership of, and rights associated with, a particular image. The TGM and APT focus primarily on the content of the image, not the ownership issues.

Establishing these knowledge organisation systems is only part of the story. If these systems are not applied to images then they remain pipe dreams. Some of this categorisation and tagging is carried out by the image creators (for example, photographers), some of it by employees of the various institutions involved, and some of it is outsourced to commercial keywording companies. Outsourced keywording may cost US$1-2 per image, which values the notional replacement cost of commercial image metadata in the hundreds of millions of dollars. Of course, this applies only to the metadata that can be identified from the image.

USERS: DELIBERATE TAGGING

The commercial image world is already complex. However, the uptake of digital cameras in the early years of this century now adds more participants and more complexity to the mix. Sites such as Flickr allow users to share their own images and to add descriptive metadata tags to them. These tags are free-text terms and users can tag the images of others as well as their own. As of January 2008, there were 20 million unique tags on Flickr.[7]

FIGURE 2 An example of an image stored on Flickr



Ireland, 9 June 2010), http://www.phmdc.org/phmdc2010/PhMdC2010--MForster-ALaBonte-Searching_for_Images.pdf.

[5] Alexander A and Meehleib T, "The Thesaurus for Graphic Materials: Its History, Use, and Future" (2002) 31 *Cataloging & Classification Quarterly* 189.

[6] Stumm D, "When is a Forest Fire a Bushfire? Towards an Australian Pictorial Thesaurus" (Paper presented at the 13th National Cataloguing Conference, Brisbane, 13-15 October 1999), http://www.pandora.nla.gov.au/pan/77226/20071011-0000/www.sl.nsw.gov.au/staff/dstumm/index.html; Kingscote A, "The Australian Pictorial Thesaurus 2 Years On" (Paper presented at the DC-ANZ Metadata Conference, Australian National University, Canberra, 26-28 February 2003), http://www.pandora.nla.gov.au/pan/77226/20071011-0000/www.sl.nsw.gov.au/staff/apt2/index.html.

[7] Oates G, "Many Hands Make Light Work" (16 January 2008), http://www.blog.flickr.net/en/2008/01/16/many-hands-make-light-work.

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

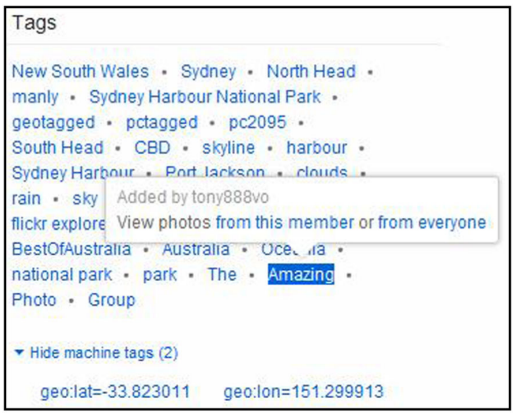
Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

Researchers are beginning to explore the motivations of those who tag.[8] Two key motivations are personal information management and resource sharing. Personal information management requires the selection of tags that are most relevant to the tagger. Resource sharing requires the tagger to identify tags that other potential users (either within a defined audience community or more generally) might use. It seems that users of Flickr have both motivations. These differing motivations are important because they lead to different kinds of tags. Tags for personal use will be more idiosyncratic than those designated with others in mind.

How useful are these user-generated tags? Several large institutions have started to put subsets of their image collections on Flickr for users to annotate. The Library of Congress is one such institution. A recent study of tags added to a sample of the Library of Congress photostream[9] indicated a significant minority of tags were misspelled or irrelevant. However, many of the tags applied were not found in the TGM or the Library of Congress subject headings. Taggers were adding new metadata terms.

