# REGIMvid at ImageCLEF2012: Concept-based Query Refinement and Relevance-based Ranking Enhancement for Image Retrieval

Rim Fakhfakh, Ghada Feki , Amel Ksibi, Anis Ben Ammar and Chokri Ben Amar

REGIM: REsearch Group on Intelligent Machines,
University of Sfax, ENIS, BP W, 3038, Sfax, Tunisia

{fakhfakhrima, ghada.feki}@yahoo.fr

{amel.ksibi, anis.benammar, chokri.benamar}@ieee.org

**Abstract.** In this paper, we present our proposed approach that deals with two important steps in the retrieval process, which are query analysis and relevance-based ranking. First, query analysis takes into account two forms of queries: textual and visual which present the form of queries provided by ImageCLEF in the task "Visual concept detection, annotation, and retrieval using Flickr photos ". The main idea in the query analysis process is to interpret the textual and visual queries and select the most appropriate concepts according to the query to concept mapping. Even the list of concepts is retained, we perform a query refinement process to improve the quality of the interpretation of the query. Second, we try to achieve a high-quality relevance-based result not only with choosing an adequate similarity measure but also with enhancing obtained scores by elaborating a random walk with restart, which is performed over inter-images semantic similarity graph.

**Keywords:** Concept based image retrieval, Query to concept mapping, Relevance based ranking, Inter-concept similarity graph, Inter-images semantic similarity graph.

## 1    Introduction

This paper presents the second participation of our group REGIMvid, within REGIM research unit, in the "Visual concept detection, annotation, and retrieval using Flickr photos " task [1].

The aim in this task is to analyze a collection of Flickr photos in terms of their visual and/or textual features in order to detect the presence of one or more concepts. The detected concepts can then be used to automatically annotate

images or to retrieve the best matching images to a given concept-oriented query. The main challenge for this task is to select the most appropriate concepts related to a query and so to ensure a good relevance-based result ranking. For the selection of the most adequate concepts from a query, we execute a query to concept mapping requiring a calculation of the semantic similarity measures between query keywords and concepts.  Concerning the ranking step, we try to achieve a high-quality relevance-based result not only with choosing an adequate similarity measure but also with enhancing obtained scores by elaborating a random walk with restart. This refinement step is based on information, which are occurred from the inter-images semantic similarity graph.
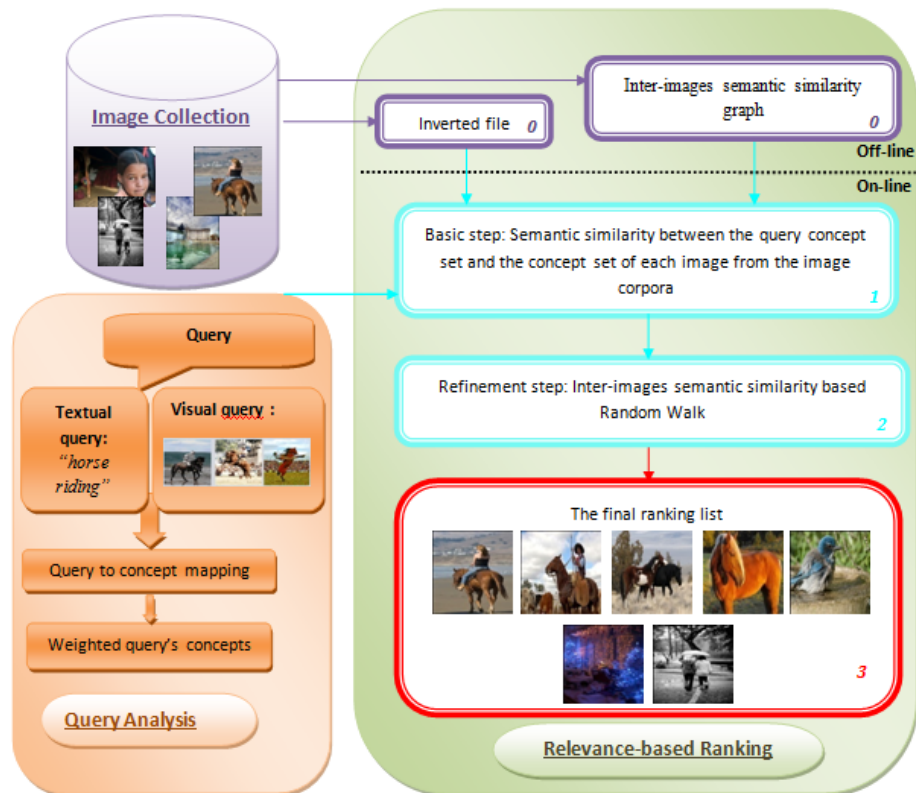


