# Image Search Process in the Web using Image Copy

Paul Nieuwenhuysen
University Library, Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussel, Belgium
Paul.Nieuwenhuysen@vub.be

**ABSTRACT: Objective** - *The reported investigation aimed to assess the performance of services that offer search by image on the Internet / WWW to find copies of a known image. The motivation is that finding copies of an image can be useful for several purposes, while search by image is a relatively new method that allows finding such images and that is offered free of charge by a few Internet search services. Furthermore, the context of each revealed image copy may yield relevant information.*

**Methods** - *Various images have been used in empirical case studies.*

**Results** - *Finding copies is positioned in a structured view of applications of reverse image searching. The concept 'copy' is clarified and sharpened. Data have been collected regarding the performance of several contemporary services to find copies of an image, namely TinEye, Google and Yandex.*

**Conclusions** - *The reported findings have revealed significant differences in the number of copies discovered and in the precision of the search results. Therefore, all this may be helpful for users / practitioners to select and apply an appropriate image search service to reveal copies of an image and even information that is related to the image.*

## 1. Introduction

### 1.1 Finding Copies of an Image
Starting from a known image, it can be interesting to find duplicates (exact copies) of that image or to find near-duplicates (images that have elements in common with that image, but which are not identical).

How to detect as many as possible near-duplicates among billions of images in an efficient way is a challenging problem for scientists and engineers. Several approaches have been developed (briefly reviewed by Wang et al., 2010). An investigation showed that about 22 % of the images on the WWW had near-duplicates, and about 8 % had more than 10 (Wang et al., 2010).

One of the suitable search services on the Internet can be applied to find copies of an image.

The following gives a few scenarios.

### 1.2 Finding copies of an image that you have created
Starting from an image that you have created or of an image affiliated with your organization, you may want to find derived images.

• In general, you may want to track how such an image is used / reused. This can reveal that your image has been copied and reused without asking permission. This means copyright infringement, unless your image has been published with an explicit message that it is free from copyright for some or all applications. This is important, for instance for professional photographers and designers of infographics (see for instance Tyrrell, 2015 or 2016).

• Even a whole web site that you have created can be copied partially or completely and can be republished on another site. This can be checked of course by a classical text search, so that this may seem irrelevant to be mentioned in this context. However, such a text search will probably not detect a translated version of your document, while a search by image can reveal that one of your documents has been copied, translated and (re)published somewhere on the WWW. This is not purely hypothetical but realistic; I have experienced this personally and detected this by using a search by image.

• In a more positive and constructive way, finding copies or derived versions of your image allows you to assess the impact of that image on a worldwide audience, which reflects the 'value' of that image. The following are examples. Curators or owners of a collection of objects can assess the impact and reuse of photos of the physical objects in their collection, on a worldwide scale; more specifically this works well for photos of objects and landscapes (Nieuwenhuysen, 2013), for images of paintings from a museum collection (Kirton & Terras, 2013, 2014) and for digitized photographs from a collection in a university (Kelly, 2015). Scientists, academics and researchers (see for instance Kousha et al., 2010) as well as photographers, artists and designers can assess the impact and reuse on the WWW of images that they have created.

• Your image can form a component of a different image on the WWW, but it does not stop there, as digital images can also be used in the more tangible, physical reality, for instance in an illustration of a printed publication; earlier I have reported some concrete cases of digital images used on a book cover, on the cover of a music album, and on posters, all detected using search by image (Nieuwenhuysen, 2013).

### 1.3 Finding other versions of an interesting image
You can know an image that you consider as interesting, but

• That you did not create, so that you do not know the original version,

• That can be the original version or a derived version,

• That may give no clue about the creator/author/owner/publisher and the related copyright.

Then searching by image may yield desirable information:

• You may find other versions of that image, which are more suitable for your application and need; for instance, a version closer to the original image at a higher level of resolution or quality or integrity.

• You may find the copyright status of the image.

• You may find he author(s) or publisher or copyright holder, which can be useful to obtain more information or to discuss possible copyright linked to the image.

• You may find a copy of the image, plus also its location on some WWW page and WWW site, which can provide you with more information about the contents of the image plus related information.

