Information retrieval using search by image

Preliminary version of slides created to support a presentation at the annual *Inforum* day
Inforum 2015 : The I&D puzzle: new pieces to solve it
organised by *ABD / BVD*, in KBR, Brussels, Belgium KBR
June 5, 2015
http://www.abd-bvd.be/inforum/inforum_2015_fr.html

by Paul.Nieuwenhuysen@vub.ac.be
Vrije Universiteit Brussel
Brussels, Belgium

Vrije Universiteit Brussel

These slides should be available from the WWW site
http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations/
(note: BIBLIO *and not* biblio)

Vrije Universiteit Brussel

contents
= summary
= structure
= overview

- Introduction / context
1. Searching for images, using text as query
2. **Searching for images, using an image as query**

Introduction / Context

Introduction: images

Images become more important as information sources, due to
--increasing number of digital cameras
--increasing number of digital photos on the WWW

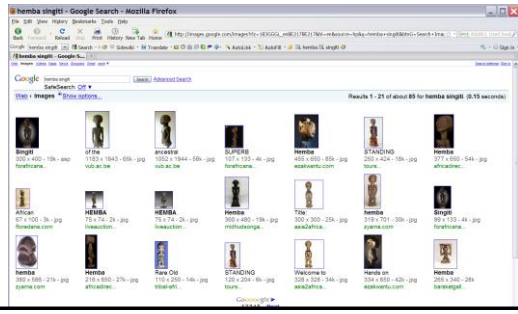
Introduction: images

→ Searching for images becomes more important

Searching for images, using *text* as query

Introduction: finding images and information

Introduction: finding images and information



Findings & Discussion

YAHOO! bing Google images

Comparison of WWW image search engines:

- Good relevance ranking
- Overlap in coverage and search results, as expected
- *Yahoo!* & *Bing* quite similar, as expected

Introduction: finding images and information

Searching for images on the WWW has become an attractive starting point to discover information:

- A search query with text is submitted to an "image search engine".
- The results come as small "thumbnail images".
- Each small image links to the document where it occurs.

Searching for images: images & their context

Image found on WWW (+ link to WWW file)

13

Searching for images: images & their context

Image & context on WWW

14

Introduction: finding images and information

- Search for images is faster than search for texts.
- When a relevant image is found, then also the context of that image is found, and that can yield relevant information.

15

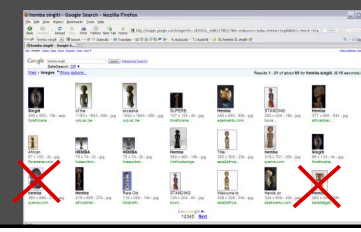
A report of a part of research work has been published:

P. Nieuwenhuysen
WWW image searching delivers high precision and no misinformation: reality or ideal?
Issues in Informing Science and Information Technology,
Vol. 7, 2010, pp.109-131.
ISSN: On-line 1547-5867; CD 1547-5859; Print 1547-5840,
peer reviewed journal, available online free of charge from
http://iisit.org/Vol7/IISITv7p109-131Nieuwenhuysen770.pdf

16

Conclusions ☹

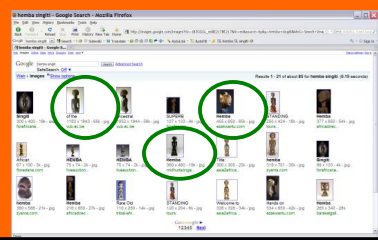
- WWW image search delivers
- a mixture of useful information with irrelevant information and even with misinformation



17

Conclusions ☺

+ Information retrieval from the WWW through image searching is attractive, simple and fast, AND can be efficient / productive.



18

Searching for images, using *an image* as query

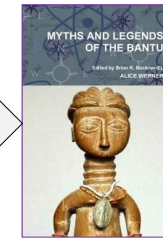
Searching by image: introduction

Besides searching for images with a query that consists of text,
some systems allow us to use as query
an image (file).

- *Searching by example*
- *Reverse image search*
- ***Searching by image !***

Searching by image: example

Search with the original image
reveals
a derived image



Original photo of a sculpture → Book cover

Searching by image: publication

A report on this subject has been published:

P. Nieuwenhuysen

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
also available free of charge from
http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations

Searching by image: various types of results

Start from a particular image;
search for images that are
not copies at all,
but semantically related,
from weak to strong

Difficult

Easy

NOT exact copies

exact copies
= duplicates

← Width indicates expected success of a search →

Research problem

Which systems are available
free of charge
for searching by image
through the WWW?

