Archive of SID

International Journal of Information Science and Management
Vol. 16, No. 2, 2018, 43-59

Improved Retrieval of Online Images via Site Operator: Findings at a Website Level

Hamid Keshavarz
Assistant Prof., Department of Information Science and Knowledge Studies,
Semnan University, Iran
Corresponding Author,
h_keshavarz@semnan.ac.ir

Saleh Rahimi
Assistant Prof., Department of Information Science and Knowledge Studies,
University of Razi, Kermanshah, Iran
s.rahimi@razi.ac.ir

Abstract

The current paper aimed to explore how image indexing and annotating could improve image retrieval via site operator command. Also was among the goals of the study to compare the effectiveness of different codes assigned to sample images in retrieval ranks of images by Google search engine. Using quasi-experimental method 100 images were selected, each image was uploaded 9 times by concept-based characteristics on site iiproject.ir. Analysis consists of images which retrieved from the site operator command. Number of images retrieved by the site operator command is 151 images of total 900 that are used in the study. The minimum number of retrieved images is related to "image titles" and the maximum to the criteria images which entitled with Q code. Chi-square statistics showed that the number of images retrieved in various codes was different. The best ranking is related to "image title" and the weakest to "image caption in Farsi". Images average ranking retrieved in 9 groups were different. Findings reflect problems and issues of image indexing and retrieval and put forward some ways to overcome the challenges identified. Some lacks in image retrieval by Google search engine at website level are identified. Different codes and descriptors show different retrieval ranks and results considerable for designers, indexers and even users.

Keywords: Visual Communication, Image Indexing, Image Storage and Retrieval, Concept-based Image Indexing, Site Operator, Google Search Engine.

Introduction

Images are among the most searched and used data available on the web. Image searchers face a double challenge when searching for an image using a textual query. Firstly, the query terms must correspond to the text associated with the image. Secondly, the language of the query must match the language of the text associated with the images (Ménard, 2009). As a result of increasing advancements in multimedia technologies, visual data systems have been developed in industrial and research areas. Government and educational institutes, museums, and commercial sectors have created databases full of images. Images are stored in digital format in various fields and areas including medicine, geography, law enforcement, art, aerospace, journalism, and media communications.

Image retrieval has been largely studied on two research communities including Database Management and Computer Science since 1970s (Fauzi and Lewis, 2008). Growing power of computers and presence of tools with high storage volume allows storage of high volumes of images. Most users are interested in semantic entities rather than in visual representations. An

www.SID.ir

image in the web is especially surrounded by semantic issues such as image title, image alternate text, image caption, page title, and metadata (Jayaratne, 2006; Patil, and Durugkar, 2015). Griesdorf and O'Connor (2002) argue that human beings evaluate images according to three levels and assign meaning to them. First level includes color, shape, and texture of the image. Second level is the subject, and includes people, locations, and events in the image. Third level, which is the most complex part, includes inferential interpretations of the image, and it is here that image viewer's subjective topic is formed.

Such issues as lack of coherent metadata for images, poor accuracy of image search engines in the web, and lack of user understanding in web image searching have caused people to perceive their favorite image content with difficulty (Lee and Neal, 2010; Patil, & Durugkar, 2015) and they just search for name or time of images key terms. Since image titles generally do not provide descriptive information about the document content, the users describe their images themselves. The best way for image retrieval is relaying on textual descriptions (Bar-Ilan Zhitomirsky-Geffet, & Shoham, 2012) and level and extent of indexing is often determined by the nature of collections and the user needs (Booth, 2001). If the intention is identifying the image, everything about it is described by the terms except color, shape, and context. Images often are used not only for indicating a specific object, but also for expressing specific feelings (Westerveld, 2000). Images contain more semantic layers compared to the text, because every image is both "From something" and "About" something, and there is usually difference between "From" and "About". Rapid advancement in communication and information technologies has led to the increasing use of visual resources more than ever.

When searching an image, it is likely to retrieve an image from a personal or organizational collection or on the web, but we may face a problem or fail in retrieving that image by just using common words, conceptual or even with the name of that image. It is clear that several factors may be involved in this failure such as content, text and keywords in storing and retrieving images, inefficiency of search engines or inability of users to retrieve images. On account of current importance of images, some people believe that our generation emphasis on texts and wittings; but our children emphasis on image due to the technology progresses (Vadivel, Sural & Majumdar, 2009; Patil, & Durugkar, 2015).

Researchers in CBIR (Content Based Image Retrieval) identify low level features color, texture, shape, statistical parameters and high level features like semantics, fuzzy logic for interactive image understanding. The basic idea of image retrieval by image example is to extract the characteristic features from target images which are then matched or compared with that of the query image. These features are typically derived from shape, texture, color properties or statistical attributes of the query and the target images. After matching, the images are ordered with respect to the query image according to their similarity measure and are displayed for viewing

Problem Statement

The purpose of the Worldwide Web like any other information system is quick access to related resources. Next to creation and recognition the web, retrieval and ranking the sources by search engines have been the most important issues for users and service providers of this global network. Given that we are living in a meta-visual age, transmission of a large amount of information is done by eyes and non-verbal tools (Mills, 2011). In some fields such as architecture and medicine, information which are transmitted via image is often more comprehensive than when transmitted via the text (Grauman, 2010). This days, images are

www.SID.ir

considered as the main media on the web, but in spite of books and periodicals, images don't have page title or other bibliographic information (Lee and Neal, 2010). Image databases are becoming more and more important in everyday life; therefore, there should be appropriate methods and techniques to enable users to uploading images in digital image databases and retrieving them.