• A search by text suffers from the fact that documents about the topic of your interest can be written in another language, so that they are not found by your simple search with words in only one or a few languages. On the other hand, a search by image does not suffer from this complication and difficulty. Therefore, a search by image with a source image that reflects your specific interest may be successful to reveal interesting, relevant documents on the web, independent of the text language.

• You may get a better view on the authenticity of an image that illustrates and supports the message / contents / claims of a document. The image has perhaps been copied from another site, from another context and perhaps it has even been modified / changed / doctored, to support the text, the claims of the author of the document. This phenomenon becomes more important, because news is distributed and consumed increasingly through more informal and less expensive channels such as social media on the WWW; however, this is accompanied by an increasing number of unverified / fake / false news stories and claims. Verifying images and contents or —in other words— detecting such untrustworthy information sources is considered now quite important, so that efforts are made to make the process of detection smooth, fast and user friendly (see for instance Elkasrawi et al., 2016; Goel, 2016).

• Related to authenticity, you may check the reality of a portrait of a person, for instance a portrait used as profile photo on a social networking site. In other words, search by image can be applied to authenticate people.

### 1.4 Images in Information Retrieval
Investigating information retrieval systems that involve images is motivated and justified by the fact that the number of images available is increasing rapidly, in parallel with the decreasing technical difficulties and costs that are related to

• Digitization of hard-copy images,

• Digital cameras and photography,

• Publication / distribution of images through the WWW and even social interaction associated with images.

Consequently, billions of images are now available on the web. This evolution makes images more important as carriers of information. So, images have also become more important in information retrieval and discovery. More concretely, digital information systems on the Internet and WWW can help in finding 'copies' of an image, as well as images that are needed, or in elucidating the origin or even the contents of a known image. A recent research paper deals with the precision of search for images on the Web with a classical text query and offers a review of searching for images with a text query (Uyar & Karapinar, 2016).

### 1.4 Search by image through the Internet
Finding duplicates or near-duplicates of an image is possible by applying a relatively new method of database searching, in which each query consists not of text but of an image file. The search results lead to images on the WWW and to related texts. Terms used for this method are

• Search(ing) by example

• Reverse image search(ing)

• Backwards image search(ing)

• Inside search(ing)

• Reverse image lookup = RIL

• Query by Image Content = QBIC

• Content-based information retrieval = CBIR

The company Google offers several search services and since 2011 also a search service of this type, named "reverse image search"; as Google products are popular, this is a term used in many texts about this search method. The author of this

contribution prefers the term "search by image" because it is more simple and clear.

The state of the art in visual information retrieval and in particular of content-based information retrieval (CBIR) has been sketched (Marques, 2016) and has been reviewed in more detail (Tyagi, 2017).

## 2. Objectives / Aims

This contribution deals with search by image to find copies of your source image. Earlier I have reported on an investigation of finding 'copies' of an image, by using

1. The pioneering and specific system and service TinEye to search by image at https://www.tineye.com, and

2. The more recent, similar service offered by the big, popular and successful company Google since 2011 at https://images.google.com/ (Nieuwenhuysen, 2013).

Afterwards, a few additional general search services that are significant in terms of their technical capacity and number of users have started to offer also some search service to search by image. Concretely, Yandex at https://yandex.com/images/ offers searching for images in the classical way with a text query and also searching by image. The user interfaces offered by Google and Yandex are similar. Yandex is mainly active and popular in Russia: "In Russia the most popular search engine is Yandex, it shares 60.4% of the market, while Google.ru has 26.2%." (Paananen, 2012).

After a first reported investigation (Nieuwenhuysen, 2013), the systems developed by TinEye and Google have evolved, and the WWW has grown significantly in recent years. TinEye stated on their user interface, that their search service deals with about 20 billion images. Numerous documents on the WWW and some in printed format only mention or deal with search by image and some applications, but only mention in a superficial way some of the services that are available on the WWW (see for instance Tyrrell, 2015 or 2016).

Furthermore, an application of search by image, can deliver various categories / types of results, as outlined below in the context of Figure 1, but most of these publications do not make a clear distinction between these categories and write only in a general way about "similar" images (for instance Adrakatti et al., 2016).

Users and potential users of search by image may and should be interested in the performance of the respective available search services, qualitatively and quantitatively, which can be expressed in terms related to the concepts that are classical in statistics in general (see for instance Stats-Lab Dublin, 2013) and more particularly also in the theory of information retrieval, namely recall and precision. However, I am not aware of a published report of a comparative assessment of contemporary search services that offer search by image. Therefore, I have investigated / assessed / compared the performance of search by image, as offered these days by TinEye, Google, and Yandex.

