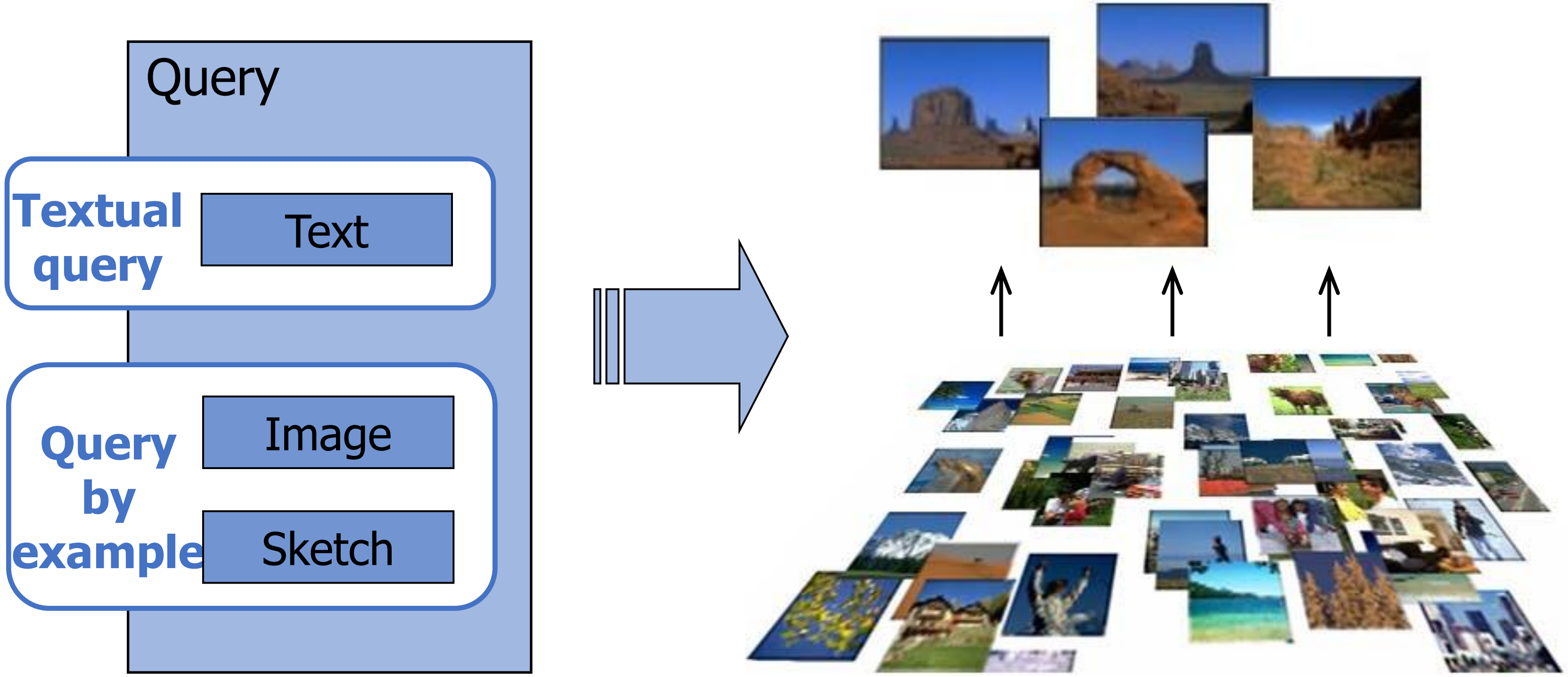
Content-based Image Retrieval

Lecture Outline

- What is and Why image retrieval?
- How to compare and retrieve images?
  - Common components of the CBIR systems
  - Main problems and research directions
- Performance measurement
  - Retrieval effectiveness

What is image retrieval?

- Description Based Image Retrieval (DBIR)
- Content Based Image Retrieval (CBIR)



DBIR vs. CBIR

	DBIR	CBIR
+	▪ Fulltext search algorithms are applicable ▪ Search results corresponds to image semantics	▪ Automatic index construction ▪ Index is objective
−	▪ Manual annotating is hardly feasible ▪ Manual annotations are subjective	▪ Semantic gap ▪ Querying by example is not convenient for a user

Levels of image retrieval

- Level 1: Based on color, texture, shape features
  - Images are compared based on low-level features, no semantics involved
  - A lot of research done, is a feasible task
- Level 2: Bring semantic meanings into the search
  - e.g., identifying human beings, horses, trees, beaches
  - Requires retrieval techniques of level 1
  - Very active and challengeable research area
- Level 3: Retrieval with abstract and subjective attributes
  - Find pictures of a particular birthday celebration
  - Find a picture of a happy beautiful woman
  - Requires retrieval techniques of level 2 and very complex logic

Why image retrieval?

- Huge amounts of images are everywhere: how to manage this data?
- “A Picture is worth thousand words”
- Not everything can be described in text

Why content-based image retrieval?

- Automatic generation of textual annotations for a wide spectrum of images is not feasible.
- Annotating images manually is a cumbersome and expensive task for large image databases.
- Manual annotations are often subjective, context-sensitive and incomplete.
- Google, Yandex and others use text-based search. Results are not perfect.
  However, now it is much better, than a couple of years ago!

