

Information search & discovery, using an image as query: a workshop

Paul Nieuwenhuysen

University Library, Vrije Universiteit Brussel, Pleinlaan 2, Brussels, Belgium
Paul.Nieuwenhuysen@vub.be

Abstract:

Introduction:

This tutorial workshop is based on a continuing investigation of the power, applicability, usefulness and limitations of search by image through the Internet / WWW.

Background and purpose:

In this relatively new method for information retrieval, a query does not consist of text but of an image file. The search results lead to images on the WWW and to related documents. Google offers this WWWsearch method as “Reverse image search(ing)”

Other terms used for this method are

- Search(ing) by example
- Reverse image lookup = RIL
- Backwards image search(ing)
- Inside search(ing)
- Content-based information retrieval = CBIR

Furthermore, a search query can also consist of a combination of an image with text.

Topics:

1. Several online services are available free of charge to search by image.
2. Differences among these services are substantial..
3. Google can reveal images present on the Internet, which are duplicates / copies of the query / source image. The success is

quite variable from case to case.

4. This recall performance is strongly correlated with the performance of a more classical Google search by text to find copies of the query/source image file on the Internet.
5. Reverse image search as offered by Google can even reveal images that are modified versions of the query / source image; more specifically, modified versions can differ from the source image in size and in colours. The system can also reveal a fragment and even a modified / edited fragment of the source image, when this is included in an image present on the internet.
6. Three services to reveal copies of an image on the Internet / WWW have been compared. The outcomes have demonstrated that this type a searching can not only reveal images in simple classical formats such as jpeg / jpg and ping / png, but even to copies of the query image that are embedded in a more complicated file such as a PDF or set of slides. Furthermore, this investigation has led to a ranking for the number of copies found, and to the same ranking for the precision of the search results, as follows:
 - Google, the well know general web search system
 - Yandex, the general web search system that is popular mainly in Russia
 - TinEye, the pioneering system that is dedicated to search by image only
7. Our tests have demonstrated that since 2014 Google reverse image search can not only find images that are visually similar to the query / source image, but can even retrieve images that are semantically similar / related to the query / source image, while they have no elements in common the query image. The search results may also include a description of the subject of the image, and this can of course be interesting if the user has not yet much knowledge about the subject, so that using a specific text query becomes possible. Furthermore, other information related to the image and relevant links may also be included in the search results.
8. The performance of search by image to find images that are semantically similar to the query/source image is improving.
9. Not only pure, simple search either with words or with a source

image is possible by the freely available search system offered by Google, but a search query can also consist of a combination of an image with words. This allows us to combine the strengths of more classical text retrieval with the more recent search by image. Our tests have shown that this allows us to obtain search results with a precision that is higher than when only one of both search methods is used.

10. The progress described above in automatic analysis of images to determine some of their contents / meaning / semantics is also reflected by the improvements in automatic categorization / classification of images based on their contents. A popular example of such a system that is freely accessible and usable is Google Photos at <https://photos.google.com/>
11. Recently, search by image is applied not only by systems that are primarily search services, but also by systems in which images are important. An example is Pinterest that allow users to collect or “pin” images in virtual, digital, so-called pin boards. Our recent comparison of Google with Pinterest shows that Pinterest has become an additional and even competing system that allows us to start from a selected source image to find related, relevant images; furthermore, any of these found relevant images may lead to information that is related to the selected source image.

Various applications can be shown:

- Starting from an image that you created or that is affiliated with your organization, you may find copies / duplicates or even modified versions on the WWW. This can reveal copyright infringements. In a more positive way, this allows to assess the impact of such images on a worldwide audience. For example: 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.
- Starting from some interesting image that you have not created, but that you consider as interesting, and that is perhaps not the original version and for which the creator/author is not indicated, you may find other and better versions that are more suitable for your application and need. Also you may find the author(s) on the WWW, which can be useful to obtain more information or to discuss possible copyright linked to the image.
- Also searching by image may allow us to discover that the image that illustrates and supports a document is NOT real /

authentic, but that is has been copied from another site, from another context and perhaps that it has even been modified / changed / doctored, to support the text, the claims of the author of the document.

- Starting from some interesting source image, you may find semantically related images; in other words, you may discover images with a subject that is related to the subject of that source image. Furthermore, including some text in the query may increase the precision of the results, even when not enough knowledge is available in advance, so that only one or a few unspecific search words can be used.
- Consider the scenario in which you have already sufficient information/knowledge in advance to formulate and submit a classical, textual, specific, focused search query; even then, including an image to the text query can be useful to increase the precision of the results.
- Furthermore, in each of these applications, you may also find related text information.

Recommendations:

The growing success of the search methods that include an image in the query to find relevant information leads us to a few recommendations:

1. To find relevant information, these recent, additional search methods should be considered besides more classical methods, by librarians and information intermediaries in general, and also by end-users of information discovery systems.
2. As a consequence, search by image deserves a place in educational courses and tutorials on information and media literacy.
3. Authors and publishers in general want to create their publications and make these available in such a way that they rank high in the results of relevant search and discovery systems. Therefore it is good practice to take into account the workings of at least the classical, popular, search services, in the creation and optimization of their website(s). Not only the texts in a website should be considered, but also images, to optimize
 - for a relatively classical search with a text query to find images,

- for a more recent search by image, or
 - for a search with a query that consists of text plus an image file.
4. More concretely, website developers should try to publish their meaningful images in such a way that these can be well harvested, analysed and included in the database index of relevant search systems in an effective way.