FIGURE 3 The descriptive keywords added to an image by the photographer and others



Analysis of comments and conversations associated with a sample of images indicated that taggers were also linking the images to other resources (for example, other photographs, Wikipedia), telling personal stories inspired by the image and discussing issues that the image triggered. Images are "social objects"[10] in that they are artifacts that people share and exchange. These exchanges of objects are also acts of relationship building between participants. A largely unexpected result of digitisation is a move from the private to the collective. The identification and management of images is a social phenomenon, not just a personal one. In his paper *Ontologies are Us*, Peter Mika notes that:

> The Semantic Web is a web for machines, but the process of creating and maintaining it is a social one. Although machines are helpful in manipulating symbols according to pre-defined rules, only the users of the Semantic Web have the necessary interpretive and associative capability for creating and maintaining ontologies.[11]

[8] Heckner M, Heilemann M and Wolff C, "Personal Information Management vs Resource Sharing: Towards a Model of Information Behaviour in Social Tagging Systems" (Paper presented at the International AAAI Conference on Weblogs and Social Media ICWSM, San Jose, California, USA, May 2009).

[9] Stvilia B and Jörgensen C, "Member Activities and Quality of Tags in a Collection of Historical Photographs in Flickr" (2010) 61 *Journal of the American Society for Information Science and Technology* 2477.

[10] Engeström J, "Why Some Social Network Services Work and Others Don't – Or: The case for Object-Centered Sociality" (13 April 2005), http://www.zengestrom.com/blog/2005/04/why-some-social-network-services-work-and-others-dont-or-the-case-for-object-centered-sociality.html.

[11] Mika P, "Ontologies Are Us: A Unified Model of Social Networks and Semantics" (2007) 5 *Journal of Web Semantics* 5. See also McLuhan M, *The Gutenberg Galaxy* (University of Toronto Press, 1962).

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

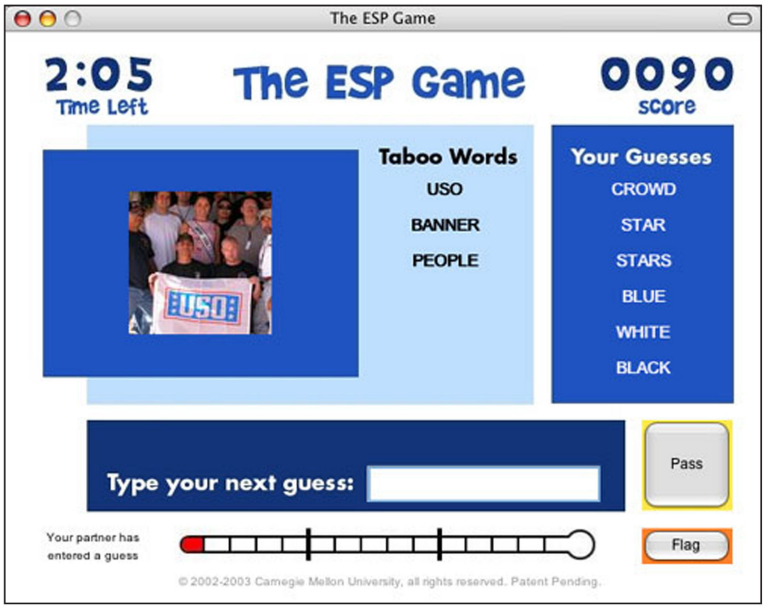
Please email any queries to
LTA.permissions@thomsonreuters.com

USERS: GAME-BASED TAGGING

The tagging of images by creators is one thing but getting someone else to add metadata to an artifact unless they receive an immediate benefit is notoriously difficult. One novel solution to this challenge can be seen in the ESP and Peekaboom games developed by Luis von Ahn.[12]

The ESP Game is an internet-based game. Each round of the game involves two players, neither of whom know who the other is. Each player is presented with the same image. Players are asked to provide words associated with that image. When the players both enter the same word, they are then moved onto the next image. Players have to complete as many images as they can in 2.5 minutes and they receive bonus points for completing 15 images or more in that time.

FIGURE 4 An example of the ESP Game



In playing the game and amassing points, players get to have fun. Meanwhile, the provider of the images gets keywords that have been validated by two users. The variety and quality of the keywords can be increased by the use of "taboo" words and "label thresholds". A taboo word is a pre-identified word that players cannot apply. It is often a common word associated with an image. The labelling of certain descriptors as taboo encourages players to explore broader associations for images. As one image can feature in a number of games, a label threshold describes the number of times a keyword needs to appear before it is accepted as a viable descriptor. The higher the number of appearances, the more likely it is to be widely accepted.