Some factors may be involved in failure to retrieve an image including content or textual and keywords elements in image storage and retrieval, inefficiency of search engines, or inability of users in image seeking. This study considers factors such as text and keywords in indexing of images. Current research considers textual and keyword factors could be assigned to given images. It attempts to investigate reasons for failure and success in image storage and retrieval in conceptual and textual aspects. Findings of the research maybe used by the users, searchers, and designers of image storage and retrieval systems as well as indexers who act in related areas. Research findings may reflect problems and issues of image indexing and retrieval and ways for overcoming the problems and issues.

In this study, we have tried to investigate the reasons of failure and ways of success in storage and retrieval of images from conceptual aspects. The results of this study can be used by users, storage and retrieval systems designers and also indexers who work in the field of image storage and retrieval. Regarding that today, many people, organizations and institutions use images for different purposes; findings of this research may be generalizable to their websites and reflect the problems, issues of indexing and retrieving images and the ways of coping with them.

Literature Review

Image search and retrieval is an important and much-used aspect of the search engine market, and works on optimizing images and their metadata for indexing and retrieval is relatively limited. There are different guidelines for image publishing on the web all of which are not necessary rules but recommendations for better image publishing. These guidelines are considered as SEO (Search Engine Optimization) which are user's experiences not indexing and retrieval algorithms of search engines like Google. There appears some research concerning text-based image retrieval methods where the text associated with an image is used to determine the image contents.

Chua, Pung & Jong (1994) studied design and implementation of concept-based image retrieval system. The main cases in this system included: using a concept based search engine for proper retrieval of concepts and images, using concept groups in indexing and updating images, and using relevance feedback to update representation of concepts and image descriptions. They tested this system in history field in Singapore using a database including 120 images. Their research findings showed coherent progress is probable in system retrieval with various types of searches.

Jang (2002) studied indexing and image retrieval using conceptual analysis. In this research, a new technique of indexing and image retrieval is proposed based on formal concept analysis (FCA) which allows quick image retrieval from databases. Efficient retrieval in this plan depends on the number of properties rather than number of images existing in the database and dynamic support for increasing new images. However, it also requires advanced knowledge in a specific area.

Azzam, Leung & Horwood (2004) focused on implicit concept-based image indexing and development of a method for indexing and image retrieval. They provided a method which enables classification of image details based on their relative importance. Image storage is

www.SID.ir

based on an implicit indexing plan rather than explicit one. Then, image retrieval is influenced by an algorithm based on this classification, which allows related images to be determined and retrieved properly and efficiently.

Smits, Plu & Bellec (2006) studied semantic indexing of personal images using textual descriptions and proposed an automatic indexing area for continuous image management software. Semantic descriptors are extracted from text descriptions along with personal images. The users naturally write annotate for their images using natural language so that major elements of the images are personally described. The main aim of the research was retrieval of individuals and locations which are given directly or indirectly in the captions alongside the textual descriptions. The system attaches new images automatically to the user text considering extracted descriptors.

Ferecatu, Boujemaa & Crucianu (2008) addressed challenges of semantic distance reduction for image retrieval via active relevance feedback of representation and retrieval of visual and conceptual content. They introduced a new axis – vector property for the terms along with the image in a collection of key concepts using external lexical database. Firstly they optimized data transfer between the user and system, and then proposed an active learning selection criterion which reduced redundancy between the selected images to the user. Experimental evaluations indicate order of using new axis – vector property and visual properties with the relevance feedback efficiently improves quality of results.

Rorissa (2008) studied user descriptions about personal images in comparison with tags of image groups using Basic level Theory. Digital libraries and search engines often rely on textual search and descriptions for provision and retrieval of image conceptual content. Thus, the users yet tend to use the text for explain images and search formulation, while images with or without text are visual information sources. The researcher used Basic Level Theory as framework for comparing descriptions of individual images and tags determined for image groups by 180 participants in three studies, and they found there is significant difference in their abstract level. Finally, some concepts for designing search interfaces of images, classifications, thesauruses, and similar tools were discussed.

Vadivel, Sural & Majumdar (2009) aimed at providing dynamic method for attaching low level properties of images in order to utilize their complementary capacities in image retrieval from the web using multiple characteristics. They investigated image meanings using lexical low level properties. Findings suggest results of image retrieval using various techniques proposed by the authors were improved. Experiential findings showed attachment method provides better results compared to concept-based and content-based retrieval techniques.

Vrochidis, Moumtzidou & Kompatsiaris (2012) introduced a method for automatic concept extraction which explains image content of patents. For evaluation of this method, a database on footwear domain was chosen and some conceptual specialists with different form compounds were trained. Research findings show combination of textual and visual data of patent images suggest higher efficiency of visual and textual features combination. Results of this text suggest the fact that determining the concept can be used in patent image retrieval domain, and it may be complementary in real world applications in favor of research in patents domain.