Searching by image: search engines

Available WWW image search engines:

TinEye Reverse Image Search

Google images

others…

since 2011

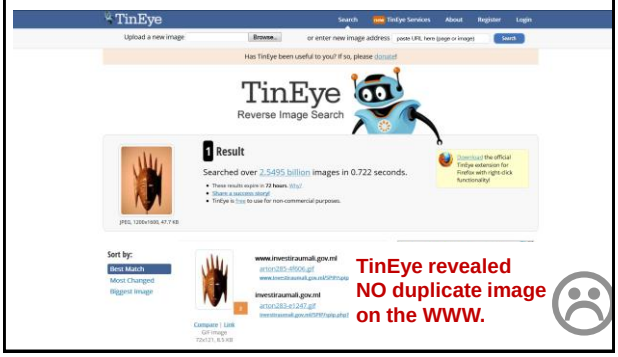
Research problem

Which differences among these systems are interesting for a user in practice?

Searching by image: Research method

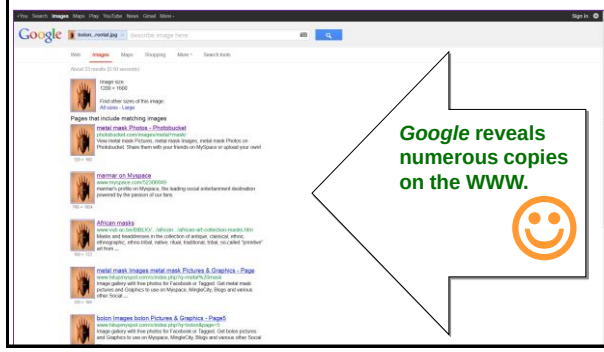
In each test query, we submit an image from the disk of a PC.
These images are accessible also on the WWW,
already for many months.
We check if the search engine finds the image copy that is accessible on the WWW.

Searching by image: Findings & Discussion: *TinEye*



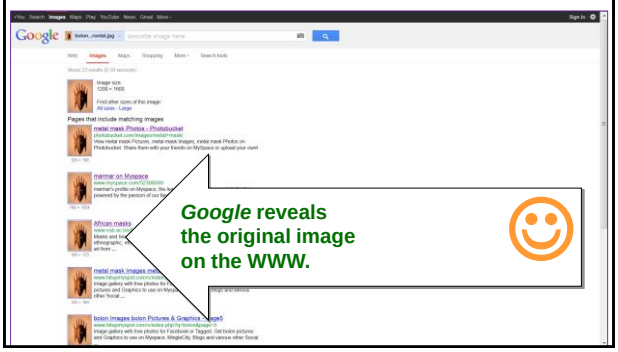
TinEye revealed NO duplicate image on the WWW.

Searching by image: Findings & Discussion: *Google*



Google **reveals numerous copies on the WWW.**

Searching by image: Findings & Discussion: *Google*



Google **reveals the original image on the WWW.**

Searching by image: Findings & Discussion

10 images that have a duplicate / exact copy present on the WWW, were submitted as a query, to both systems:

TinEye Reverse Image Search

Google images

TinEye **revealed only 3/10.**

Google **revealed 7/10.**

31

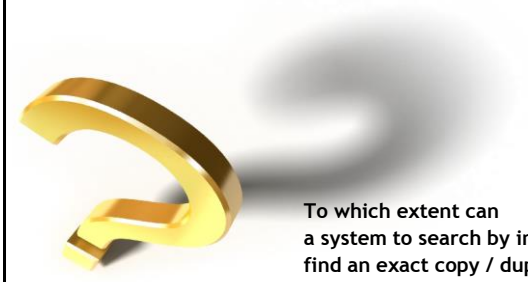
Searching by image: Conclusion

TinEye Reverse Image Search

Google images

32

Research problem



To which extent can
a system to search by image
find an exact copy / duplicate
that is present on the WWW ?

33

Searching by image: searching for exact copies

Start from a particular image;
search for images that are
not copies at all,
but semantically related,
from weak to strong

Difficult

NOT exact copies

Easy

exact copies
= duplicates

← Width indicates expected success of a search →

34

Searching by image: Findings

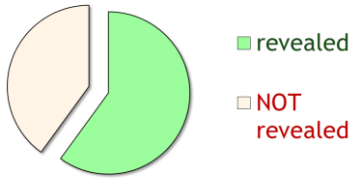
25 images that have a duplicate / exact copy
present on the WWW, were submitted as a query to
images.google

Google images

35

Searching by image: Findings

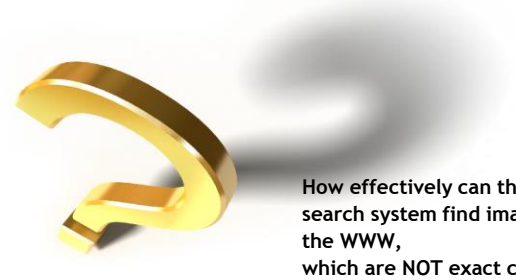
search by image for copies
(that exist on the WWW)