Image retrieval by Google

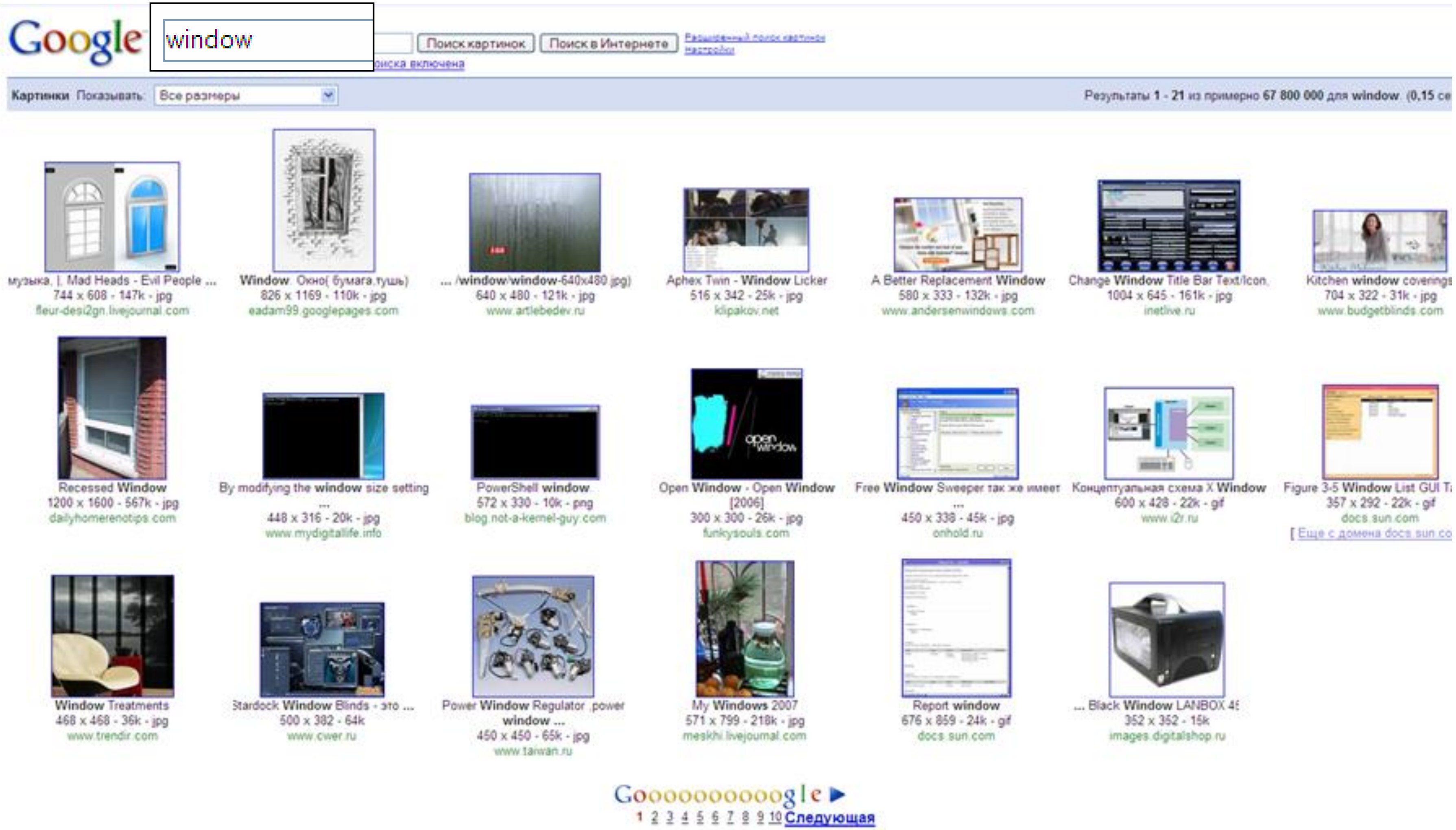
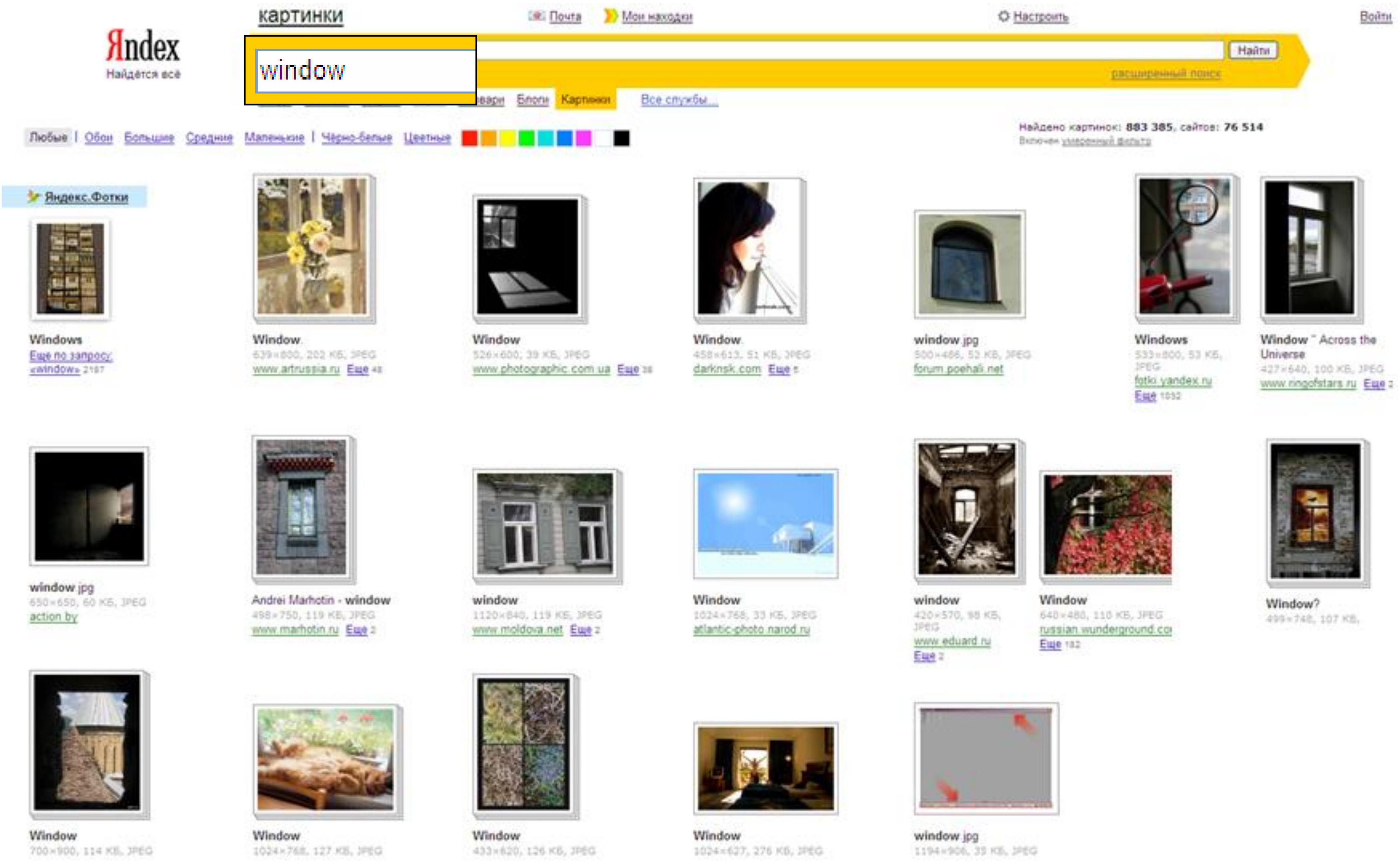