Fauzi & Belkhatir (2013) discussed a user-based plan of automatic multifaceted concept-based indexing framework which analyzes meaning of textual data of web images. They categorize the plan into five conceptual major semantic conceptual facets including signal, object, abstract, scene and relational which and determines conceptual meaning between concepts. Results of testing web image collection described by human and associated

www.SID.ir

Archive of SID

textual data denote that this method outperforms related frameworks like TF-IDF experimental frameworks and location-based TF-IDF weighing plans as well as N-gram indexing in a recall/precision based evaluation framework.

Previous studies indicate methods used in indexing and image retrieval area in two past decades have faced many changes. In comparison to content-based indexing, research in concept-based indexing of images is developed by Information Science researchers. Current work attempts to investigate importance of using text concepts, image title, image alternate text, and caption in increasing rate of image retrieval using Google search engine which wasn't taken into consideration in previous works.

Research Questions

Current research aimed to answer the following questions:

A. Does use of controlled language make any difference in extent of image retrieval in research sample?

B. Does use of free language make any difference in extent of image retrieval in research sample?

C. Does use of file name make any difference in extent of image retrieval in research sample?

D. Does use of image title make any difference in extent of image retrieval in research sample?

E. Does use of image alternate text make any difference in extent of image retrieval in research sample?

F. Does use of image caption in Persian make any difference in extent of image retrieval in research sample?

G. Does use of image caption in English make any difference in extent of image retrieval in research sample?

H. Does use of property tag including subject and title make any difference in extent of image retrieval in research sample?

Research Methodology

Current work is an applied research and we focused on image retrieval method via site operator. Technology-based research method (Powell, 1997) and quasi-experimental method (post-test plan with case and control group) were used. Using site operator command, images of Ahwaz Shahid Chamran University website (http://www.scu.ac.ir) were searched in September 15, 2015. Using site operator, all images available in a specific website can be retrieved, and images out of search website is prevented. Retrieved images were stored on personal computer. Google shows 1,000 initial documents for each search by default. Thus, 1,000 images were retrieved with site operator command. 100 images were selected as research sample.

Considering that research sample size was selected based on conceptual indexing i.e. individual, subjects, objects, and texts around the image and author's observation from the population, purposive sampling was used. Using FastStone Photo Resizer software, image resolution was set as 640*480 which is suitable standard for website. Then, for surface distinction, some codes were assigned to images so that they can be distinguished in their retrieval from Google search engine. The codes assigned to images were placed in bottom left right side of images.

www.SID.ir

Archive of SID

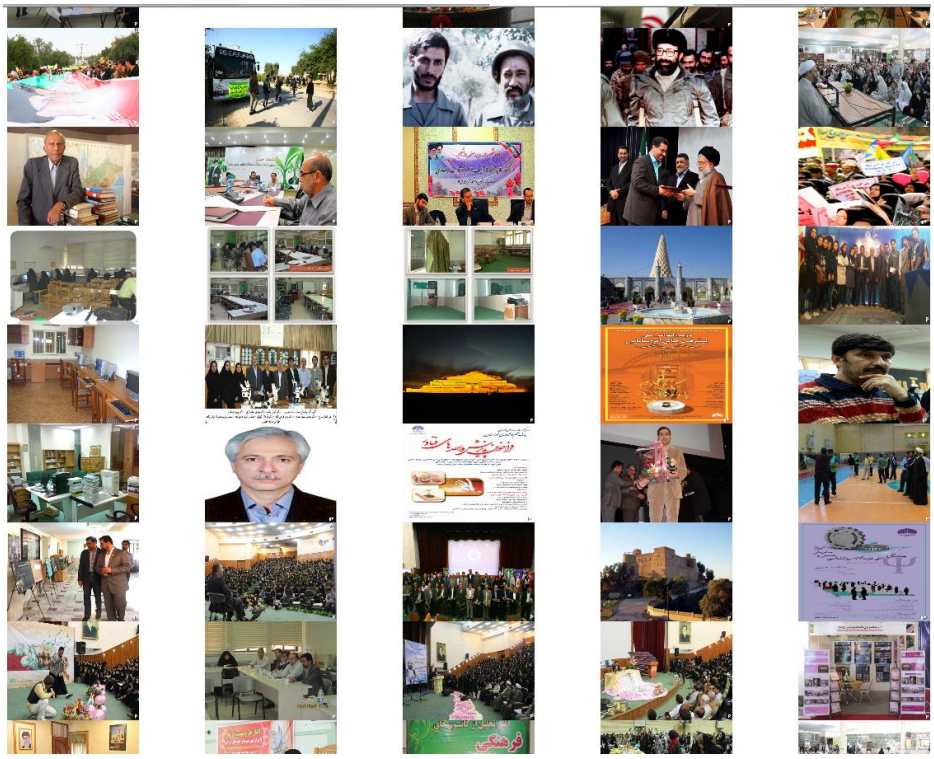

Figure 1. A part of a page of the nine pages used for the study available at: http://iiproject.ir/PropertiesP.html (accessed 25 July, 2016)

Since the research design is post-test with control and quasi-experimental type, minimum 30 images suffice for each group (Powell, 1997). However, since this work is conducted for the first time in Iran, it was attempted to use higher numbers of images. Thus, research sample of this study include 100 images and each image was loaded nine times, and overall 900 images were loaded on a specific research website with URL iiproject.ir which is available right now (May 7, 2016). Special conceptual properties were considered for each image and in fact each image had different conceptual characteristics and properties.

