A Novel Approach for Image Semantic Search Engine

**Abstract**

We introduce ISSE, as a fully featured semantic search engine for images on the internet. This system gives the user the ability to search with both textual queries and image contents. It relies on ontology and the semantic meaning of the queries to present the user related searches and the results. Additionally, it uses the SIFT algorithm to analyze images for their key points and get visually similar images from database as well as images with contents related to the meaning of the content of the input image. The implementation of the proposed system is introduced and tested with 6 million input images.

# 1. Introduction

Most of the people today rely on search engines to find the information they need from the Internet as the biggest source of information to increase their productivity. The more information each day poured into the mix, The harder it gets to achieve this task. Users need a tool to extract the information they need from the internet with its help more effectively. The semantic search engines are more efficient because they allow the users to search for the information based on a set of rules and their meanings in contrast to only search by the textual information on a web document which is done in the standard search engines. With these kinds of tools it is possible to retrieve documents which do not include the keywords introduced by the user.

One of these kind is the semantic image search engine. Today there are many standard search engines which offer image search functionality. Google [[[1]](#endnote-1)], Bing [[[2]](#endnote-2)], Yahoo! [[[3]](#endnote-3)], etc, service to search for images but unfortunately most of them look for the images based on the HTML tags and keywords on web pages. A semantic image search engine is a tool to look for their meanings and relationships. In other words, these kinds of tools do not only look for information on HTML documents but they look for the semantic relationship between queries and the images even if it is not mentioned in the source website. There are two different approaches to express this semantic relationships:

**Search based on related keywords:** This search method uses information related to images such as keywords, metadata and other textual sources to find images related to the query.

**Content-based image retrieval (CBIR):** In this method, instead of using metadata and other information related to images, the image’s content is used to find other related images. This content includes shape, color, geo-location and many other information extractable from the images. There are innumerous tools and algorithms to extract this information; however, one of the bests in this area which is used in this paper is the SIFT algorithm [[[4]](#endnote-4)]. More information on this algorithm is presented in the section 3.

It is worth to mention that when either of the method above is used, a semantic search engine, should find the semantic relationship between the images and present the result. For instance, if a user search for “flower”, in the first method, the search engine finds the terms with a meaning related to “flower” – for example types of flower – and suggests them to the user. If the second method is chosen and the user submits a flower image to the system, will present the other dimensions of that image along with other images related to the subject of the input image (which in this case is “flower”).

Currently, many search engines offer image search functionality. In addition some of them (such as Google search engine) offer even the content-based image search. However, most of them do not look for relationship between the images and the queries and the semantic meaning of the queries. Instead, they only look for the textual terms in the HTML tags and web pages or the same images with different dimensions based on the search method used.

If a semantic image search engine is implemented correctly, in equal situations, the result would be more clear and satisfied and with less garbage compared to a standard image search engine. Because, based on analyzes made on the image contents, the results would be more related to the user’s need. The standard image search engines could present the user inaccurate results. For instance if an image has the title attribute of “a garden without flower” and the user enters the query “flower”, the standard search engine would return it as a result while the users obviously indents to find a flower image.

Additionally, the standard search engines have lack of the ability to find images with synonyms, subsets and terms including the query. Based on these flaws, the need to create a semantic image search engine becomes more obvious. One such search engine is introduced in this paper with specific features such as an ontology graph and RDF representation of information which will be covered more in detail in the section 3.

The rest of this paper is organized as follows. Information about semantic images search engines is presented in section 2. The details of implementing the ISSE system are described in section 3. Section 4compares related works with ISSE. Finally section 5contains conclusion and future work.

# 2. Semantic Image Search Engines

The semantic image search engine is a type of a semantic search engine which only searches among images; hence, in this section the main characteristics of a semantic search engine will be reviewed and when needed, the specific information related to images will be provided.

As mentioned before in section 1, there are two search methods feasible in the semantic image search engine: textual and content-based image search. In this section these methods will be discussed in more detail. After that, a representation of the semantic information called RDF will be presented. The discussions will include the brief introduction of the YAPTCHA tool created by us for the purpose of labeling images along with ontology and its usage in the proposed system.

## 2.1 Textual Image Search

Currently the most popular search method is processing a textual query submitted by the user. In a semantic search engine, a few more steps are required to process this query. First it’s meaning and relationships with the entities in the database should be retrieved. The various relationships include: synonyms, anonyms, subset, supersets, etc. After extracting the related terms, the results for these terms will be appended to the query result. The entire result will be ranked and presented to the user.