The initial version of the ESP game was run for a four-month period in 2003. A total of 13,630 people played the game during this time, generating 1,271,451 labels for 293,760 different images. In 2006, Google licensed von Ahn's technology to create the Google Image Labeler.[13] Unfortunately, Google has released little information about the success or otherwise of this application. It appears that some players began to game the system with irrelevant tags, such as "carcinoma", "congenita" and "diphosphonate", however this was detected and prevented.

[12] von Ahn L and Dabbish L, "Labeling Images With a Computer Game" (Paper presented at the ACM Conference on Human Factors in Computing Systems, Vienna, Austria, 24-29 April 2004); von Ahn L and Dabbish L, "Designing Games With a Purpose" (2008) 51 *Communications of the ACM* 58.

[13] See Wikipedia, http://www.en.wikipedia.org/wiki/Google_Image_Labeler.

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

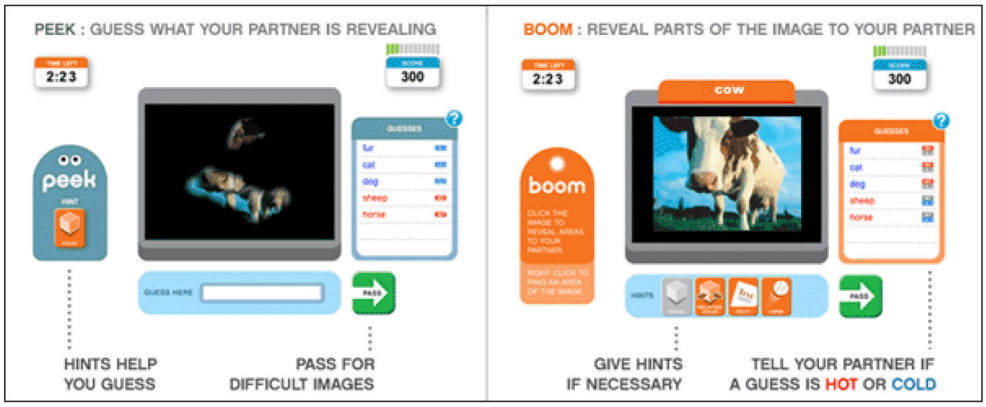
Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

The Peekaboom game takes the gaming principles of the ESP game and uses them to tag specific areas of an image. This is also a two-player game. One player (Boom) starts with an image and an associated tag (identified by the ESP game). The other player (Peek) starts with a blank screen. The goal of the game is for Boom to reveal parts of the image to Peek, so that Peek can guess the associated word. Boom reveals circular areas of the image by clicking. A click reveals an area with a 20-pixel radius. Peek, on the other hand, can enter guesses of what Boom's word is. Boom can see Peek's guesses and can indicate whether they are hot or cold. When Peek correctly guesses the word, the players get points and switch roles.[14]

FIGURE 5 An example of Peekaboom



The ESP and Peekaboom games demonstrate that users can be persuaded to add descriptive metadata to large image sets as entertainment. Ingenious game design can raise the quality of added tags. However, it remains uncertain as to whether this approach is scalable and sustainable. It is reliant on a pool of users who find the game enjoyable. "Gamefication" (or the introduction of game mechanics in an attempt to solve social and business problems) is a currently fashionable topic. Who will play all these games?

MACHINES

While human beings find image interpretation comparatively easy, it has proved surprisingly difficult for computers. Instructing a computer to search for the word "ball" in a portion of text is far easier than training a machine to recognise a ball in a photograph. The actual meaning of "ball" might vary from one context to another but the text string is the same. The pattern of pixels that represents a ball is not. Visual recognition and processing is an activity that can be summarised by Moravec's paradox: "it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."[15]

Nevertheless, progress is being made in this area. One example of this is Imense, a Cambridge UK-based company. Imense uses a combination of content-based image retrieval (CBIR) technologies to process images.[16]

[14] von Ahn L, Liu R and Blum M, "Peekaboom: A Game for Locating Objects in Images" (Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Montréal, Québec, Canada, 22-28 April 2006).