Following loading images on the allocated website, and their indexing by Google search engine, site operator command was used so that all images of research sample were retrieved at once by Google. Hence, order of placement of images is specified. That is, properties which are more important for Google are placed at higher ranks.

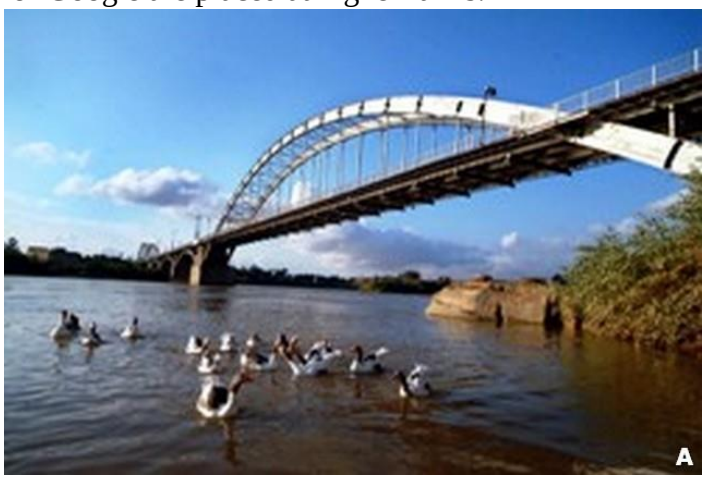

Figure 2. Example of a coded image: Sign A is placed as Alt text or image alternate text at bottom right side (http://iiproject.ir/ImagealtA.html) May 7, 2016

www.SID.ir

For investigation of research sample, hard indexing (Krause, 1988) or first level indexing introduced by Panofsky (1955) was used for artistic images, which later used under title of *Ofness* indexing by Layne (1986). In this type of indexing, indexer emphasizes objects and subjects visible in the image, and uses existing objects and events in the image for determining description level of images, unlike soft indexing or second and third level indexing by Panofsky, which Layne introduced it as *about* indexing, and includes subjective evaluation and interpretation of indexer individual. Krause (1988) distinguishes hard indexing (description of what visible by indexer in the image) and soft indexing.

A domain was registered in January 15, 2013 for conducting research, and research data were loaded in 72 MB volume. HTML 4, Notepad ++, and Cascading Style Sheets Css 2.3 were used for specifying HTML text format traits. iiproject.ir (image indexing project) website was prepared using allocated domain and host in January 14, 2013, and indexed images were loaded in January 20, 2013. Images were developed in the form of 9 links, and each links included 100 images with separate descriptive information. In order to optimize website search, proper website map with website subjects was prepared and website address was introduced to Google, Yahoo, and Bing webmasters.

During the time images were indexed by Google, we discovered that Google has indexed the website that we uploaded our images sooner than other search engines and browsers like yahoo and Bing, so we were interested to see whether passing the time has effected on the image retrieving rate or not. The results will be the same if we undertake a second similar research.

We paid attention to the concept based image indexing instead of content-based image indexing so the difference and variety of sample images weren't considered. In the former method, the emphasis is on human edited descriptions to images while in the latter automatic and computerized assignments and descriptions to the images are taken into attention when publishing images online (Chu, 2001). The content method allows researchers to locate images on digital libraries based on their physical characteristics while the next method locates them on the basis of their concepts. As a result, research on indexing images includes two methods; the first one is essentially rooted in Computer Science and the second one is in Information Science (Chu, 2001). Given the above issues, in this research, we have considered concept-based image indexing which is in the field of the Information Science research agenda. In other words, image retrieval would be improved when indexing images is to be done by human editors and indexers. By keeping such issues in mind, we preferred to use concept based image retrieval than the other method when managing the research project.

Actually, users searching for images from Google by keywords than other available mechanisms so the emphasis is more on keyword search and retrieval in comparison with content-based methods. A challenging issue for image databases and retrieval systems is how the same images by different text could be indexed and retrieved which itself is a potential area for further research. It is somehow a weakness for Google and other image databases to retrieve differently the same pictures with different descriptive texts.

After about two months of indexing images by Google, every image combined with its tags was searched from google and the results were recorded. Some well-known Persian thesauri like ASFA and NAMA were used to designate standard keywords like captions, file names, ALT texts to each image. The four collections of images were not retrieved by Google so we couldn't take them into consideration. It is also a concern why Google could not retrieve such collections with appropriate texts.

www.SID.ir

There are different guidelines for image publishing on the web all of which are not necessary rules but recommendations for better image publishing. These guidelines are considered as SEO (search engine optimizing) which are basically user's experiences not indexing and retrieval algorithms of search engines like Google.

Because database management systems like CMS (Content Management System) or DAM (Digital Asset Management System) are more content based database which uses computerized and pixel methods, we preferred to use another method by which we could test the effect of concept based and manmade indexing. Furthermore, many of users don't know how to browse in such collection. We thought that we have more control over the work in a website than a CMS or other templates.

Images were indexed using image caption in Persian, image alternate text, image title, file name, free language, controlled language, and image property tag. Persian cultural thesaurus (ASFA) was used in controlled language indexing, and NAMA thesaurus (Scientific-Technical Information Exchange System) was used in cases where terms were not found in ASFA thesaurus.

Since rank single variable data (image retrieval) in several independent groups are compared in this research, Chi-square test is used for investigating difference of frequency of retrieved images in the groups, and Kruskal-Wallis test is used for investigating mean rank difference in nine groups. It should be noted *Mann Whitney* U *test* is used for comparing mean rank of retrieved images as pair-wise.