### 2.1.1 The Need to Labeling Images

After the list of related terms is extracted, the images related to them should be retrieved from database. Due to do this, there should be some way to relate the images to the terms. One of the most popular methods to do this is to label images. These labels usually will be saved in databases, XML files, as metadata or some other ways. but one thing common between them is using humans to do the labeling process since currently there is no efficient way to recognize objects in images.

### 2.1.2 Using YAPTCHA for Labeling Images

YAPTCHA is a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) which is used to label images. A CAPTCHA is a program which is designed based on Automated Turing Test[[[5]](#endnote-5)]. Alan Turing in 1950 [[[6]](#endnote-6)] introduced it to test a machine’s ability to exhibit intelligent behavior. A machine will pass the test successfully if a person outside of a room cannot distinguish it from a human in the room only based on some textual interactions and their replies.

YAPTCHA is a multilingual CAPTCHA, designed and implemented with the goal of resolving its counterparts’ security problems while presenting a pleasant user experience and added value of labeling unrecognized images.

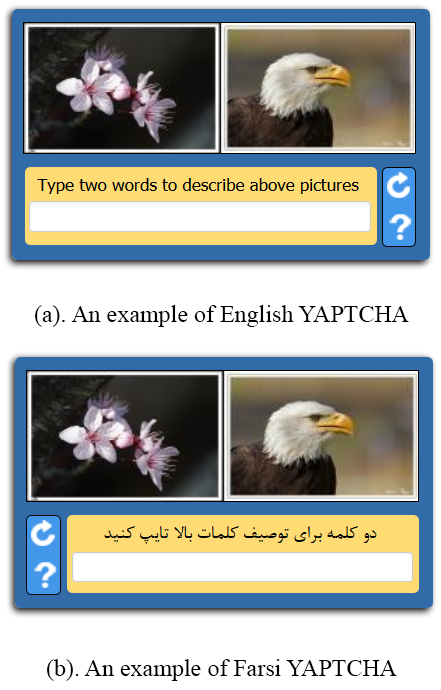


Fig.1Sample of YAPTCHA in 2 languages

As it is seen from the Fig. 1, in YAPTCHA, two images will be shown to the user. One of them is labeled and the other is known to the system. A user has to enter both words to bypass the system since the labeled image is not recognized from the non-labeled. If the user enters the correct word for the labeled image, this labeled is added to the system for further verification by the system. If enough users enter the same label for the image, it will be confirmed. It is worth mentioning that in a 6 hour workshop participating 60+ students from Shahed University of Tehran, this system had a 98.18% success rate in labeling images.

## 2.2Content-Based Image Search

Although many times it is easier to enter a textual query for initiating a search, in some cases it is needed to run the search through image’s contents, including:

* When there is not enough information about the source of image and we are looking for similar images
* When the database is very large and labeling the image is not preferred
* When all images are in the same class or are automatically generated and some attributes or portions of them is needed. For instance if someone is looking for a person in a survey camera.

In these cases, some algorithms are used to analyze the images and find similar attributes from them. One of such algorithms is the SIFT algorithm. This algorithm is presented in 1999 by David Lowe. Using this algorithm, it is possible to extract feature vectors of images and eventually find similar images even with different view angles, noise and rotation. To use this algorithm it is important to know the keypoint concept. In every raster image, there are a few points in one or more pixel widths which represent the key features of objects in the images. However these objects are in another image, even under some changes such as rotation, scaling, noise, etc. because these points are fixed, they’re detectable.

In the proposed semantic image search engine, after the image is submitted by the user, first the image is analyzed, then the feature vector is extracted and after the keypoints are analyzed, the similar images will be sent to the user.

### 2.2.1 Ontology and Its Usage in a Semantic Image Search Engine

One thing to note in YAPTCHA is that, in contrast to reCAPTCHA and many other CAPTCHA systems in which there is only one correct word for the distorted text, in YAPTCHA there are a few acceptable labels for each image. These labels should be analyzed using ontology and if they are in the ontology, they will be accepted. For instance, when an image of a passerine is sent to the user,when the user submits each of the words “passerine”, “bird” or “animal”, he/she is expected to successfully pass the CAPTCHA. In this regard, based on the results of our experiment in the workshop, the users tend to enter more general words instead of more specific ones. Meaning, a user is likely to enter the word “bird” instead of “passerine”.