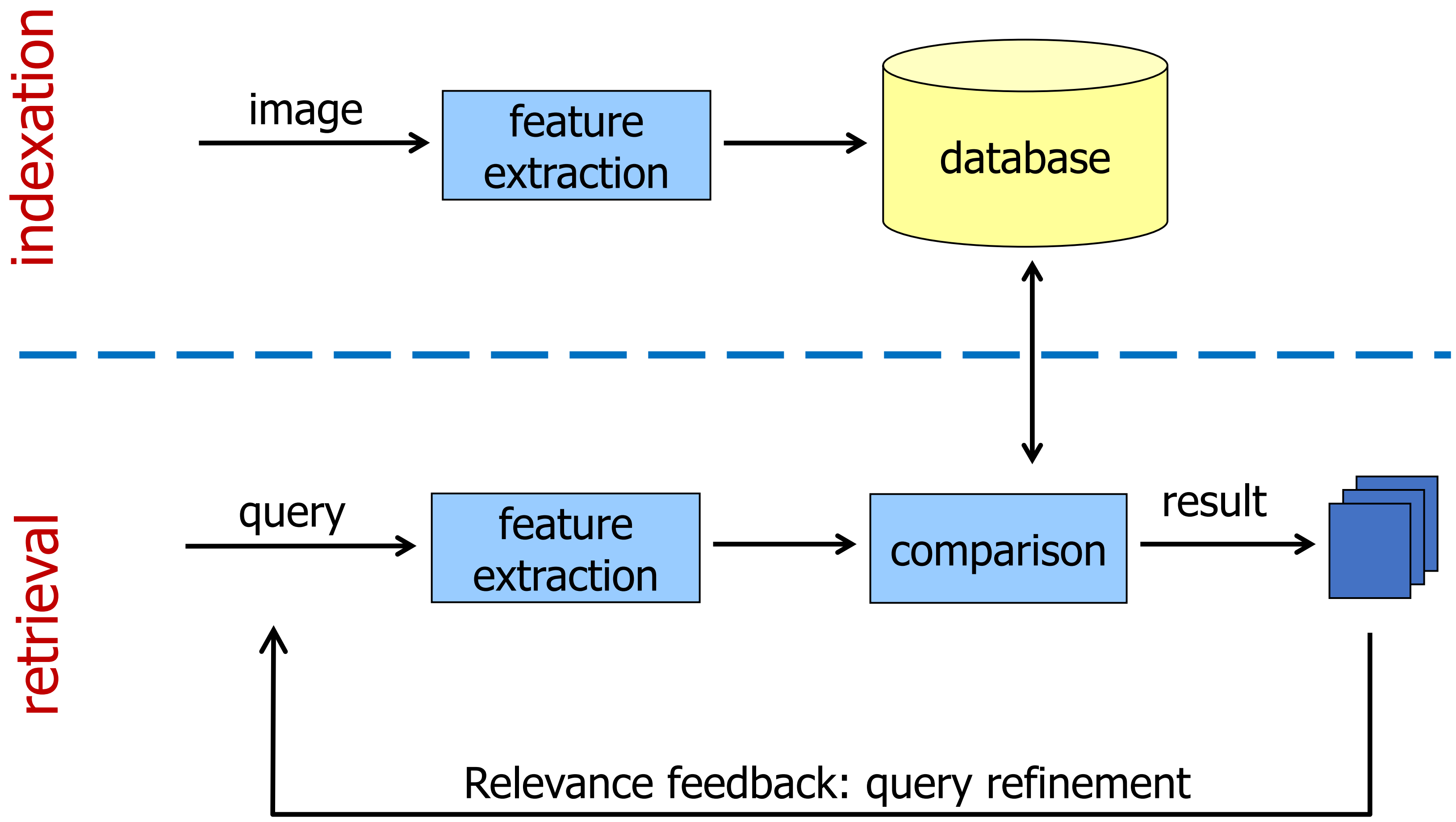
Image retrieval by Yandex



Lecture Outline

- What is and Why image retrieval?
- How to compare and retrieve images?
  - Common components of the CBIR systems
  - Main problems and research directions
- What are applications?
- Performance measurement
  - Retrieval effectiveness

Common components of CBIR system



Lecture Outline

- What is and Why image retrieval?
- How to compare and retrieve images?
  - Common components of the CBIR systems
  - Main problems and research directions
- Performance measurement
  - Retrieval effectiveness

Problems and directions

- Feature extraction
  - How to represent an image in a compact and descriptive way?
  - How to compare features, and, thus, images?

How to: Image features

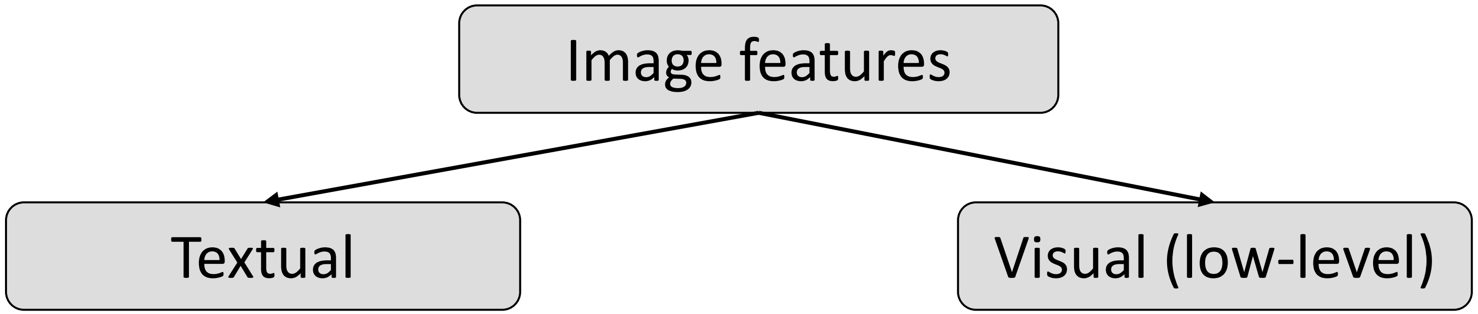
Levels of image content

- Semantics
- Shape
- Texture
- Color, lightness

Textual/metadata features

Low-level features / visual features (signatures, descriptors)

How to: Image features



Annotations and metadata:

- tags/keywords;
- creation date;
- geo tags;
- name of the file;
- photography conditions (exposition, aperture, flash…).

Features extracted from pixel values:

- color descriptors;
- texture descriptors;
- shape descriptors;
- spatial layout descriptors.

How to: Image features

Low-level features

Global

Local

Describes the whole image:

- average intensity;
- average amount of red;
- …

All pixels of the image are processed.