Research Findings

In this section, differences in retrieved images are calculated and effectiveness of codes allocated to images in extent and rank of image retrieval in research sample is discussed.

Table 1
Descriptive indexes for image search using site operator

Statistics	Frequency	Mean	Min.	Max.
Numerical value	151	16.8	7	39

Table 1 provides descriptive indexes for investigating extent of retrieval of images from sample website in terms of site operator. In this table, the number of retrieved images is 151 images from a total of 900 images. It also indicates mean, minimum and maximum retrieved images in each of nine codes including image title (T), image alternate text (A), English image caption (E), Persian image caption (G), file name (N), free indexing language (F), controlled indexing language (C), image tag information (P), and images with no change (Q). Minimum and maximum frequency of retrieved images was 7 and 39, respectively.

Considering type and purpose of the research, one subject is raised which explores specific properties and conditions of the research sample. The subject includes 8 questions which attempt to answer the respective subject. Thus, considering various methods used for indexing images in the research sample, eight indexing methods are compared with Standard situation (images with no changes) in this research.

www.SID.ir

Archive of SID

Table 2
Observed frequency and expected frequency for nine groups

Image code	Observed frequency	Expected frequency	Remaining
image tag information (P)	10	16.8	-6.8
controlled indexing language (c)	11	16.8	-5.8
file name (N)	11	16.8	-5.8
free indexing language (F)	11	16.8	-5.8
image alternate text (A)	15	16.8	-1.8
English image caption (E)	13	16.8	-3.8
Persian image caption (G)	34	16.8	17.2
Image title (T)	7	16.8	-9.8
images with no change (Q)	39	16.8	22.2
Total	151		

Table 2 indicates observed frequency and expected frequency and remaining for nine groups. As observed in above table, minimum code (7 cases) is related to image title (T) code and maximum number of image retrieval (39 cases) was related to code Q (images with no change) and code G (Persian image caption). According to above table data, allocation of image alternate text and English and Persian image caption to the images in the website is significant in image retrieval from the website.

Table 3
Frequency difference of retrieved images

Statistics	Numerical value
Chi-square	48.755
Degree of freedom	8
Significance level	0.000

Table 3 gives chi-square, degree of freedom and significance level of chi-square for investigating frequency difference of the number of retrieved images in terms of different codes. Chi-square statistics (48.75) is significant at significance level ($P < 0.000$), and suggests that the number of retrieved images was different in different codes.

Table 4
Retrieved rank of images in terms of nine codes

Statistics	Frequency	Mean	SD	Min	Max
Numerical value	151	75.99	43.732	1	151

As observed in Table 4, mean rank in 151 retrieved images is 75.99 in all codes, and SD is 43.73 and minimum rank and maximum rank is 1 and 151, respectively.

Table 5
Mean rank of retrieved images in nine codes

Image code	Frequency	Mean rank
image tag information (P)	10	65.70
controlled indexing language (c)	11	67.55
file name (N)	11	47.45
free indexing language (F)	11	55.55

www.SID.ir

Archive of SID

Image code	Frequency	Mean rank
image alternate text (A)	15	62.83
English image caption (E)	13	58.27
Persian image caption (G)	34	103.68
Image title (T)	7	36.71
images with no change (Q)	39	88.74
Total	151	

Table 5 gives mean rank of retrieved images in terms of each code. The best ranks goes to images with code T (image title) as 36.71, and poorest rank goes for images with code G (Persian image caption) as 103.68. Although image title had lowest retrieval in terms of frequency, it is in better situation in terms of retrieval rank compared to other codes. Persian image caption in terms of number is in more suitable situation compared to other codes, but it at lowest rank of retrieval among other codes. Thus, allocation of image title and Persian image caption leads to increasing rank and number of image retrieval in site operator stage.

Table 6
Mean difference of image retrieval ranks

Statistics	Numerical value
Chi-square	34.131
Degree of freedom	8
Significance level	0.001

Table 6 indicates Kruskal-Wallis for mean difference of image retrieval ranks in nine groups. As observed, chi-square (34.12) with degree of freedom as 8 indicates significance level $p < 0.001$. Thus, mean image retrieval rank in nine groups is different.

Considering chi-square tests for measurement of frequency difference of retrieved images in nine groups and Kruskal-Wallis for investigating mean difference of retrieval ranks in nine groups were both significant, it is necessary to investigate groups in pairwise manner in terms of frequency and rank so that pairwise differences or pairwise comparisons are done. Hence, chi-square test was used for comparing frequency of retrieved images in groups in pairwise manner, and *Mann Whitney* U *test* was used for investigating image retrieval rank difference as pairwise manner.

Table 7
Chi-square and significance level for frequency difference of retrieved images in nine codes

Code	C	F	N	T	E	G	A	P
Q	17.16	15.16	15.68	22.26	10.66	0.342	13.0	17.16

In Table 7, Chi-square test for measurement of frequency difference of retrieved images is significant for code C and code Q at significance level ($P < 0.0001$). Codes F and Q were significant at significance level $P = 0.0001$. Codes N and Q were significant at significance level $P = 0.0001$. Codes T and Q were significant at significance level $P = 0.0001$. Codes A and Q were significant at significance level $P = 0.001$. Codes G and Q were not significant at significance level $P = 0.55$. Codes E and Q were significant at significance level $P = 0.0001$. Chi-square was significant for codes P and Q at level $P = 0.0001$.