[15] Moravec H, *Mind Children* (Harvard University Press, 1988) p 15.

[16] Town CP, "Giving Meaning to Content through Ontology-based Image Retrieval" (Paper presented at the ISKO Conference on Content Architecture: Exploiting and Managing Diverse Resources, London, England, 22-23 June 2009); Town CP and Harrison K, "Large-scale Grid Computing for Content-based Image Retrieval" (Paper presented at the ISKO Conference on Content Architecture: Exploiting and Managing Diverse Resources, London, England, 22-23 June 2009); Oonk M, "A Cambridge Team That's 2 Years Ahead of Google", *Fast Media Magazine* (5 October 2009), http://www.fastmediamagazine.com/blog/2009/10/05/a-cambridge-team-thats-2-years-ahead-of-google; Oonk M, "Google and Imense on Visual Search at PACA 2009", *Fast Media Magazine* (2 November 2009),

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

Image segmentation

The image is automatically segmented into a covering set of non-overlapping regions, and sets of properties such as size, colour, shape and texture are computed for each region. This yields salient parts of the image corresponding to objects or object parts. The number of segmented regions depends on image size and visual complexity, but has the desirable property that most of the image area is usually contained within a few dozen regions which closely correspond to the salient features of the picture.

Region classification

Segmented regions are then automatically classified in a predefined set of material and environmental categories, such as "grass", "sky", "wood", "water" etc. Statistical machine learning methods are employed to yield a highly reliable probabilistic classification of the image. This may be regarded as an intermediate level semantic representation which serves as the basis for subsequent stages of visual inference and composite object recognition.

Scene classification

A second stage of classifiers is applied to analyse image content at a higher scene level. Examples of scene categories include "indoor", "beach", "sunset", "nighttime", "autumn" etc.

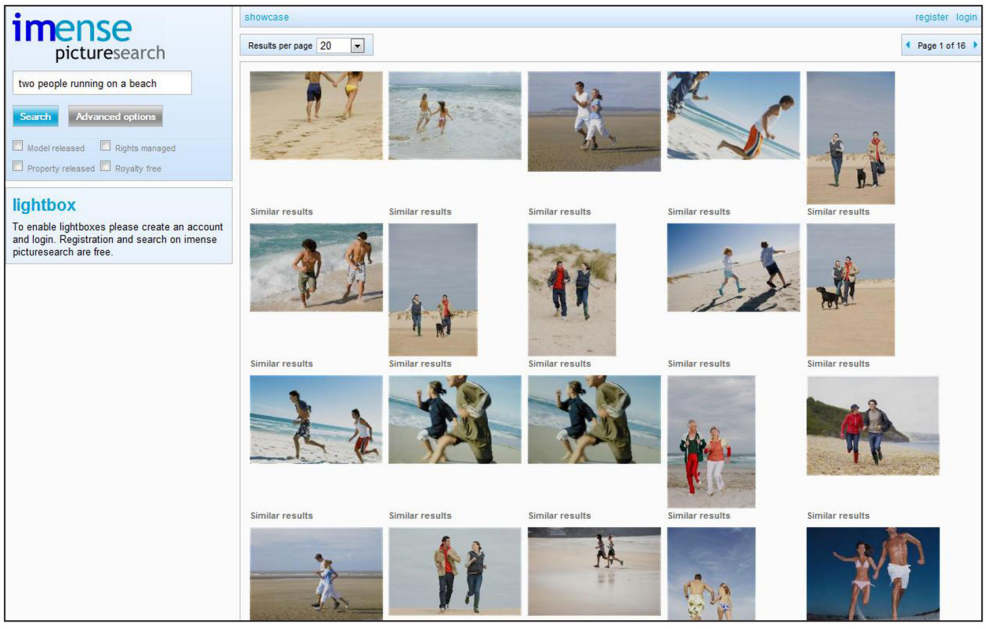
Object detection and recognition

The image analysis also features detectors for common objects. For example, human faces are automatically detected and classified according to personal attributes such as gender, age and facial expression.