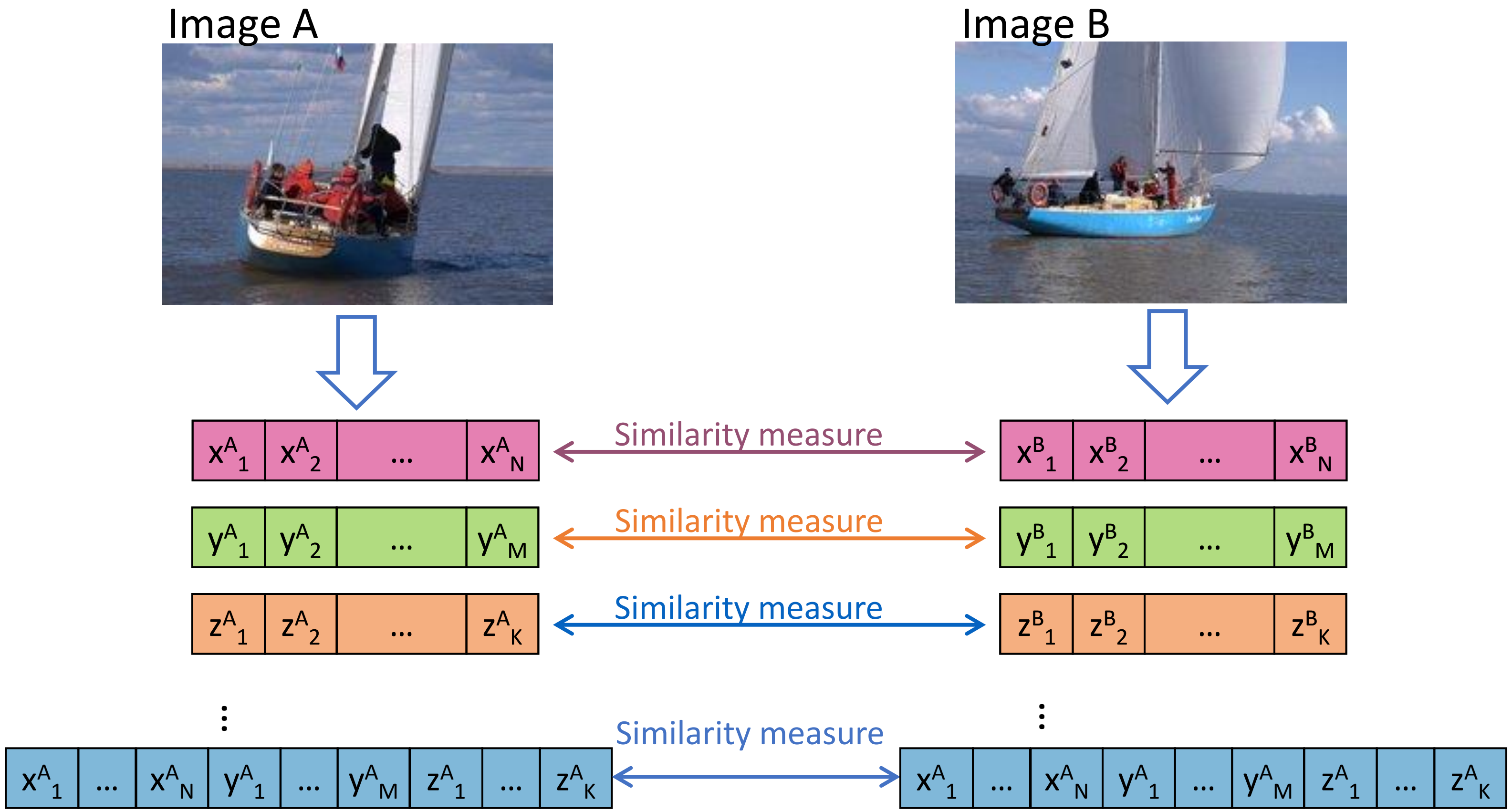
Describes one part of the image:

- average intensity for the left upper part;
- average amount of red in the center of the image;