## 3. Methods

In each search query by image, one source image was submitted from the computer disk as query to the chosen search service.

The images used in the tests were photos in colors.

These photos have been present on the Internet on a public access web server already for several years, as part of a classical, simple web page and site and not as part of a container file such as a PDF or Word file or of a database, so that they can be accessed and harvested/copied by human users as well as by automated harvesting robots, without difficulty, in standard ways.

Each image file was submitted as a query, keeping the same name of the file as on the Internet. In earlier reported investigations of the capacity of Google to reveal images that are semantically similar, the meaningful, informative file names were changed to neutral, insignificant names, to avoid providing / revealing information about the contents / subject of the image to the search service (Nieuwenhuysen, 2014). This is not needed in this investigation. On the contrary, by providing the file name as present on the Internet, I hope / assume that this may help the search services to reveal the original image on the WWW; in the tests, I noted if the search service did include a link in the search results to the original web site.

Each test with a source image was executed on the same day with the various search services, to avoid measuring differences that are due to the evolution of the Internet and / or the search service. Only exceptionally a search has been repeated a few times to test the stability of the performance of the search service over time, as mentioned in the text below.

Each link that was given as a search result by a search service was manually activated and it was checked if the link leads indeed to retrieval of a document from the WWW. Most links do lead to a fetched document as expected; then this fetched document was inspected, to check if a 'copy' of the image was indeed included.

In this investigation, the relevance of a retrieved image / document / file is judged with the simplest bimodal model (relevant or irrelevant, yes or no, 0 or 1). More concretely, relevant means here that a 'copy' of the source image is indeed included, while the concept of 'copy' is clarified in a dedicated chapter of this paper. This simple judging is considered as sufficient here and has also been used in earlier similar investigations (see for instance Kousha et al., 2010; Nieuwenhuysen, 2013; Kelly, 2015; Adrakatti et al., 2016). This approach is more straightforward than in an investigation of the retrieval of semantically similar images, because then the results may contain images that are relevant in various aspects and at various levels, (see for instance Nieuwenhuysen, 2014, 2016a, 2106b).

## 4. Results and Discussion

### 4.1 Search by Image can Yield Various Types of Search Results
Investigating the applications of search by image has resulted in understanding that retrieved images (plus related information) should be considered as belonging to various types / classes / categories, as sketched in Figure 1.
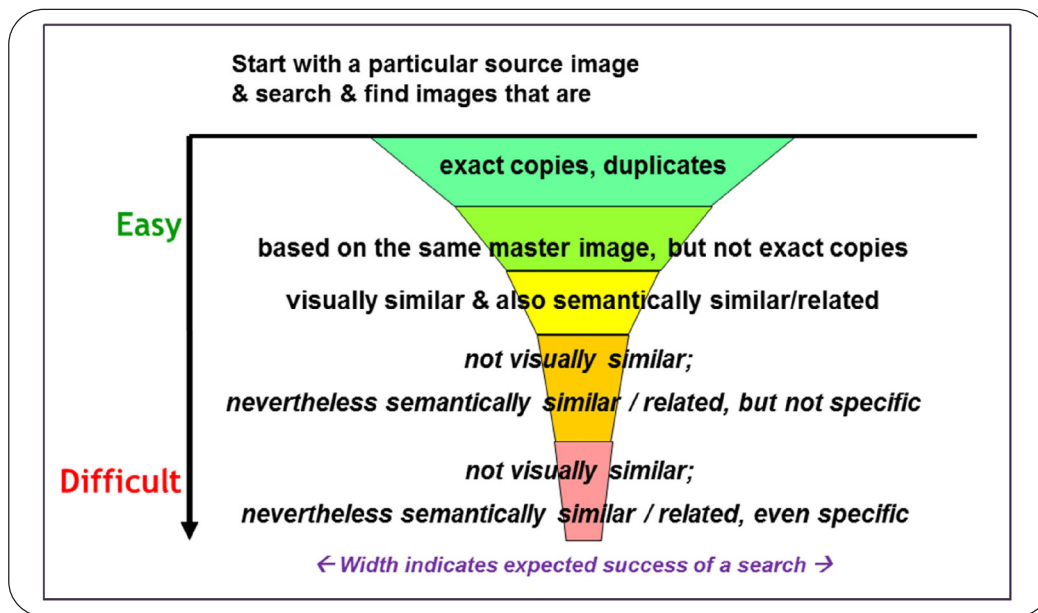