Successful
but not completely

36

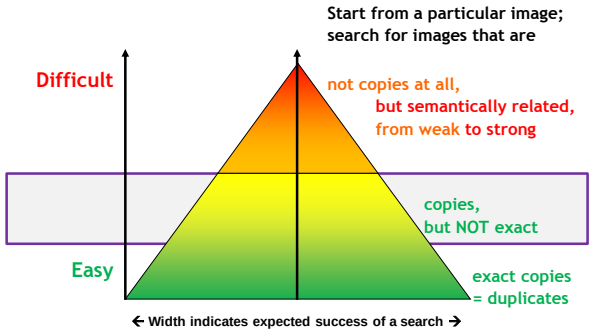
Research problem



How effectively can the
search system find images on
the WWW,
which are NOT exact copies
of a particular image,
but which do have elements
in common?

37

Searching by image: searching for edited copies



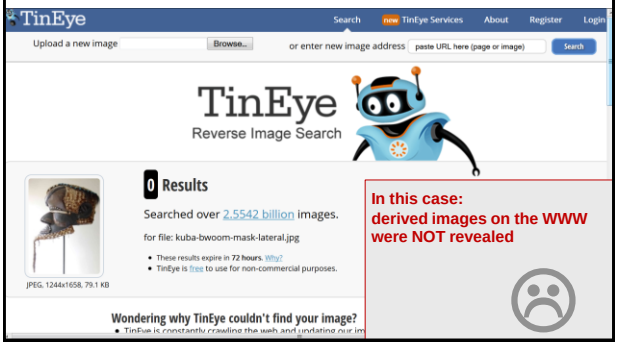
38

Research method

In each test query, we submit an image from the disk of a PC.
These images are accessible also on the WWW,
already for many months.
We check if the search engine finds the image copy that is accessible on the WWW.
We also check if the search engine finds *derived* = *modified* copies on the WWW.

39

Searching by image: Findings & Discussion: *TinEye*



In this case:
derived images on the WWW were NOT revealed

☹

40

Searching by image: Findings & Discussion

10 images were submitted as a query to 2 search systems:

TinEye Reverse Image Search

Google images

TinEye revealed
NO images on the WWW,
which include common elements, in any query.

Google revealed images on the WWW,
which include common elements
in 7 of those 10 queries.

41

Searching by image: Findings & Discussion

Example:

Google images

Original revealed the derived image:
→

42

Searching by image: Findings & Discussion

Example:

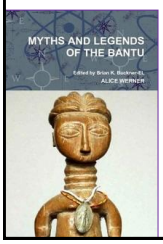
Google images

Original revealed the derived image:
→

Searching by image: Findings & Discussion

Example:

Google images



Afterwards,
the derived image
revealed the original:
→

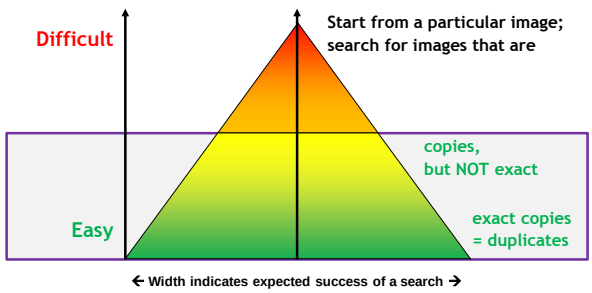
Searching by image: Findings & Discussion

Example:

Google images

Original image
in colours
revealed the derived image
in black & white, on a poster
→

Searching by image for copies: *applications*



Applications of searching by image: *Finding copies of your image*

- We can start from a particular image that we have created.
  Then we can search by image,
  to investigate if this image is used and made available from another WWW site.

Applications of searching by image: *Finding copies of your image*

- This can be interesting in several ways:
  - » Copyright infringements / plagiarism can be discovered.
  - » In a more positive/constructive way:
    allows us to investigate the impact of some image !
    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.

Applications of searching by image: *Finding other versions of an interesting image*

- We can start from an image
  that we consider as interesting,
  but that we did not create
  and that is perhaps not the original version
  and for which the creator/author is not indicated.