Index generation

Once all image analysis stages have information from the various classifiers and recognisers, it is combined into a special indexing format which supports fast content-based image retrieval.

FIGURE 6 An example of an image search using Imense



http://www.fastmediamagazine.com/blog/2009/11/02/google-and-imense-on-visual-search-at-paca-2009.

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

This approach means that the composite elements of an image can be specified – for example, a picture of one person in the distance on a beach may be "beach" 98%, "sky" 70%, "person" 5%. The ontology underlying the image analysis also allows the system to make inferences. It cannot identify a camel purely graphically but the ontology links the concept of "camel" to the concepts of "animals with fur" and "animals that live in deserts". The image analyser can identify fur and deserts in images and can infer that a camel may be present.

CONCLUSION

At the moment, non-textual materials pose a special challenge to information managers. When I was studying librarianship in the mid-1990s, photo libraries were a topic slightly outside the mainstream. Corporate information managers have primarily dealt with textual documents in the last decade. Yet we are creating and distributing images at a rate that is hard to fully conceive. The canny information manager cannot wait for the perfect solution to be developed (it may not exist) but rather has to grapple with the current reality. Therefore, an information manager needs to be aware of tools they have at their disposal and the strengths and weaknesses of those tools.

When designing a knowledge organisation system for images, established metadata schemes and thesauri for images provide a good place to start. However, it is important to understand the origins and initial purpose of each scheme. This will determine whether it is suitable for your use. It may also indicate changes that will have to be made for this use to be as effective as possible.

When applying these to an image set, there are inevitable trade-offs between the quality of metadata applied and the effort expended. Higher quality metadata will require more effort. The first questions should be: "what quality level do you need" and "how good is 'good enough'"? Each field is an additional overhead unless its collection can be automated.

The next questions are then "who adds this metadata" and "how will this addition benefit them"? If they are the creator of the image and they are tagging for their own personal information management purposes then the outputs will be different to a situation where they wish to share the material. Creators may be amateurs or professionals – and with the arrival of online "microstock" sites, that line is blurred. Would they use a controlled vocabulary or is that asking too much of them?

Tags added by those other than creators can add value to images. Again, questions of motivation are important. Should it just be left to experts? In case of sensitive images or specialist domains (for example, medical imaging), the answer may be "yes". However, many institutions, from the Library of Congress to Sydney's Powerhouse Museum, have opened up their image collections to the broader public. However, this then raises further questions. If you have many people tagging, then is consistency an issue? It is then that gaming platforms and other techniques that incentivise people to collaborate in new ways become useful.

Finally, to what extent can automated metadata and retrieval play a role? These tools are still developing. They do not do intangibles well. We want images that describe a mood rather than ones that just contain specific objects and our software cannot feel for us. While we cannot expect them to replace people just yet, we can now expect software to assist us in the retrieval and analysis of images.

APPENDIX 1: IPTC CORE METADATA

Photo metadata: IPTC core

Contact section

- Creator
- Creator's job title
- Contact info (address, city, state/province, postal code, country, phone, email, website)

Image section

- Date created
- Intellectual genre
- IPTC scene code

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com

- Geographic fields

Content section
- Headline
- Description
- Keywords
- IPTC subject code
- Description writer

Status section
- Title
- Job identifier
- Instructions
- Credit line
- Source
- Copyright notice
- Rights usage terms

THOMSON REUTERS

© 2011 Thomson Reuters (Professional) Australia Limited
for further information visit **www.thomsonreuters.com.au**
or send an email to **LTA.service@thomsonreuters.com**

Please note that this article is being provided for research purposes and is not to be reproduced in any way. If you refer to the article, please ensure you acknowledge both the publication and publisher appropriately. The citation for the journal is available in the footline of each page.

Should you wish to reproduce this article, either in part or in its entirety, in any medium, please ensure you seek permission from our permissions officer.

Please email any queries to
LTA.permissions@thomsonreuters.com