Figure 1. Search by image can yield various types of retrieved images

### 4.2 The Concept 'copy' of an Image
This contribution deals with finding 'copies' of an image by using reverse image search. This is a relatively easy task for a computer system, as shown in Fig. 1. In the text above, the word 'copy' is used loosely, in an intuitive way. In several more superficial documents about search by image, the words "copy" and "similar" are used without definition. One result of this investigation is that it has become clear that the concept of 'copy' is not straightforward and that a sharper, structured view has been formed, as shown in Figure. 2. In this context, we mean by 'copies' those images that are related to an original / authentic master image file and thus to each other. The source image used in a search query, as well as the relevant images that are found using search by image belong to the image files represented in Figure. 2, where each circle represents an image file that includes elements from the master image on top.
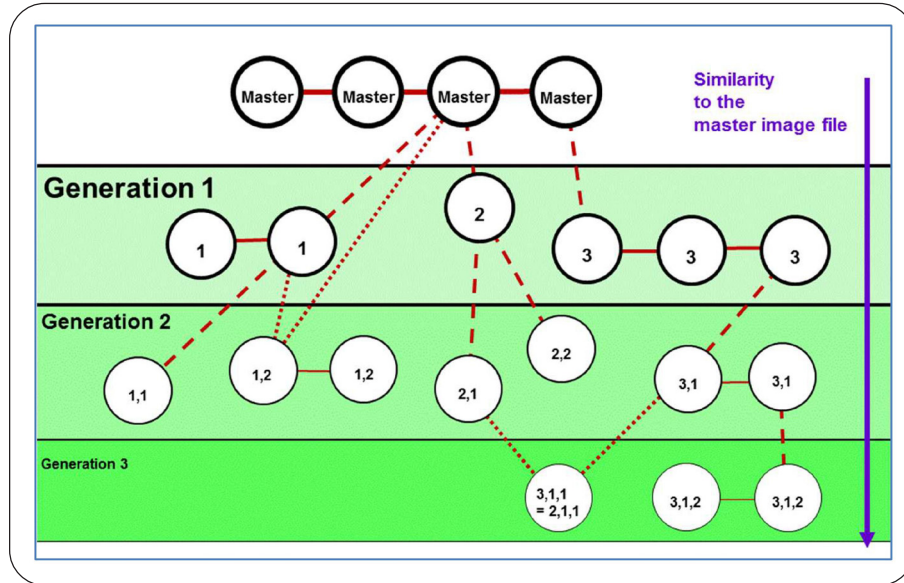
Figure 2.

Figure. 2. The image files and their relations that form the targets of search by image, as investigated in this paper, here shown in the form of an example that is limited to three generations. A full connecting line connects exact-duplicate image files. A dashed connecting line connects an image file to a parent image file of a previous generation from which it has been derived or from which it has at least inherited some elements. A connecting line that consists of only dots indicates the less common relation between an image file and more than one image file of a previous generation, namely to several parent image files from which it has inherited some elements. (This is not only possible in theory, but also in reality; indeed, I have found an image on the web, which includes two of the images that I had created and published on my website; this example is shown in the case studies reported below.).

In view of the considerations above, not the words copy / copies, but 'copy' / 'copies' is used in this paper, to indicate in a brief and simple way the image files that are related as clarified above.

**4.3 User Interfaces and Usability for Search by Image, offered by the Various Search Services**
*TinEye:*
This is an exceptional web search service, as it offers only search by image, while other search services offer several types of web search. Consequently, the user interface is simple and finding out how to start a search is easy and fast. TinEye offers the user various ways to rank the results: Best match / Most changed / Biggest image / Newest / Oldest.

This is a useful feature. For instance,

• To discover efficiently images that are not pure duplicates, you can select the option "Biggest change";

• To discover efficiently images that are closer to the original / authentic master image, you can select the option "Oldest".

This option is not offered by the other search services investigated.