To prevent the users from entering duplicate words for the images, the forbidden words concept is used in which, after some time of introducing an image to the database, the accepted words in the top of the ontology are added to the forbidden list and will be sent to the user to prevent him/her to submit them. Fig.2 and 3 show the usage of ontology and their relative images.

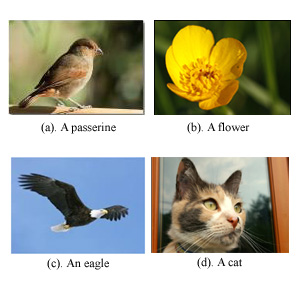


Fig.2 Four of YAPTCHA images

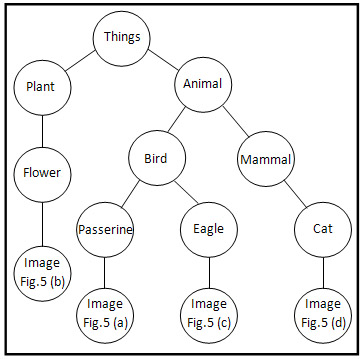


Fig.3 A sample ontology from Fig. 3

# 3. Proposed Method(ISSE)

Based on the system introduced in this paper, a semantic image search engine named ISSE (Image Semantic Search Engine) is implemented. This search engine is capable of running searches based on both textual and content-based image search methods. Fig 4 shows a summary the components of ISSE.

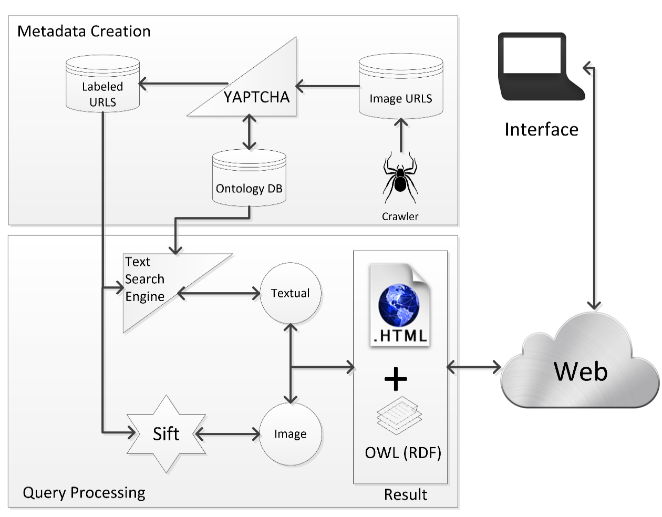


Fig.4 ISSE's components

First, in the metadata creation unit, the crawler program submits the non-labeled URLs to the database. As mentioned earlier, URLs are labeled with YAPTCHA which uses ontology database to verify submitted terms and then labels the URLs and saves them in the database.

In the query processing unit, the user submits the query using system’s interface which is implemented in open web technologies such as PHP, HTML5 and JavaScript. Depending on the search method the query follows two different paths in the system.

## 3.1Image Source and Ranking Strategy

The ISSE uses an image database which are used for [[[7]](#endnote-7)] and are downloadable for research purposes. In this database, the Bounding Box Annotation is used to focus on only one object. For each object a few attributes are annotated and the images are labeled manually. If a few persons have verified an image, it was accepted in the system.

The images used in database have initial ranking; however, based on the system’s requirements, the ranking of the images is update constantly.

**Ranking in Textual Search:** If the user searches a term in the system, after clicking on an image, its ranking will be increased slightly. In this way, the image will be possibly displayed at a higher position the next time.

**Ranking in Content-Based Search:** If the user searches for the content of an image, in addition to the previous ranking system, the number of keypoints and visual similarity attributes have a high ranking influence. This means, the results will first be sorted by keypoints with higher priority and then the images with exact or near exact rank will be sorted by previous method. This is due to the fact that when searching for images, the user is more likely looking for similar images than finding images that where popular among other users.

## 3.2Textual Query Processing

When the query is textual, first the textual search engine retrieves similar images for the query from the ontology database. Then it retrieves the image URLs for related terms from image database. After that, the semantic information for each image and the submitted query will be appended in the HTML and presented to the user in the OWL language in the RDF (Resource Description Framework) format.

