COMPARATIVE EVALUATION OF WEB IMAGE SEARCH ENGINES FOR MULTIMEDIA APPLICATIONS

Keon Stevenson
Clement Leung
School of Computer Science and Mathematics
Victoria University, Footscray, Vic 3011, Aus
keon182@yahoo.com
clement.leung@vu.edu.au

ABSTRACT

While text-oriented document searching are relatively mature on the Internet, image searching, which requires much more than text matching, significantly lags behind. The use of image search engines significantly enlarges the scope of images to users accessibility. This paper provides an understanding of current technologies in image searching on the Internet, and points to future areas of improvement for multimedia applications. We develop a systematic set of image queries to assess the competence and performance of the major image search engines. We find that current technology is only able to deliver an average precision of around 42% and an average recall of around 12%, while the best performers are capable of producing over 70% for precision and around 27% for recall. The reasons for such differences, and mechanisms for search improvement, are also indicated.

1. INTRODUCTION

There are a number of different types of Image Search Engines (ISEs) currently in use in terms of technology [1, 2, 6-8, 10, 11]. Currently most major search engines fall into three types:

- ISE such as Google and Yahoo are of basic design having large databases and relying on mixed retrieval for accuracy. The technology is very basic; it is indexed in terms of text label/words coupled with an image.
- Specialized ISEs that are dedicated to indexing images or multimedia such as Corbis & Getty Images. These sites are often experimental and have limited databases restricted by size when compared with enormous sites such as Google.
- Finally there are image search engines that are called Meta-Search engines, which send users requests to multiple search engines and then display the 'multiple' results.

Content-based systems use technologies that refine searches by focusing on content and objects within the image rather then text. Visual-meta tagging is another method like content-based retrieval that organizes images into groups of relevance using their visual content. Finally, whilst most current images search technologies on the Internet are based on 2D images, it is worth mentioning that a method that is being developed [4,5,9] for 3D image search that uses algorithms and mathematical models of each shape to determine image content.

Whilst current major image search engines in use are limited in terms of high functionality (precision and recall), the need for a performance evaluation of current engines will be of great benefit to users. It will provide them with an understanding of current technologies, advanced searches and ways of making use of the results for specific multimedia applications.

2. THE SEARCH MECHANISMS

With most ISEs, the options available require that an online user will type in a keyword search and it will match the keywords and present them in a thumbnail manner with the appropriate links. By specifying the file type the ISE can detect the tag associated with the file e.g. JPEG or GIF. In matching a keyword search with a file type an ISE can approximate the content of the file. Engines look for and index online Web sites where the titles of the files accurately describe the content of the image. In using such a method it can lead to error but it also means it can support a larger database as they are catalogued more easily rather than looking at its actual content. As an example, in a search for "cat", the first few result pages will be relevant but it will also match other images that include cat i.e. "cat scan" "different felines" and so on. Other strategies that can assist an ISE can be human involvement where they search the World Wide Web manually and look for images and catalogue them according to their relevant content. In cataloguing images this way more accurate results are achieved but as there are millions of images on the World Wide Web it requires very intense labour to do so. A positive aspect of current image search engines is the way in which they display results from a search. The results are displayed in thumbnail form with an associated link of where the image is originally located and some file information i.e. size and date.

When text accompanying an image has little relevance to the content they still can be indexed with the use of ALT tagging which is now a requirement in the US, with the Americans with Disabilities Act [6]. ALT tagging is known as alternative text, which is method used to actively describe an image through text. This allows users that are visually impaired to use voice activation software that informs them with the use of ALT tagging within an image.

If a keyword search within an image search engines is broad then current systems work quite well. Problems begin to surface as searches become more specific.