*Google*
This company offers several ways to search the web. Consequently, it is less easy to find out where and how to start a search by image than with TinEye. Search by image is possible on the Google web page that offers also the more classical search for images with a text query. Here, the user interface is like the one offered by TinEye.

*Yandex*

This general search system offers search by image in their module / section for image search, like Google, but the search form is somewhat hidden and should be activated explicitly by the user.

**4.4 Finding the Original Image on the WWW**
The tests made in this investigation have shown that none of the search by image systems yield for each test case a link to the original source image on the WWW, even in searches that generate numerous links to 'copies'.

**This agrees with my earlier Findings:** Then TinEye gave almost no links to the original images the WWW, while Google gave a link to the original image in most but not in all tests (Nieuwenhuysen, 2013). Also, in an investigation of reuse of digitized photos, using TinEye and Google, the original image uploads were not returned as results, except in one instance (Kelly, 2015).

**4.5 Image 'copies' in Complicated Formats can be Detected**
The tests carried out in this investigation have demonstrated that search by image can even reveal / find / detect a 'copy' of a source image that is not simply present on the web as an individual image file in one of the classical formats (gif, jpg, png), but even when the image is embedded in a more complicated file format. More concrete examples:

• Google revealed a 'copy' as a part of a file in the format PDF. In an investigation of the reuse of digitized photos, Google also revealed a PDF file that includes reused images (Kelly, 2015). These observations agree with the message that at least since 2015 Google extracts images also from harvested PDF files, so that also these can be indexed and used in the search services offered by Google, which involve images (Chitu, 2015).

• Google revealed a web page that showed mainly a video including an image derived from the source image as part of a larger image.

• Yandex revealed a 'copy' in a series of slides.

So, it has become even more productive, efficient and attractive to apply searching for images.

**4.6 Number of Image 'copies' found by the Search Actions**
Several empirical case studies (tests) have yielded the results that are summarized in Table1.

| Name of test case | *TinEye* | *Google* | *Yandex* |
|---|---|---|---|
| Bwoom | **1** | **1** | 0 |
| Congo mask | 1 | **11** | 2 |
| Head 1 | 0 | **4** | 0 |
| Head 2 | 0 | 1 | **1** |
| Heads | 0 | **5** | 4 |

Table 1. Measured Number of Relevant Results

The best performances are indicated by bold numbers.

These data together with the earlier comparison of TinEye and Google (Nieuwenhuysen, 2013) can be summarized in the form of the following ranking of the performance of the search services concerning the number of image 'copies' revealed: 1. Google, 2. Yandex, 3. TinEye.

In a first attempt to use the image matching capability of TinEye as a potential tool for informetric analysis, researchers reported that "It seems likely that TinEye could only find a small fraction of the total number of copied images but it is not clear how small this fraction would typically be." (Kousha et al., 2010). All the test cases of this newer investigation show that nowadays still only a small fraction of the existing 'copies' are revealed by TinEye. In my earlier comparative assessment, TinEye gave also a smaller recall than Google (Nieuwenhuysen, 2013). In an investigation of the reuse of images of famous paintings, this was also observed: "Google Images returns a significantly larger number of results for each search than TinEye does." (Kirton & Terras, 2013, 2014). In an investigation of the reuse of digitized photos, "TinEye did not find a single reused image among the set", while

Google Image Search did reveal some reuse on the web (Kelly, 2015).

A recent report mentions only one test of the performance of Google, TinEye and Yandex, using only one source image (Adrakatti et al., 2016). That publication does not refer to any earlier, published test case, and does not make an explicit distinction between retrieval of 'copies' and semantically similar images. Anyway, in the test, Google delivered many more so-called "relevant results" than TinEye and Yandex. So that single test agrees also with the findings reported here.

### 4.4 Precision of Search Results

Another aspect of performance evaluation in information retrieval is the "precision". This is also used in visual search, as reviewed recently (Tyagi, 2017, section 4.3). In this type of application, the precision is less important than the number of 'copies' found, but the measured values are given anyway as follows.