- …

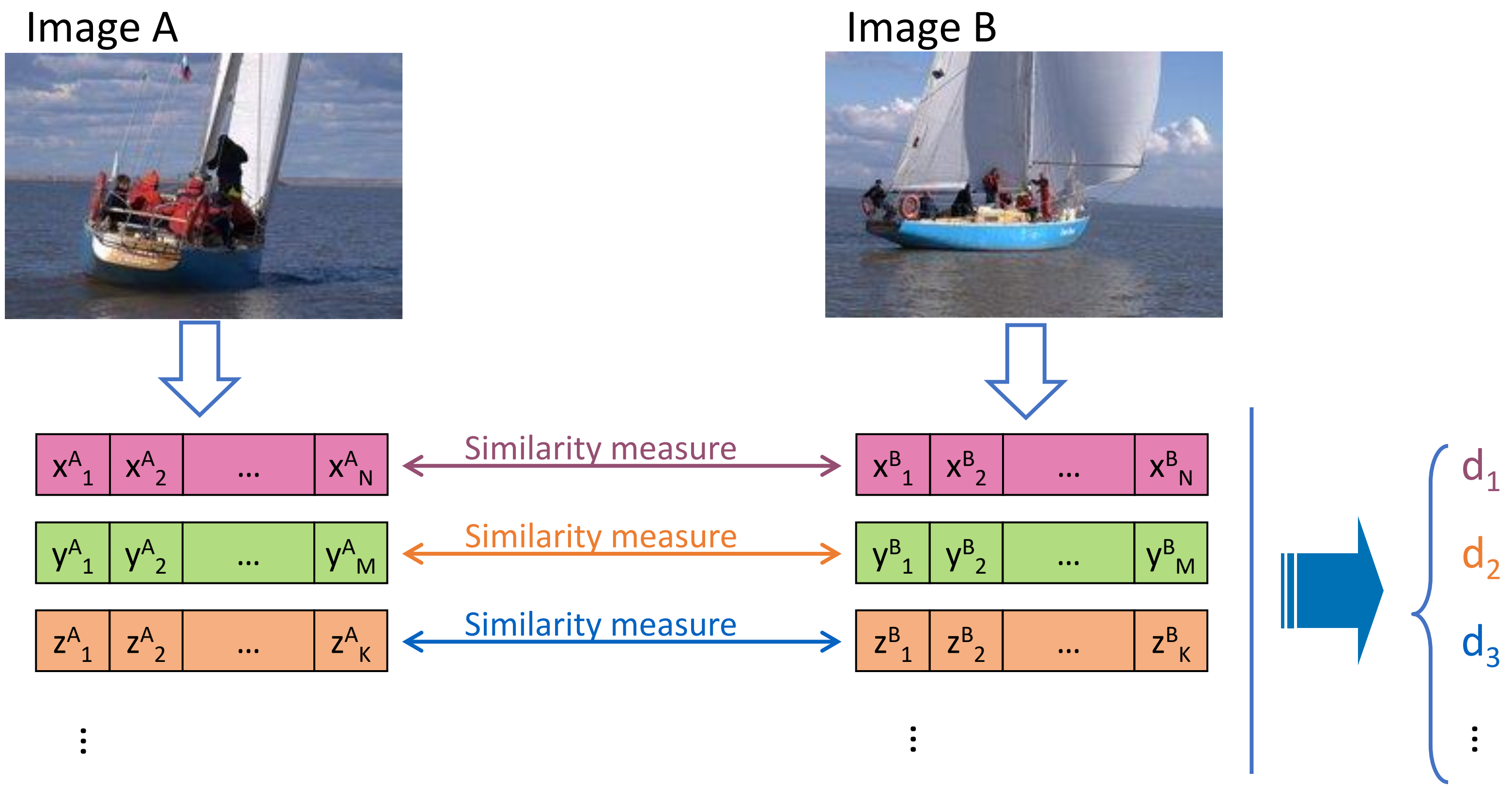
Segmentation of the image is performed, pixels of a particular segment are processed to extract features.