Question A: Does use of controlled language make any difference in extent of image retrieval in research sample?

www.SID.ir

Archive of SID

Table 8
Image retrieval rank in terms of codes C and Q

Group	Frequency	Mean rank	Sum of ranks
C	11	20.73	228.00
Q	39	26.85	1047.00
Total	50		

Table 8 gives rank of image retrieval in research sample in terms of codes C and Q (controlled language images and images with no applied changes). Mean retrieval rank of code C is 20.73 and mean retrieval rank of code Q is 26.85. Thus, images of controlled language code have better retrieval rank compared to images of code Q.

Table 9
Mean difference of ranks in groups C and Q

Statistics	Numerical value
Mann Whitney U *test*	162.000
Z	-1.230
Sig. level	0.219

Table 9 indicates *Mann Whitney* U *test* for measurement of significance o mean difference in ranks of two groups as P = 0.219, thus image retrieval rank in terms of codes C and Q (controlled language images and images with no applied changes) have no significant difference. In other words, difference between two codes is not statistically significant and difference between two codes is randomly.

Question B: Does use of free language make any difference in extent of image retrieval in research sample?

Table 10
Image retrieval rank in terms of codes F and Q

Group	Frequency	Mean rank	Sum of ranks
F	11	20.64	227.00
Q	39	26.87	1048.00
Total	50		

Table 10 gives rank of image retrieval in research sample in terms of codes F and Q (free language images and images with no applied changes). Mean retrieval rank of code F is 20.64 and mean retrieval rank of code Q is 26.87. Thus, code F has better retrieval rank.

Table 11
Mann Whitney U *test* on retrieval rank difference in groups F and Q

Statistics	Numerical value
Mann Whitney U *test*	161.000
Z	-1.253
Sig. level	0.210

Table 11 indicates *Mann Whitney* U *test* on retrieval rank difference of codes F and Q (free language images and images with no applied changes) which shows statistics 161.0 and significance level as P = 0.21. Thus, retrieval rank of F and Q has no significant difference.

www.SID.ir

Archive of SID

Question C: Does use of file name make any difference in extent of image retrieval in research sample?

Table 12
Image retrieval rank in terms of codes N and Q

Group	Frequency	Mean rank	Sum of ranks
N	11	20.18	222.00
Q	39	27.00	1053.00
Total	50		

Table 12 gives retrieval rank in terms of codes N and Q (file name and images with no applied changes). Mean retrieval rank for code N is 20.18 and 27.0 for code Q.

Table 13
Mann Whitney U *test* on retrieval rank difference in groups F and Q

Statistics	Numerical value
Mann Whitney U *test*	156.000
Z	-1.370
Sig. level	0.171

In Table 13, *Mann Whitney* U *test* on retrieval rank difference in groups F and Q (fine name and images with no applied changes) is 156.0 and significance level is P = 0.17. Thus, there is no significant difference between N and Q in terms of retrieval rank.

Question D: Does use of image title make any difference in extent of image retrieval in research sample?

Table 14
Image retrieval rank in terms of codes T and Q

Group	Frequency	Mean rank	Sum of ranks
T	7	17.43	122.00
Q	39	24.59	959.00
Total	46		

Table 14 gives retrieval rank in terms of codes T and Q (image title and images with no applied changes). Mean retrieval rank for code T is 17.43 and 24.59 for code Q.

Table 15
Mann Whitney U *test* on retrieval rank difference in groups T and Q

Statistics	Numerical value
Mann Whitney U *test*	94.000
Z	-1.300
Sig. level	0.204

In Table 15, *Mann Whitney* U *test* test on retrieval rank difference in groups T and Q (image title and images with no applied changes) is 94.0 and significance level is P = 0.20. Thus, there is no significant difference between T and Q in terms of retrieval rank.

Question E: Does use of image alternate text make any difference in extent of image retrieval in research sample?

www.SID.ir

Archive of SID

Table 16
Image retrieval rank in terms of codes A and Q

Group	Frequency	Mean rank	Sum of ranks
A	15	22.67	340.00
Q	39	29.36	1145.00
Total	54		

Table 16 gives retrieval rank in terms of codes A and Q (image alternate text and images with no applied changes). Mean retrieval rank for code A is 22.67 and 29.36 for code Q.

Table 17
Mann Whitney U *test* on retrieval rank difference in groups A and Q

Statistics	Numerical value
Mann Whitney U *test*	220.000
Z	-1.400
Sig. level	0.161

In Table 17, *Mann Whitney* U *test* on retrieval rank difference in groups A and Q (image alternate text and images with no applied changes) is 220.0 and significance level is P = 0.16. Thus, there is no significant difference between A and Q in terms of retrieval rank.

Question F: Does use of image caption in Persian make any difference in extent of image retrieval in research sample?

Table 18
Image retrieval rank in terms of codes G and Q

Group	Frequency	Mean rank	Sum of ranks
G	34	33.82	1150.00
Q	39	39.77	1551.00
Total	73		

Table 18 gives retrieval rank in terms of codes G and Q (image caption in Persian and images with no applied changes). Mean retrieval rank for code G is 33.82 and 39.77 for code Q.