The goal of the semantic web which is managed in W3C, is to present the content of the web in a format which is recognizable for the machine. Based on this fact, all the results of the ISSE are accessible in the OWL language in the RDF format.

In addition to sending the information in a machine-recognizable format, the information of ontology is presented to the user in an interactive graph above the image results to make it easy for the user to search for related words. This graph is generated using the HTML5’s canvas tag and JavaScript. It is designed and implemented with the goal of compatibility with all platforms. The results for the query “flower” is shown in Fig. 5.

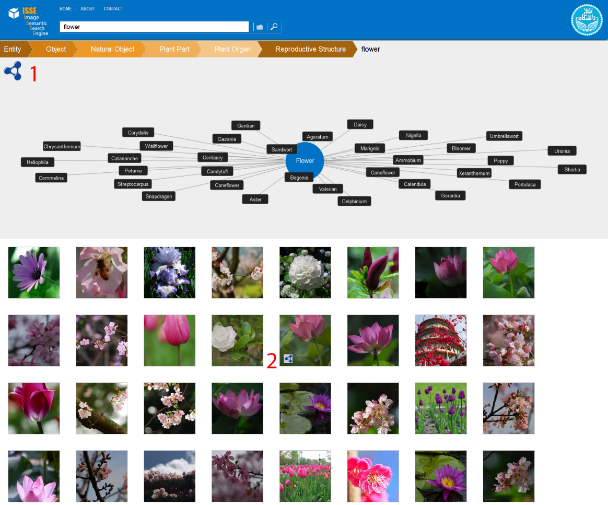


Fig.5Results for the word “flower” in ISSE

It’s good to be noticed that in addition to existing OWL information in RDF format, a link to these information is placed in the result page which is specified using red numbers in the Fig. 6. Number 1 shows the link to RDF for the query, while number 2 shows the link to RDF describing the image. Fig. 6 shows a portion of the RDF file for the “flower” query.



Fig.6 Excerpt of resulting RDF for "flower" query in ISSE

## 3.3Content-Based Query Processing

In the case that user prefers to search by image content, he/she can access the image upload form by clicking the camera icon (Fig 7).

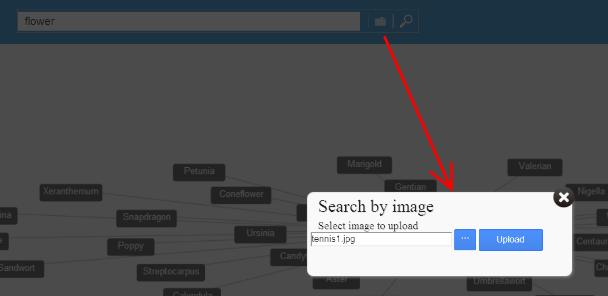


Fig.7 Image upload form in ISSE

After selecting the image and uploading it to the server, the image will be saved to the server and then sent to the SIFT-enabled program. This program will analyze the image for its keypoints and find the matching images from database and return the result. Then This result will be combined with the images with keywords similar to keywords of the images returned by the SIFT-enabled program which are visually similar images from database. Fig 8 the result for an uploaded image.

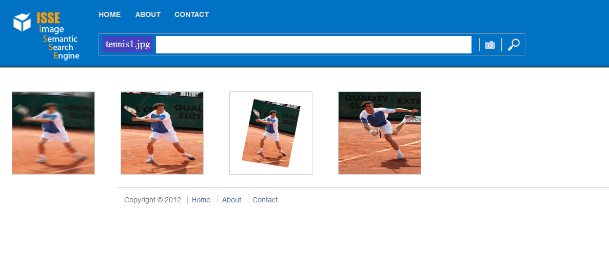


Fig.8 A sample result page for the content-based image search in ISSE

In the Fig. 8, since there were no image with the context of “tennis” in the database, only the visually similar images are shown. The images in the result contain blurred, noisy, rotated, etc. images of the source image.

# 4. Comparing ISSE with Related Works

Because the semantic web is a somewhat new concept, there are not many semantic image search engines available. In this section ISSE is compared with five of most popular and used of them.

## 4.1Pixolu

Pixolu [[[8]](#endnote-8)] is a semantic search engine which searches the images from Yahoo and Flicker. After the user submits the query, about 150 to 200 images will be presented to the user. Then the system asks the user to select his/her favorite images to retrieve similar images. Fig 9 shows the results for the “flower” query.