How to: Feature spaces

- Feature vector – a vector of features, representing one image.
- Feature space – the set of all possible feature vectors with defined similarity measure.



How to: Combine results



$$D = \sum_i c_i d_i$$

Similarity Measure - Manhattan Distance

- It is also called L_1 distance. If u(x1, x2) and v(y1, y2) are two points, then the Manhattan distance between u and v is given by:
  - $MH(u, v) = |x_1 - y_1| + |x_2 - y_2|$

- Instead of two dimensions, if the points are represented by n-dimensions, such as $a = (x_1, x_2, \ldots, x_n)$ and $b = (y_1, y_2, \ldots, y_n)$ then the Manhattan distance between a and b is calculated as follows:
  - $MH(a, b) = |x_1 - y_1| + |x_2 - y_2| + \ldots + |x_n - y_n|$ = $\sum_{i=1}^{n} |x_i - y_i|$

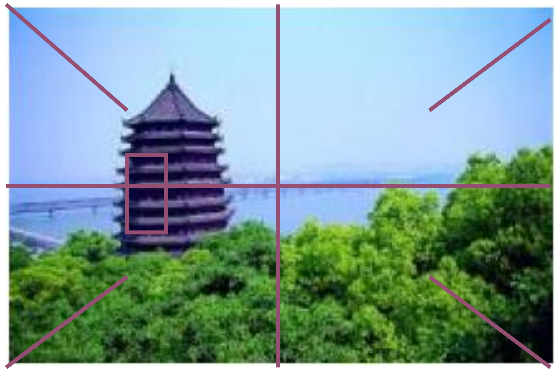
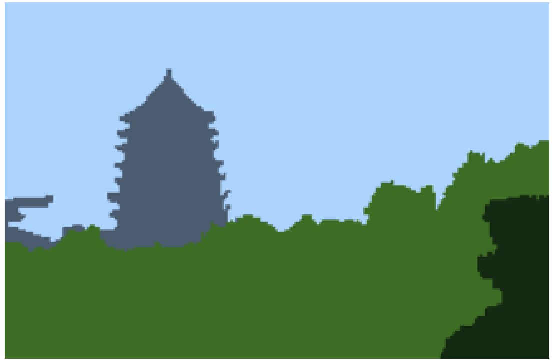
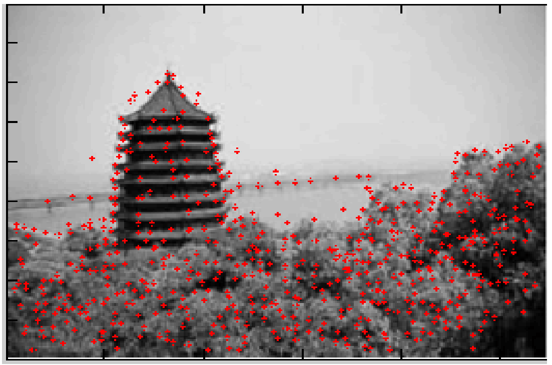
Similarity Measure - Euclidean Distance

- It is also called L_2 distance. If u(x1, x2) and v(y1, y2) are two points, then the Euclidean distance between u and v is given by:
  - $EU(u,v) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2}$