The empirical case studies (tests) of this investigation have yielded the results that are summarized in Table 2.

| Name of test case | TinEye | Google | Yandex |
| --- | --- | --- | --- |
| Bwoom | 1/1 → 100% | 1/1 →100% | Non-Applicable |
| Congo mask | 1/1 →100% | 11/13 → 85% | 2/2 → 100% |
| Head 1 | Non-Applicable | 4/4 → 100% | Non-Applicable |
| Head 2 | Non-Applicable | 1/1 → 100% | 1/2 → 50% |
| Heads | 0/1→ 0% | 5/5 → 100% | 4/18 → 22% |

Table 2. Measured Precision of Search Results

If a search leads to zero relevant results, then the concept of "precision" makes no sense and is Non-Applicable, as mentioned in this table.

The best performances are indicated in bold numbers.

As a summary this gives the following ranking of the performance of the search services, in terms of precision: 1. Google, 2. Yandex, 3. TinEye.

This ranking is identical to the ranking above in terms of recall. This is not a priori expected; the tests had to be carried out to come to this conclusion. For instance, changing algorithms in the search system may increase recall (which is often desirable), but can also lead to a decrease in precision (which is not desirable); so, the algorithms are developed to result in a compromise that is acceptable / satisfying for the user. The greater recall of Google is not accompanied by a smaller precision; therefore, the greater recall is probably due to more extensive harvesting and coverage of documents on the web.

### 5. Conclusions

### 5.1 Finding 'copies' of a known Image, using Search by Image

To find 'copies' of a known image, the method of search by image is suitable. Furthermore, the leading services offer a friendly user interface, deliver results quite fast, and (at least in our case studies) the best performing search service gave results with a high precision; all this is convenient, satisfactory and time saving.

### 5.2 Choosing a Search Service to find 'copies' of an Image

To choose one of the services for your searches, you can consider criteria such as the usability, the number of 'copies' found and the precision of search results. This investigation leads to the following general ranking:

1. Google

2. Yandex

3. TinEye.

Google web services are not available in some countries; then the alternative web search services can be considered.

## References

[1] Adrakatti, A . F., Wodeyar, R. S., Mulla, K. R. (2016) Search by Image: A Novel Approach to Content Based Image Retrieval System, *International Journal of Library Science*, ISSN 0975-7546, 14 (3). Available free of charge from https://www.researchgate.net/publication/305683970_Search_by_Image_A_Novel_Approach_to_Content_Based_Image_Retrieval_System

[2] Chitu, Alex. (2015). Google Indexes Images from PDF Files. August 9, 2015. In the blog Google Operating System - Unofficial news and tips about Google http://googlesystem.blogspot.be/2015/08/google-indexes-images-from-pdf-files.html

[3] Elkasrawi, Sarah., Dengel, Andreas., Abdelsamad, Ahmed., Bukhari, Saqib, Syed. (2016). What You See is What You Get? Automatic Image Verification for Online News Content. 12th IAPR Workshop on Document Analysis Systems (DAS) 11-14 April 2016, IEEE.

[4] Goel, Sonal., Sachdeva, Niharika., Kumaraguru, Ponnurangam., Subramanyam, A.V., Gupta, Divam. (2016). PicHunt: Social Media Image Retrieval for Improved Law Enforcement. arXiv:1608.00905 [cs.MM]

[5] Kelly, Joan., Elizabeth. (2015) Reverse image lookup of a small academic library digital collection. Codex: The Journal of the Louisiana Chapter of the ACRL,  3 (2) 80-92. Available free of charge from http://journal.acrlla.org/index.php/codex/article/view/101

[6] Kirton, Isabella., Terras, Melissa. (2013). Where Do Images of Art Go Once They Go Online? A Reverse Image Lookup Study to Assess the Dissemination of Digitized Cultural Heritage. *In*: MW2013: Museums and the Web 2013, The annual conference of Museums and the Web | April 17-20, 2013 | Portland, OR, USA. Available free of charge from http://mw2013.museumsandtheweb.com/paper/where-do-images-of-art-go-once-they-goonline- a-reverse-image-lookup-study-to-assess-the-dissemination-of-digitized-culturalheritage/

[7] Kirton, Isabella., Terras, Melissa. (2014). Digitization and Dissemination: A Reverse Image Lookup Study to Assess the Reuse of Images of Paintings from the National Gallery's Website. *J. Digital Humanities*,  3 (1) Spring 2014. Available free of charge from http://journalofdigitalhumanities.org/3-1/reverse-image-lookup-paintings-digitisation-reuse/

[8] Kousha, Kayvan., Thelwall, Mike., Rezaie, Somayeh. (2010). Can the impact of scholarly images be assessed online? An exploratory study using image identification technology. JASIST, 61(9) September 2010, 1734–1744.