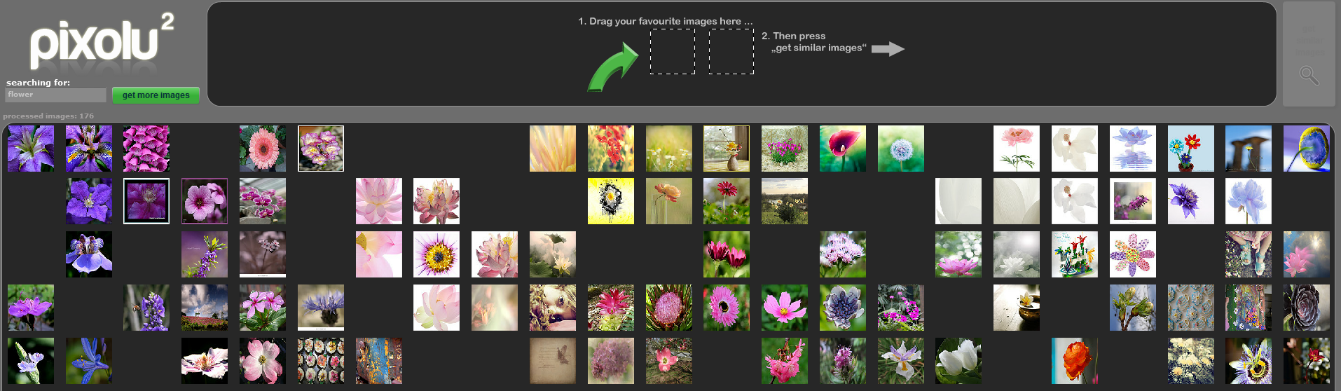


Fig.9 Results for "flower" in Pixolu

ISSE and Pixolu both search for images but in addition to different sources for their images, there are a few important differences between them which worth mentioning:

* Unlike ISSE, Pixolu doesn’t offer information from the ontology to the user. Finding similar queries for the first one submitted by the user and navigating through them to limit and refine the results is not possible in this system.
* The results’ information is not presented in RDF.
* The classification of the images is sufficed to user inputs. This may sometimes conflict with the favorite notion which is the term used to the user. An image may be pleasant (and hence favorite) by the user but not relevant to the input query.
* Arguably, the user interface, the service’s implementation and presentation of the results are more pleasant to the service instead of user. User cannot easily retrieve the images he/she likes. Instead he/she should spend too much time to classify the images. Meaning, it seems like the service’s goal is to classify images rather than finding the images for the user. And the user should make a lot of effort to find a simple image which isn’t pleasant. On the other hand, in ISSE, every feature is easily accessible to the user while the information required from the user is gained through YAPTCHA as another service. And it’s not integrated in the search system.
* In Pixolu, the classification is an optional operation. If the user finds the image he/she is looking for from the first search, he/she can not continue with the service. In YAPTCHA user has to enter labels for the images in order to pass the system.
* In Pixolu the user interface is implemented in Adobe Flash. While this platform is very popular, still there are a lot of users which don’t use it and don’t have it installed on their devices. The user interface in ISSE is implemented using the open web technologies such as HTML5, CSS, JavaScript and PHP. These are available in almost all modern devices.

## 4.2Nachofoto

Nachofoto [[[9]](#endnote-9)] is used to search for images currently trending in the internet. For example, during the Japan’s tsunami, searching the word “Japan” would result to images showing Japan’s tsunami (Fig. 10).