Conclusions:

Information discovery is enhanced by recent methods that involve images:

- Search by image is evolving to a powerful, additional method to tackle information needs that are difficult to handle with more classical methods.
- Information discovery is assisted by automatic classification of images and by recommendation services based on image similarities.
- Furthermore, using a combination of text with an image in a search query can increase the precision of the search results, in comparison with a more classical pure text search or with a pure search by image.

Goals and learning objectives:

Participants learn about state-of-the-art applications and limitations of reverse image search on the internet and WWW. So they will be motivated and enabled to apply this relatively new method to discover information and to support other potential users.

Format and method:

This workshop can be organized and offered as a lecture illustrated with practical examples. Interactions with the participants can make the lecture more dynamic. The workshop can include some time for practical tests and applications, if the participant has access to a computer and internet.

The target audience: No restrictions. For everyone interested in the topic.

Time: 3 hours, including applications by the participants.

Keywords: search by image; reverse image searching; Internet; WWW; search systems; search services; TinEye; Google; Yandex; precision;

semantic gap

References

Marques, Oge (2016) Visual Information Retrieval: The State of the Art.

IT Professional, Volume: 18, Issue: 4, DOI: 10.1109/MITP.2016.70 (accessed in 2016).

Nieuwenhuysen, Paul (2013) Search by Image through the WWW: an Additional Tool for Information Retrieval.

Full text published 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” held at Pullman Khon Kaen Raja Orchid Hotel, in 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> and full text available also online free of charge from https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions

Abstract published in printed Abstract Book, 55 pp., on p. 38.

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 pp, ISBN 978-7-5439-6289-7. pp. 145-155.

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, ISBN: 978-81-907999-6-6, pp. 179-194.

Nieuwenhuysen, Paul (2016) 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 Shiyuan Ou.. <http://aliep2016.nju.edu.cn>, PDF file, pp. 99-112. Available free of charge from <http://aliep2016.nju.edu.cn/files/A-LIEP%202016%20Proceedings.pdf>

Nieuwenhuysen, P. (2017) Information discovery: A picture is worth a thousand words?

In the proceedings of the International Conference on the Future of Libraries (ICFL), From Promises to Practice, Editor-in-chief M. Krishna Murthy, Editors B. Ramesha, B. Subhash Reddy, organized by DRTC = Documentation Research and Training Centre, Indian Statistical Institute, Bangalore @ ISI = Indian Statistical Institute in Bangalore = Bengaluru, Karnataka, India, 15-17 November 2017, Published by DRTC, Indian Statistical Institute, 2017, 602 pp., ISBN 978-81-92179-79-7, pp. 313-324.

https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions

Nieuwenhuysen, P. (2018a) Information discovery based on the emerging technology to analyse digital images. In Reshaping libraries: emerging global technologies and trends. ICRL-2018. Papers of the First International Conference on Reshaping Libraries: Emerging Global Technologies and Trends organized jointly by DELNET-Developing Library Network, Ambedkar University Delhi in Association with the Society for Library Professionals (SLP) and Special Libraries Association (SLA, USA) Asian Chapter, February 1-3 at Jaipur, India. Chief Editor: H.K. Kaul, Editors P.K. Jain, Debal C. Kar, Sangeeta Kaul. Published by DELNET, New Delhi. ISBN 978-93-82735-13-7. 230 pp., pp. 1-14. & available from https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions & to be republished in the printed DELNET Newsletter in 2019.

Nieuwenhuysen, P. (2018b) Information Discovery and Images A Case Study of Google Photos.in: 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), in Noida, India, 21-23 Feb. 2018, INSPEC Accession Number: 18150464, DOI: 10.1109/ETTLIS.2018.8485238, Publisher: IEEE <https://ieeexplore.ieee.org/document/8485238>

Nieuwenhuysen, P. (2018c) Information discovery using search by image: a comparison of Google and Pinterest. In The Proceedings of the 9th Shanghai International Library Forum = SILF 2018, Library for all: towards a smarter and inclusive society, October 17-19, 2018, at the Shanghai Library, in Shanghai, China, 464 p., pp. 400-412, ISBN 978-7-5439-7748-8, <http://www.sstlp.com> & available from https://www.researchgate.net/profile/Paul_Nieuwenhuysen/contributions

Nieuwenhuysen, P. (2019) Finding copies of an image: a comparison of reverse image search systems on the WWW. Accepted for publication in the proceedings of the 14th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 19th COLLNET Meeting, at the University of Macau, Macau, 5 to 8 December 2018, <http://socio.org.uk/collnet2018/>

Thompson, S. and Reilly, M. (2017) "A picture is worth a thousand words": Reverse image lookup and digital library assessment. *Journal of the Association for Information Science and Technology*, 68, pp. 2264–2266. doi:10.1002/asi.23847