0-7803-9332-5/05/$20.00 ©2005 IEEE

3. EVALUATION METHODOLOGY

In testing the search engines it is necessary to ensure that each ISE is tested to its fullest capacity, most engines have an advanced link to specify different criteria to refine searches. In determining each ISE's capabilities, the method used to determine just how good each ISE is. It involves 12 or so objects in the query, ranging from easy words like cat to more difficult searches using multiple words and finally progressing to specific searches that are uncommon.

In determining the results, the precision and recall are measured. *Recall* gives the ratio of the number of relevant records retrieved to the total number of relevant records in the database. Measuring the recall presents a challenge because here we are dealing with open image collections rather than traditional closed ones, and the number of images can be regarded as potentially infinite. We estimate the recall is follows. Let R_i be the set of relevant images relating to database i, with $| R_i |$ denoting the number of images in the set. The recall for ISE i is estimated by the formula

$$\text{Recall of ISE } k = \frac{|R_k|}{|R_1 \cup R_2 \cup \ldots \cup R_N|}$$

where N is the number of ISE's under evaluation.

Precision gives the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. In evaluating the results a good method is to find the relevant results from first 400 results of a query greater than 400 images. The reason for this is that it was found that after 400 images of a query the results are not related to the subject matter at all or very little. For pages less than 400 images, the precision and recall for all those images retrieved are determined.

The test queries performed range from broad searches to more specific uncommon searches, and the same queries are used to evaluate all the different engines.

One Word Test Queries
- Test 1 – "cat"
- Test 2 – "foot"
- Test 3 – "basketball"
- Test 4 – "quiksilver"

Two Word Test Queries
- Test 1 – "cat scan"
- Test 2 – "australia victoria"
- Test 3 – "university life"
- Test 4 – "football shoes"

Three Word Test Queries
- Test 1 – "tool hammer drill"
- Test 2 – "basketball shaq lakers"
- Test 3 – "football shoes helmet"
- Test 4 – "golf tiger nike"

3.1 Test Image Search Engines

The ISEs that we have tested include the following.

General / Major Image Search Engines
- Google (www.google.com)
- Yahoo (www.yahoo.com) *

Specialized Image Engines
- Ditto (www.ditto.com)
- Corbis (www.corbis.com)
- WebSeek (www.ctr.columbia.edu/webseek)
- Getty Images Creative (http://creative.gettyimages.com/)
- Picsearch (www.picsearch.com)
- Ithaki (http://images.ithaki.net/)

4.RESULTS / COMPARISIONS

4.1 Precision

One Word Test Queries

The average precision for all image search engines was 55%. Figure 4.1 illustrates which image search engines give the best the precision when being tested with a one-word query.

Two Word Test Queries

The average precision for all image search engines was 50.6%. Figure 4.2 illustrates which image search engines has the best the precision when being tested with two-words in a test query.

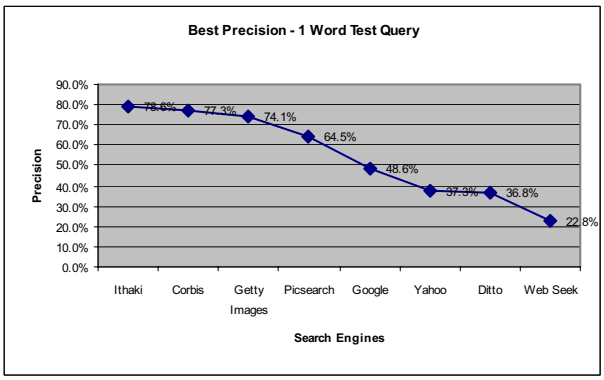

Figure 4.1 – Best Precision: 1 Word Test Queries

Three Word Test Queries

The average precision for all image search engines was 20.7%. Figure 4.3 illustrates which image search engines give the best the precision when being tested with three-words in a test query.