[9] Marques, Oge. (2016). Visual Information Retrieval: The State of the Art. *IT Professional,* 18 (4).

[10] Nieuwenhuysen, Paul. (2013). Search by Image through the WWW: an Additional Tool for Information Retrieval. *In:* Proceedings of the international conference on Asia-Pacific Library and Information Education and Practices = A-LIEP 2013 "Issues and challenges of the information professions in the digital age" Khon Kaen City, Isan, Thailand, 10-12 July 2013 [online] http://aliep2013.com/index.php/table-of-contents http://aliep2013.com/images/download/pdfs1/PaperNo41.pdf

[11] Nieuwenhuysen, Paul. (2014). Search by image through the Internet: applications and limitations. In Libraries in the Transition Era: New Space –New Services – New Experience. The Proceedings of the Seventh Shanghai International Library Forum, organized by the Shanghai Library, in Shanghai Library, China, 9-11 July 2014, http://www.libnet.sh.cn/silf2014/english/index.htm, Shanghai : Shanghai Scientific and Technological Literature Press, http://www.sstlp.com, 476 p.

[12] Nieuwenhuysen, Paul. (2015). Search by image through the Internet: an additional method to find information. In Transforming Libraries and Librarianship, Delhi : KBD Publication, 502 pp. Edited by Sanjay Kataria, John Paul Anbu, Shri Ram, Nirmal Kumar Swain, Naresh Singh Bhandari, 2015. 179-194.

[13] Nieuwenhuysen, Paul. (2016a) Information discovery on the Internet, using a search query that consists of text and an image. *In:* Libraries: Enabling Progress", Proceedings of the 8th Shanghai International Library Forum, organized by the Shanghai Library, in Shanghai Library = SILF 2016, China, 6-8 July 2016 <http: //www.libnet.sh.cn /silf2016/English /index.htm> Shanghai : Shanghai Scientific and Technological Literature Press, http://www.sstlp.com, ISBN 978-7-5439-7063-2, 442 pp., 153-162.

[14] Nieuwenhuysen, Paul. (2016b) Information discovery on the Internet, using a search query that consists of text & an image. *In:* A-LIEP 2016, proceedings of the 7th Asia-pacific Conference on Library and Information Education and Practice, in University of Nanjing, Nanjing = Nanking, Jiangsu province, China, Edited by Jianjun Sun, Qinghua Zhu, Christopher Khoo Shiyan Ou.. http://aliep2016.nju.edu.cn, PDF file, 99-112, Available free of charge from http://aliep2016.nju.edu.cn/files/A-LIEP%202016%20Proceedings.pdf

[15] Paananen, Anna. (2012). Comparative Analysis of Yandex and Google Search Engines. Helsinki Metropolia University of

Applied Sciences, Master's Degree Information Technology, Master's Thesis, 26 May 2012. Available free of charge from https://www.theseus.fi/handle/10024/46483

[16] Reilly, Michele., Thompson, Santi. (2016). Reverse Image Lookup: Assessing Digital Library Users and Reuses, *Journal of Web Librarianship*.

[17] Stats-Lab Dublin. (2013). Accuracy, Recall and Precision. Video available free of charge from https://www.youtube.com/watch?v=VPZiJGNX4_s

[18] Tyagi, Vipin. (2017). Content-Based Image Retrieval: Ideas, Influences, and Current Trends, Springer, Singapore, 378 p.

[19] Tyrrell, Katherine. (2015 or 2016). How to do a reverse image search. Available free of charge from http://www.artbusinessinfo.com/how-to-do-a-reverse-image-search.html

[20] Uyar, Ahmet., Karapinar, Rabia (2016) Investigating the precision of Web image search engines for popular and less popular entities. *Journal of Information Science*, April 27, 2016.

[21] Wang, Xin-Jing, Zhang, Lei, Liu, Ming, Yi Li, Ma, Wei-Ying. (2010). ARISTA Image Search to Annotation on Billions of Web Photos, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). http://ieeexplore.ieee.org/abstract/document/5540046/ DOI: 10.1109/CVPR.2010.5540046