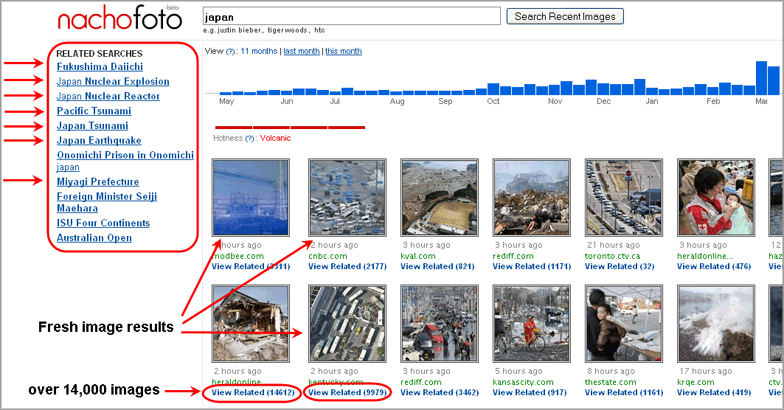


Fig.10 Seraching for Japan during tsunami

The strategy of Nachofoto is not very much related to the ISSE’s, because the focus of Nachofoto is on current events and their images, while ISSE tries to find related queries with related meanings and their images and present them to the user and machines (through RDF). Nachofoto is not looking for semantic relationshiop between the submitted query and other possible queries, it looks for relationship of the submitted queryand current point in time.

## 4.3Google

Today, the most used popular search engine is Google. This search service, like ISSE, offers both textual and content-based search methods, but the main difference is the way the query is processed in each of them.

### 4.3.1 Textual Queries

Regarding the fact that Google is a standard search engine and not a semantic one, the search is based on the keywords in the HTML of the source page. When Google crawls a web page, the information of that page such as the URL of the page it is found on, the alt tag of HTML tag and the path URL of the image will be stored. These are the information that will be used when the query is being processed, while ISSE uses the labels describing the image’s content to offer more accurate results.

In addition, although Google offers similar search queries to the user, ISSE offers more relevant queries and extracts these information from meanings of the submitted queries, while Google retrieves them from similar search queries submitted by other users.

### 4.3.2 Content-Based Queries

When searching for image contents, Google and ISSE process the queries similarly with the difference that ISSE in addition to offering the visually similar images, adds the images with similar subjects to the result set. For example, if an image of a flower is submitted, in addition to the images with visual similarity, images with “flower” keyword will be presented.

## 4.4TinEye

TinEye search engine [[[10]](#endnote-10)], works like Google’s content-based image search. When the user submits the image or its URL to the websites, the results will be presented to the user. Considering the similarity of this search engine and Google, the comparison mentioned in section 4.3 is also true for this search engine. The only difference of this with Google is its presentation of results, sorting the results and source databases.

## 4.5MUFIN

MUFIN (Multi-feature Indexing Network) [[[11]](#endnote-11)] is a project for presenting an optimized method for searching a huge dataset. In this project, we are about to demonstrate system’s functionality, a search engine is implemented which is capable of running both keyword-based and content-based searches like ISSE.

The images in the database of this search engine is provided by Profimedia [[[12]](#endnote-12)]. These images contain (as mentioned in the MUFIN’s website):

* A thumbnail image
* A link to the corresponding page on the Profimedia web-site
* Two types of image annotation: a title (typically 3 to 10 words) and keywords (about 20 keywords per image in average) mostly in English (about 95%)
* Five MPEG-7 visual descriptors extracted from the original image content: Scalable Color, Color Structure, Color Layout, Edge Histogram and Region Shape. 4.5.1 Textual Queries

The rest of this section describes the differences between ISSE and this search engine in each search method.

### 4.5.1 Textual Queries

The textual query in this search engine, like ISSE, is based on the keywords of each image. However, not only it doesn’t provide related keywords for each query, also this search engine is not a semantic search engine and suffices to only to a simple database query to find relative images and doesn’t use ontology to find similar images. Fig 11 shows the results for “flower” in this search engine.

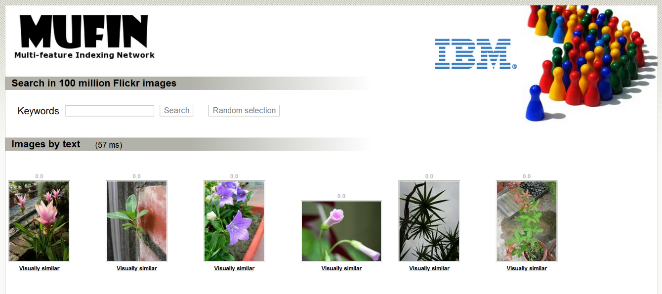


Fig.11 Results for "flower" in MUFIN

In addition, the database for keywords and images doesn’t grow overtime and it is only a static database containing a limited number of images. In ISSE, however, the design and implementation is in such a way that it will be growing overtime and can start with a small dataset.