→

49

Applications of searching by image: *Finding other versions of an interesting image*

Then searching by that image may allow us to find

- a more suitable version of that image
- the creator/author
- another version of the image

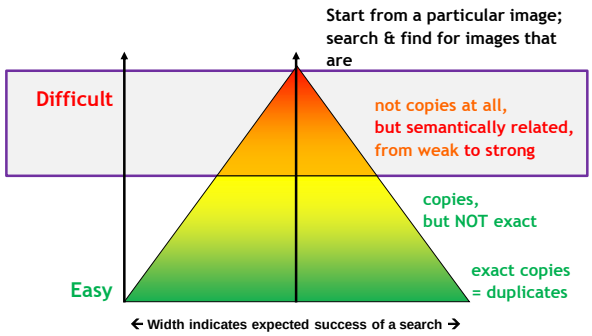
and in this way also its location on some WWW page and site that can provide us with more information about the image.

Searching by image: Research problem

Can search by image reveal images that are '*semantically*' or 'theme' or 'content' related to the image that is submitted as a query?

51

Searching by image: searching for related images



52

Searching by image: publication

P. Nieuwenhuysen

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, 476 pp, ISBN 978-7-5439-6289-7. pp. 145-155.
full text also available free of charge from
http://www.vub.ac.be/BIBLIO/nieuwenhuysen/presentations/

53

Searching by image: Findings: Example in 2013

→

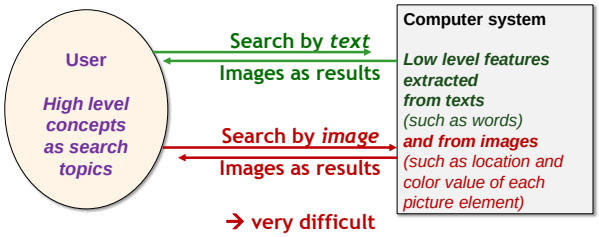
54

Searching by image: Findings in 2013

- 20 test queries by image file were executed.
- For each test, *Google* showed images that were visually similar in
  - colors,
  - shapes and/or
  - textures.
- For 19 / 20 tests, NO image was found that is *semantically* related.
  = Only 1 / 20 test yielded a semantically related image.

Searching by image: Discussion

Semantic searching and the semantic gap in the case of searching for images

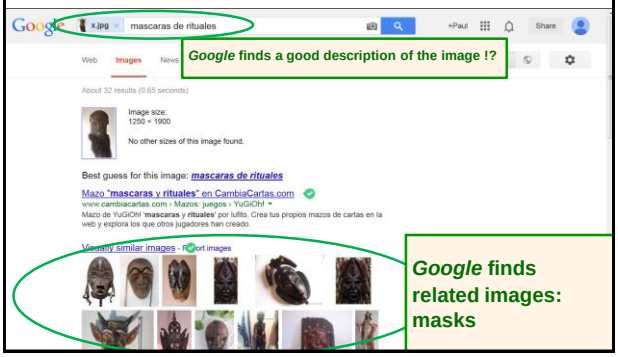


→ very difficult to reveal semantically related images

Searching by image: Findings: Example in 2014

- Source image, renamed to x.jpg to remove any relation in the form of text to images on the WWW

Searching by image: Findings: Example in 2014



Searching by image: Discussion

Step 1:
Source image
+ *Google* retrieval system
→ "best guess for the image" in words !

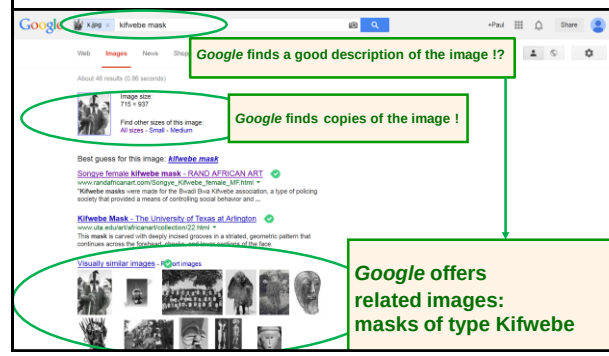
Step 2:
Source image
+ *Google* retrieval system
+ "best guess for the image" in words !
→ semantically related images !!

Looks like only 1 search action in the eyes of the user

Searching by image: Findings: Example in 2014

- Source image = famous photo of Kifwebe mask
- Renamed to x.jpg to remove any relation in the form of text to images on the WWW

Searching by image: Findings: Example in 2014



61

Applications of searching by image: *Finding semantically similar images*

- Starting from a source image, search by image can (even) be successful to find
  - » ! a suitable description in words of the image !
  - » !! images that are semantically related !!

62

Searching for images, using *an image* as query

General conclusion

63

Searching by image: *General conclusion*

- Searching by image is evolving to a powerful, additional method to meet information needs.
- This method exploits the increasing number of images on the WWW plus the related texts.

64

Searching by image: *General conclusion*

→
To assist their clients,
librarians and other information specialists
can apply search by image
to tackle some information problems.