Table 19
Mann Whitney U *test* on retrieval rank difference in groups G and Q

Statistics	Numerical value
Mann Whitney U *test*	555.000
Z	-1.194
Sig. level	0.232

In Table 19, *Mann Whitney* U *test* on retrieval rank difference in groups G and Q (image caption in Persian and images with no applied changes) is 555.0 and significance level is P = 0.23. Thus, there is no significant difference between G and Q in terms of retrieval rank.

Question G: Does use of image caption in English make any difference in extent of image retrieval in research sample?

Table 20

www.SID.ir

Archive of SID

Image retrieval rank in terms of codes E and Q

Group	Frequency	Mean rank	Sum of ranks
E	13	21.62	281.00
Q	39	28.13	1097.00
Total	52		

Table 20 gives retrieval rank in terms of codes E and Q (image caption in English and images with no applied changes). Mean retrieval rank for code E is 21.62 and 28.13 for code Q.

Table 21
Mann Whitney U *test* on retrieval rank difference in groups E and Q

Statistics	Numerical value
Mann Whitney U *test*	190.000
Z	-1.342
Sig. level	0.180

In Table 21, *Mann Whitney* U test on retrieval rank difference in groups E and Q (image caption in English and images with no applied changes) is 190.0 and significance level is P = 0.18. Thus, there is no significant difference between E and Q in terms of retrieval rank.

Question H: Does use of property tag including subject and title make any difference in extent of image retrieval in research sample?

Table 22
Image retrieval rank in terms of codes P and Q

Group	Frequency	Mean rank	Sum of ranks
P	10	20.40	204.00
Q	39	26.18	1021.00
Total	49		

Table 22 gives retrieval rank in terms of codes P and Q (property tag and images with no applied changes). Mean retrieval rank for code P is 20.40 and 26.18 for code Q.

Table 23
Mann Whitney U *test* on retrieval rank difference in groups P and Q

Statistics	Numerical value
Mann Whitney U *test*	149.000
Z	-1.141
Sig. level	0.264

In Table 23, *Mann Whitney* U *test* on retrieval rank difference in groups P and Q (tag property in English and images with no applied changes) is 149.0 and significance level is P = 0.26. Thus, there is no significant difference between P and Q in terms of retrieval rank.

Discussion and Conclusion

In order to investigate effect of indexing methods on optimization of image retrieval using site operator, differences obtained in retrieval of images were calculated and effectiveness of properties allocated to images as well as retrieval extent and rank in research

www.SID.ir

sample was discussed. Although different in research methodologies, the findings of the current research in comparable with of some other researchers like Setchi, Tang, & Stankov (2011), Fadzli & Setchi (2012) or Smits, Plu & Bellec (2006) in which annotation properties of the images have influential effect on improving image retrieval. The findings also reveal the fact that intellectual assignment of annotation by human users could remarkably change the way a given image would be retrieved.

The number of retrieved images in site operator is 151 images from a total of 900 images. Minimum and maximum frequency of retrieved images was 7 and 39, respectively. Minimum number of image retrieval (7 cases) is related to image title (T) code and maximum number of image retrieval (34 cases) was related to code Q (images with no applied change) and code G (Persian image caption). Thus, allocation of image alternate text and English and Persian image caption to the images in the website is significant in image retrieval from the website.

Chi-square statistics (48.75) for investigating frequency difference of retrieved images based on different codes is significant at significance level ($P < 0.000$), and suggests that the number of retrieved images was different in different codes. Mean rank of 151 retrieved images in all codes is 75.99 and SD is 43.73. The best ranks goes to images with code T (image title) as 36.71, and poorest rank goes for images with code G (Persian image caption) as 103.68. Although image title had lowest retrieval in terms of frequency, it is in better situation in terms of retrieval rank compared to other codes. Persian image caption in terms of number is in more suitable situation compared to other codes, but it at lowest rank of retrieval among other codes. Thus, allocation of image title and Persian image caption leads to increasing rank and number of image retrieval in site operator stage. Kruskal-Wallis test was performed for mean difference of image retrieval ranks in nine groups. Chi-square (34.13) with degree of freedom as 8 is significant at significance level $p < 0.0001$. Thus, mean image retrieval rank in nine groups is different.

Chi-square test for measurement of frequency difference of retrieved images is significant for code C and code Q at significance level ($P < 0.0001$). Rank of retrieved images was not significantly different based on above codes. Codes F and Q were significant at significance level $P = 0.0001$. Rank of retrieved images was not significantly different based on above codes. Codes N and Q were significant at significance level $P = 0.0001$. Rank of retrieved images was not significantly different based on above codes. Codes T and Q were significant at significance level $P = 0.0001$. Rank of retrieved images was not significantly different based on above codes. Codes A and Q were significant at significance level $P = 0.001$. Rank of retrieved images was not significantly different based on above codes. Codes G and Q were not significant at significance level $P = 0.55$. Rank of retrieved images was not significantly different based on above codes. Codes E and Q were significant at significance level $P = 0.0001$. Rank of retrieved images was not significantly different based on above codes. Chi-square was significant for codes P and Q at level $P = 0.0001$. Rank of retrieved images was not significantly different based on above codes.