Best Image Search Engine: Precision

The overall average for precision amongst the search engines was 42.1%. Figure 4.4 illustrates that Corbis keeps on proving its system to be the best, showing the best precision with its large database of catalogued images. The use of good advanced search features and well catalogued images via descriptive names make it the best system. Getty Images comes in 2nd with excellent precision if a word or query exists in its database. Ithaki proves that bringing existing search

* Note: Yahoo's image search engine is also used by All The Web (www.alltheweb.com), Lycos (www.lycos.com), AltaVista (www.altavista.com.au).

engines together works, with an above average overall precision. Part of the reason for this is that it only displays small total results for a query, hence the good precision. Picsearch keeps proving itself as a solid system with just above average precision, through the use of simple, easy to use advanced features make it a very useable system. Google and Yahoo are a little disappointing, Ditto and Web Seek round out the remaining image search engines with very poor precision, which is due to a number of factors, which include limited database of catalogued images and basic or non-existent advanced search features.

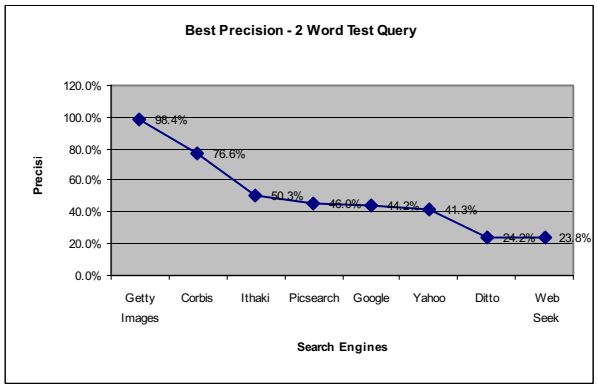

Figure 4.2 – Best Precision: 2 Word Test Queries

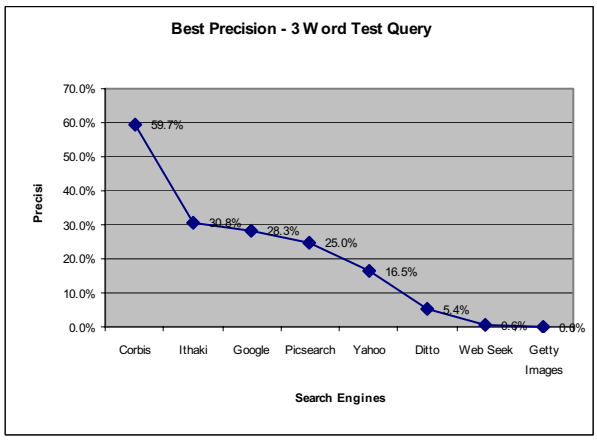

Figure 4.3 – Best Precision: 3 Word Test Queries

4.2 Recall

One Word Test Queries

Figure 4.5 illustrates which image search engines gives the best the recall when being tested with a one-word query. Getty Images, Corbis, Yahoo and Google had the majority of the recall due to their great precision and mainly because they have larger databases.

Two Word Test Queries

Figure 4.6 illustrates which image search engines give the best the recall when being tested with two-words in a test query.