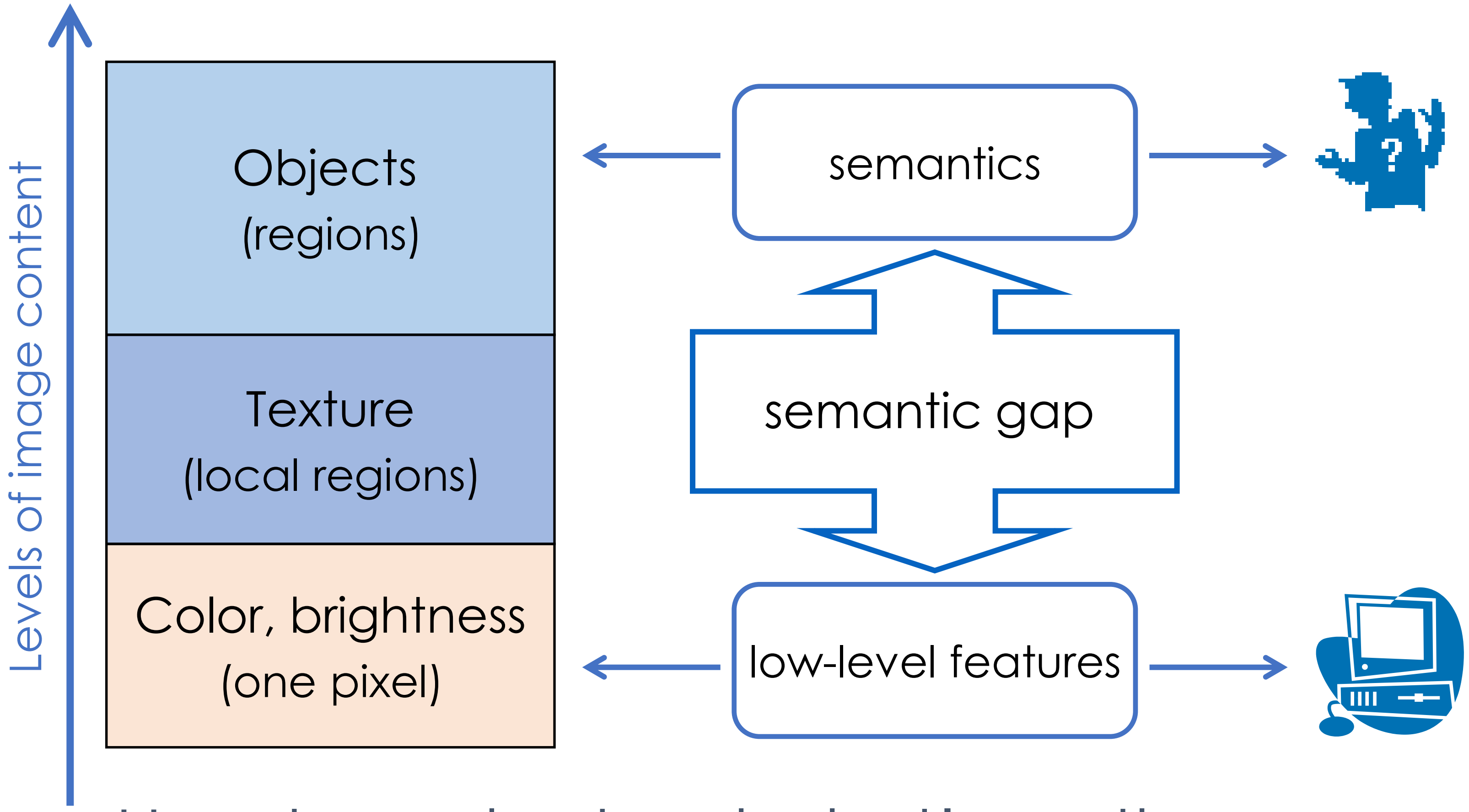
- Instead of two dimensions, if the points are represented by n-dimensions, such as $a = (x_1, x_2, \dots, x_n)$ and $b = (y_1, y_2, \dots, y_n)$ then the Euclidean distance between a and b is calculated as follows:
  - $EU(a,b) = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \dots. + |x_n - y_n|^2}$
    $= \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$

How to: Image segmentation

- Fixed regions
  - The same region boundaries for all images.
- Segmentation
  - Boundaries depends on image content.
- Key points (point of interest) detection
  - Points of particular interest in the image, feature extraction for areas around key points.