Overall it seems Google search engine is planned in a complex manner and it causes that images with some codes have better retrieval rank, and some images have better retrieval number. It seems that methods of image retrieval in Google for different parts of image (image title, image alternate text, image caption and etc.) is not set as fixed manner, and Google performs retrieval action differently for different image properties. Considering algorithms of indexing in search engines especially Google are among security issues of these companies, thus expressing comments on them requires further studies. Google search engine is capable in indexing and retrieval of images, but it is not adequately capable in retrieval of

www.SID.ir

images from a website. If image collections lack suitable descriptive annotates, they will not be retrieved. Hence, image database aware users about suitable descriptive annotates of images, and allows adding various tags to images by the users.

Since image analysis techniques demand large time overhead for the online retrieval process, it is advised some semantically-enabled content recognition technology to aid in semi-automating the annotation process of caption-poor images. Additional studies are also needed to investigate how image searchers diverge from text-based searchers.

References

Azzam, I.A.A., Leung, C.H.C. & Horwood, J.F. (2004). Implicit Concept-based Image Indexing and Retrieval, *10th International Multimedia Modelling Conference*, Brisbane, Australia, 5-7 January, p. 354.

Bar-Ilan, J., Zhitomirsky-Geffet, M., Miller, Y. & Shoham, S. (2012). Tag-based retrieval of images through different interfaces –a user study. *Online Information Review*, 36(5), 739-757.

Booth, P.F. (2001). *Indexing: The manual of good practice.* Munich: K. G. Saur.

Chu, H. (2001).Research in Image Indexing and Retrieval as Reflected in the Literature. *Journal of the American Society for Information Science and Technology*, 52(12), 1011-1018.

Chua, T.S., Pung, H.K., Lu, G.J. & Jong, H.S. (1994). A Concept-Based Image Retrieval System. *Proceedings of the 27th Annual Hawaii. International Conference on System Sciences,* Wailea, Hawaii, January 4-7.

Fadzli, S.A. & Setchi, R. (2012). Concept-based indexing of annotated images using semantic DNA. *Engineering Applications of Artificial Intelligence*, 25(8), 1644–1655.

Fauzi, F, & Belkhatir, M. (2013). Multifaceted conceptual image indexing on the World Wide Web. *Information Processing & Management*, 49(2), 420-440.

Fauzi, M.F.A, & Lewis, P.H. (2008). A multiscale approach to texture-based image retrieval. *Pattern Analysis and Applications*, 11(2), 141-157.

Ferecatu, M., Boujemaa, N. & Crucianu, M. (2008). Semantic interactive image retrieval combining visual and conceptual content description. *Multimedia Systems*, 13(5-6), 309-322.

Grauman, K. (2010). Efficiently Searching for Similar Images. *Communications of the ACM*, 53(6): 84-94.

Greisdorf, H, & O'Connor, B. (2002). Modelling what users see when they look at images: A cognitive viewpoint. *Journal of Documentation*, 58(1), 6-29.

Jang, T.S. (2002). *Image indexing and retrieval using formal concept analysis*. MA thesis, Windsor University. Canada.

Jayaratne, L. (2006). *Enhancing retrieval of images on the web through effective use of associated text and semantics from low-level image features*. PhD Dissertation, School of Computing and Mathematics, University of Western Sydney.

Krause, M.G. (1988). Intellectual problems of indexing picture collections. *Audiovisual Librarian*, 14(4), 73-81.

Layne, S. S. (1986). Analyzing the subject of a picture. A theoretical approach. *Cataloging and classification quarterly*, 6(3), 39-52.

Lee, H.J, & Neal, D. (2010). A new model for semantic photograph description combining basic levels and user-assigned descriptors. *Journal of Information Science*, 36(5), 547-565.

 www.SID.ir

Archive of SID

Ménard, E. (2009). Images: indexing for accessibility in a multi-lingual environment–challenges and perspectives. *The Indexer*, 27(2), 70-76.

Mills, R. (2011). Bluff your way in image management. *Multimedia Information & Technology*, 37(1), 24-25.

Panofsky, E. (1955). *Meaning in the Visual Arts*, Anchor, NewYork, NY.

Patil, R.C. & Durugkar, S. R (2015). Content Based Image Re-ranking using Indexing Methods**,** *International Journal of Emerging Technology and Advanced Engineering*, 5(8), 447-453.

Powell, R.R. (1997). *Basic research methods for librarians*. 3rd ed. Greenich, CT: Ablex Publishing.

Rorissa, A. (2008). User-generated descriptions of individual images versus labels of groups of images: A comparison using basic level theory. *Information Processing and Management*. 44 (5), 1741–1753.

Setchi, R. Tang, Q. & Stankov, I. (2011). Semantic-based information retrieval in support of concept design. *Advanced Engineering Informatics*, 25(2), 131-146.

Smits, G., Plu, M. & Bellec, P. (2006). Personal Semantic Indexation of Images Using Textual Annotations. *SAMT, LNCS* 4306, 71–85.

Vadivel, A., Sural, S. & Majumdar, A. K. (2009). Image retrieval from the web using multiple features. *Online Information Review*, 33(6), 1169-1188.

Vrochidis, S., Moumtzidou, A. & Kompatsiaris, I. (2012). Concept-based patent image retrieval. *World Patent Information*, 34(4), 292–303.

Westerveld, T.H.W. (2000). Image Retrieval: Content versus Context. In *Proceedings of the Conference on Content-Based Multimedia Information Access*, RIAO, Paris, April, 12-14, 276-284.

www.SID.ir