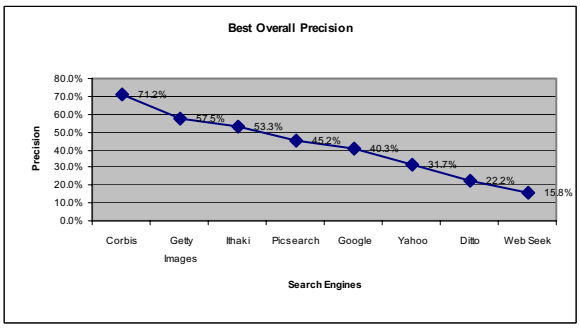

Figure 4.1.4 – Best Image search engine: Precision

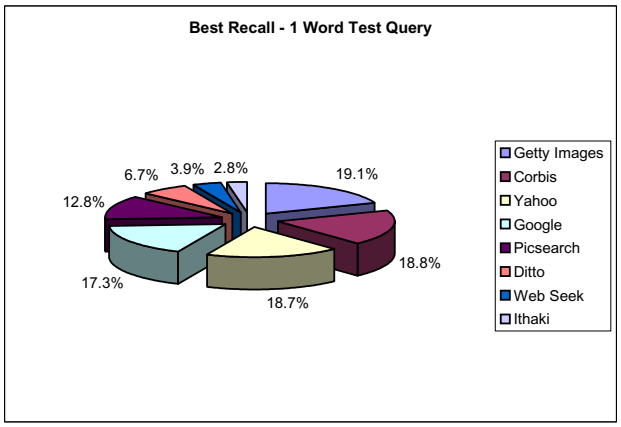

Figure 4.5 – Best Recall: 1 Word Test Queries

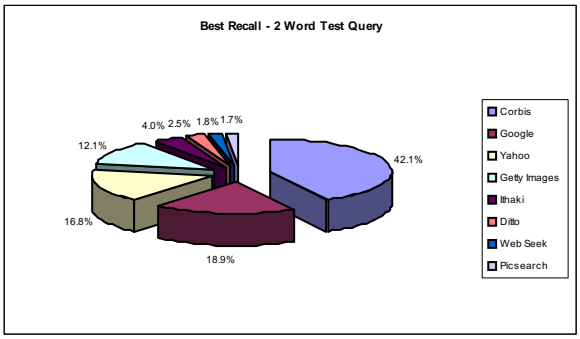

Figure 4.6 – Best Recall: 2 Word Test Queries

Three Word Test Queries

Google is now beginning to shine and show why it has the largest index and widest range of images available. Figure 4.7 illustrates which ISEs had the best the recall when being tested with three-words in a test query.

BEST IMAGE SEARCH ENGINE: RECALL

Google and Corbis have the majority of the recall due to their high precision as shown in Figure 4.7 (images retrieved that are relevant divided by the all relevant images). Google has produced excellent recall mainly due to the fact that it has the most extensive database of catalogued images (880 million), and the fact that it performs well in a refined search with 3 words in a query or more. Corbis has a large portion of recall due to the fact that it performs searches quite well under all conditions of searches especially with a two word query as the images a catalogued and refined using two words.

Yahoo comes in third which should be expected as it has the 2nd largest database (630 million), it also has good advanced features hence the third position. Getty Images would have had great precision if all the terms in the test queries existed in their database; it also failed at finding any 3 word queries. The remaing four image search engines (Ithaki, Picsearch, Ditto and Webseek) fail to show any significant recall due to the fact that they have much smaller databases and advanced features to refine searches.

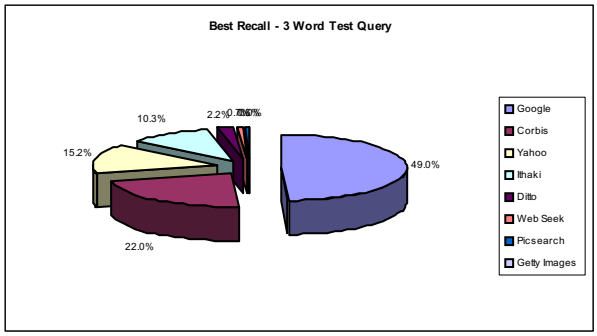

Figure 4.7 – Best Recall: 3 Word Test Queries

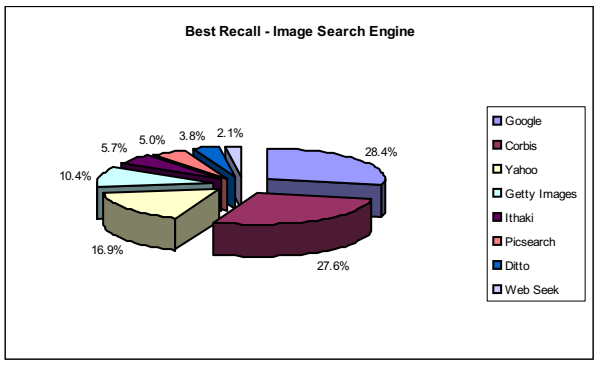

Figure 4.8 – Best image search engine: Recall

Table 4.1 summarizes the overall performance of the ISEs evaluated. A General Evaluation Criteria was also created and separated into the following areas and rated. These tend to be more qualitative and include the following:

- system range,
- system advanced features,
- system performance and
- system presentation.