### 4.5.2 Content-Based Queries

The MUFIN search engine does not offer the image upload functionality to the users and the content-based image search can only be initiated using the “visually similar” link at the bottom of each thumbnail on the result page. This means it can only search for initial images in its database and not an image provided by the user. Fig. 12 shows the results for the second result image presented in the Fig. 11.

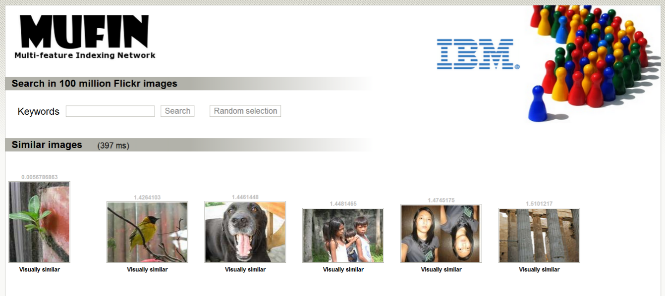


Fig.12 Content-based image search result for second image of Fig. 11

As it is comprehensive from Fig. 13, the MUFIN search engine does not usually offer an acceptable result set in content-based image searches. Additionally, the lack of semantically checking the contents of the images for relevant data is obvious. In Fig 13, the only flower (which is the context of the image the search is based on), is the first image which is the image the search is initiated from. This could be easily prevented since all the images in the database have keywords and the initial image has keywords itself. Comparing these results with the ones using algorithms such as SIFT shows this method’s weaknesses.

# 5. Conclusion

A semantic search engine is a type of search engine which in addition to standard methods to retrieving data, adds the value of meanings of queries and target items to consideration. One type of such search engine could be used to find images which is called a semantic image search engine. There are two methods, textual and content-based,for running searches in a semantic search engine

**Refer**ence

In the proposed search engine, ISSE, the textual search queries is validated through keywords of images and the ontology database which all are gathered using YAPTCHA service. The related queries is presented to the users to make it easier to refine his/her search. The content-based image search is ran through SIFT algorithm and after that the related results with similar contexts are added to the result set.

The details of design and implementation of ISSE and comparisons between it and other similar counterparts shows the accuracy and ease of use of this search engine.

**Acknowledgement**

The authors wish to thank the 60+university students in Bs., MSc. and PhD levels who helped in the testing process for the first YAPTCHA prototype.

1. [] Google Images, <http://www.google.com/imghp?hl=en&tab=wi>. [↑](#endnote-ref-1)
2. [] Bing Images- Search the web for pictures, photos & images, <http://www.bing.com/images>. [↑](#endnote-ref-2)
3. [] Yahoo! Image Search, <http://images.search.yahoo.com/>. [↑](#endnote-ref-3)
4. [] It’s 1999. Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157. DOI:10.1109/ICCV.1999.790410. [↑](#endnote-ref-4)
5. [] Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford. CAPTCHA: Using hard AI problems for security. In Eli Biham, editor, *Advances in Cryptology – EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques, Warsaw, Poland, May 4-8, 2003, Proceedings*, volume 2656 of *Lecture Notes in Computer Science*, pages 294–311. Springer, 2003. [↑](#endnote-ref-5)
6. [] [Elie Bursztein , Matthieu Martin , John Mitchell, Text-based CAPTCHA strengths and weaknesses, Proceedings of the 18th ACM conference on Computer and communications security, October 17-21, 2011, Chicago, Illinois, USA](http://dl.acm.org/citation.cfm?id=2046724&CFID=88791706&CFTOKEN=98198191) . [↑](#endnote-ref-6)
7. [] O. Russakovsky and L. Fei-Fei, Attribute Learning in Large-scale Datasets. Proceedings of the 12th European Conference of Computer Vision (ECCV), 1st International Workshop on Parts and Attributes. 2010. [↑](#endnote-ref-7)
8. [] Pixolu – find what you imagine, <http://www.pixolu.de/>. [↑](#endnote-ref-8)
9. [] Nachofoto, http://nachofoto.com. [↑](#endnote-ref-9)
10. [] TinEye Reverse Image Search, http://www.tineye.com/. [↑](#endnote-ref-10)
11. [] MUFIN : The MUFIN Project, <http://mufin.fi.muni.cz/tiki-index.php>. [↑](#endnote-ref-11)
12. [] fotobanka Profmedia – fotografie, ilustrace & video, <http://www.profimedia.cz/>/ [↑](#endnote-ref-12)