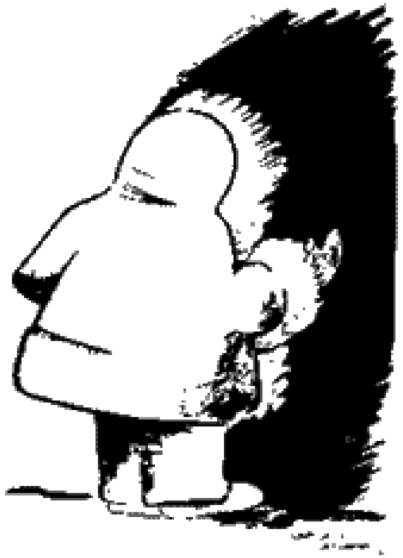




Problems: semantic gap



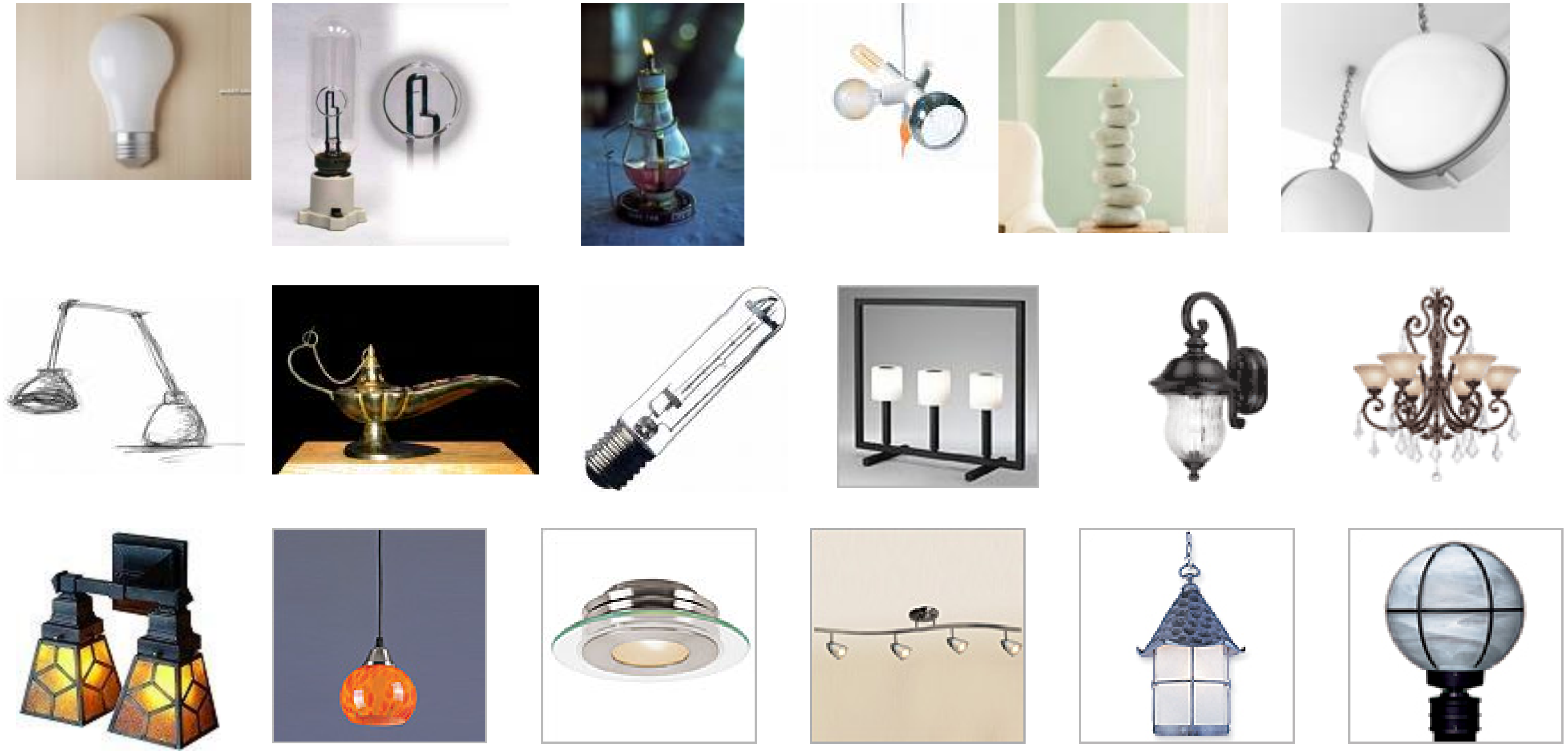
How to understand what’s on the images?

Problems: what’s on the images?



- Sometimes it is not easy to understand the image even for humans!
- What do we want from machines?

Problems: what’s on the images?



- How do we now that all these objects are lamps?

Lecture Outline

- What is and Why image retrieval?
- How to compare and retrieve images?
  - Common components of the CBIR systems
  - Main problems and research directions
- Performance measurement
  - Retrieval effectiveness

Performance measurement

Performance concerns

- Efficiency
  - Important due to the large data size
- Retrieval effectiveness
  - No similarity metric which exactly conforms to human perception

Problems in effectiveness evaluation

- Define a common image collection
  - Corel Photo CDs
  - Brodatz texture collection: http://www.ux.uis.no/~tranden/brodatz.html
  - CoPhIR: http://cophir.isti.cnr.it/whatis.html
  - Participate in ImageCLEF, TRECVID, imageEVAL, ROMIP
- Obtain relevance judgement
  - Use of collections with predefined subsets (Corel collection)
  - Image grouping (medical)
  - User judgements
    - Pooling
    - Different types of judgement data (relevant – not relevant, ranking, …)

Effectiveness measurement

- “You can see, that our results are better”

Effectiveness measurement (1)

- Recall and precision

$$precision = \frac{\text{No. relevant documents retrieved}}{\text{Total No. documents retrieved}},$$

$$precision = \frac{true\ positive}{true\ positive + false\ positive}$$

$$recall = \frac{\text{No. relevant documents retrieved}}{\text{Total No. relevant documents in the collection}}$$

$$recall = \frac{true\ positive}{true\ positive + false\ negative}$$

- F-measure
  - F1 score is the harmonic mean of precision and recall

$$F1\ Score = 2 * \frac{precision\ * recall}{precision + recall}$$

Effectiveness measurement (2)

- Precision at N (P@n)
  - Precision takes all retrieved documents into account, but it can also be evaluated at a given cut-off rank, considering only the topmost results returned by the system. This measure is called precision at n or P@n.

- Recall at N
  - Recall at N is the same as the normal recall, but Recall at N is applied into the top-n retrieved image.

Effectiveness measurement (3)

- Error rate

$$\text{Error rate} = \frac{\text{No. non-relevant images retrieved}}{\text{Total No. images retrieved}}$$

Effectiveness measurement (4)

- **Mean Average Precision (MAP)**
  - The mean average precision which averages precision across all queries Q = $\{q_1, \ldots, q_\mathrm{n}\}$.
  - $MAP = \frac{\sum_{q_i \in Q} precision\ (q_i)}{|Q|}$

Effectiveness measurement (5)

- **Mean Reciprocal Rank (MRR)**
  - the mean reciprocal rank evaluates how good the search mechanism is in ranking the relevant images.
  - The main focus of this metric is to find the order of the first relevant image among the retrieved images.
  - Assume that r_q is the rank of the first relevant image for a query q, then the reciprocal rank of q is computed as follows.
    - $RR(q) = \begin{cases} 0 : not\ relevant \\ \frac{1}{r_q} : \ \ otherwise \end{cases}$
  - By considering the multiple participants' responses of each query, the final reciprocal rank per query averages the responses among all users. Thereafter, the mean reciprocal rank for all queries is calculated as follows.
    - $MRR = \frac{1}{|Q|} \sum_{q \in |Q|} RR(q)$