For each criterion listed below a score from 1-10 is given with its associated reasons and methodologies. Further details of this is may be found in [3].

5. SUMMARY & FUTURE DIRECTIONS

In most image search engines, images are not indexed by their appearance but by text, which can be found in the context of the image. Current search engines are technological basic using keyword searches that accompany an image. We develop a systematic set of image queries to assess the competence and performance of the major image search engines. These queries are human oriented as they aim to retrieval contents which are humanly meaningful as opposed to machine-oriented features such as colour, textures, and shapes. We find that current technology is only able to deliver an average precision of around 42% and an average recall of around 12%, while the best performers are capable of producing over 70% for precision and around 27% for recall. With limitations in current automatic image recognition technology, The key to making a successful ISE will require more human involvement and model-based indexing of image contents.

Search Engine	General Evalution Criteria	Precision	Recall	Best Overall
Corbis	83.0%	71.2%	27.6%	**60.6%**
Getty Images	76.0%	57.5%	10.4%	**48.0%**
Google	74.0%	40.3%	28.4%	**47.6%**
Yahoo	70.0%	31.7%	16.9%	**39.5%**
Picsearch	67.0%	45.2%	5.0%	**39.1%**
Ithaki	42.0%	53.3%	5.7%	**33.6%**
Web Seek	54.0%	15.8%	2.1%	**24.0%**
Ditto	43.0%	22.2%	3.8%	**23.0%**
AVERAGE	**63.6%**	**42.1%**	**12.5%**	**39.4%**

Table 4.1. Summary of Overall Performance

6. REFERENCES

[1] Image Retrieval Projects, 2004, Retrieved: 27 May, 2004, from http://compass.itc.it/links.html

[2] J.R. Griffiths, F.C. Johnson, R.J. Hartley, 'A framework for the evaluation of Internet search engines', Devise, The Council for Museums, Archives and Libraries, pp1-58, 2001

[3] K. Stevenson. 'Performance and Comparison of Image Search Engines' Honours Thesis, Victoria University, Melbourne, 2004.

[4] P. Min, 'A 3D Model Search Engine', A Dissertation Presented to the Faculty of Princeton University in Candidacy for the Degree of Doctor of Philosophy, pp. 1-158, 2004.

[5] P. Shilane, P. Min, M. Kazhdan, T. Funkhouser, 'The Princeton Shape Benchmark', Shape Modeling International, Genova, Italy, pp. 1-12. Retrieved June 13, 2004 from, 2004.

[6] RLG Digi new, 2001, Retrieved: 23 May, 2004, from http://www.rlg.org/preserv/diginews/diginews5-6.html#faq

[7] S. Sclaroff, 'World wide web image search engines', BU CS Technical Report, No. BU TR95-016, pp.1-2, 1995.

[8] Technical Advisory Service for Images, 2003, Retrieved: 10 May, 2004, from http://www.tasi.ac.uk/resources/searchengines.html

[9] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, D. Jacobs, 'A Search Engine for 3D Models', ACM Transactions on Graphics, 22(1), pp. 1-28, 2003.

[10] T. Gevers and AVM Smeulders, 'Image search engines An Overview, Emerging Topics in Computer Vision' Prentice Hall, pp. 1-54, 2004.

[11] Y. Wen, C. Meadow, G. Markowsky, 'A Search Engine for Images' Project Report, Department of Computer Science, University of Maine, 2003.