**Fig. 1.** General scheme

The rest of the paper is organized as follows. Section 2 describes our proposed relevance-based approach; Section 3 discusses the experimental results.

## 2 Relevance-based approach: Visual concept retrieval using Flickr photos task

Relevance computing in our approach implies 3 phases : First, we analyze the query to enhance the user request understanding . Second, we compare query concepts with concepts that annotate each image from the collection. Finally, scores obtained in the previous step are refined by the means of random walk with restart.

The following notations will be used. Given a set of query concepts $C_q$ = $\{c_1^q, c_2^q, .., c_{|C_q|}^q\}$, denote by $D_q$ = $\{x_1^q, x_2^q, .., x_{|D_q|}^q\}$ the collection of images that are associated with the set of query concepts $C_q$. This collection, which is a part of the large collection $D = \{x_1, x_2, .., x_{|D|}\}$, is obtained by the inverted file construction in an off-line stage.

For image $x_i$ , denote by $C_i = \{c_1^i, c_2^i, .., c_{|C_i|}^i\}$ the set of its associated concepts. The relevance scores of all images in D are represented in a vector $r = [r(x_1), r(x_2), .., r(x_{|D|})]^T$, whose element $r(x_i)$ denotes the relevance score of image $x_i$ with respect to the set of query concepts $C_q$.

$$\begin{cases} r(x_i) > 0 \; if \; x_i \in D_q \\ r(x_i) = 0 \; \text{otherwise} \end{cases} \qquad (1)$$

Relevance score $r(x_i)$ reflects the degree of the existence of a given concept in the image $x_i$. This score is normalized that we range it from 0 to 1.

### 2.1 Query analysis

Query analysis in a retrieval process is a fundamental and critical step. The main challenge is to enhance the interpretation of user request in order to improve retrieved result quality. The proposed process takes into account the query format: title, description and images (fig2). Firstly, a textual analy-

sis is performed on title and description components in order to deduct the main concepts within text. Secondly, a matching process is performed based on a concept within query and image data. The obtained results are, thirdly, merged. Finally, a refinement process is executed.

The following figure shows the query analysis process which goal is the selection of the most appropriate concepts related to each query.
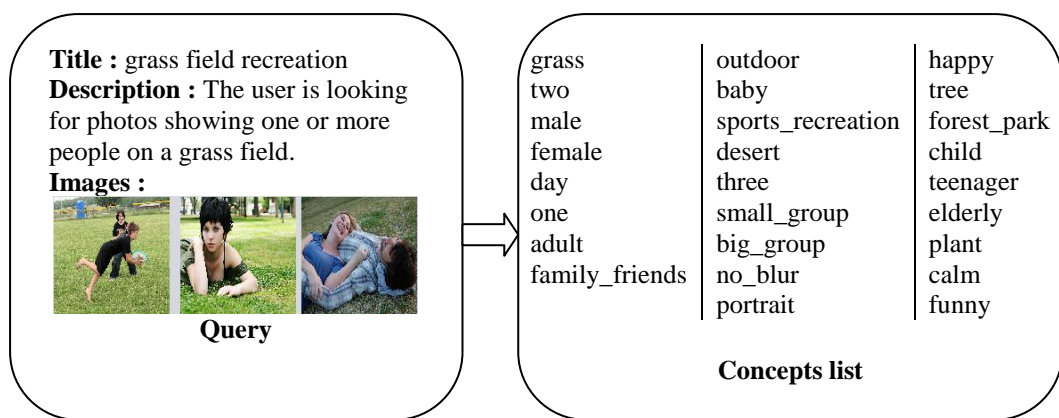


**Fig. 2.** : Overview of query analysis process

**Textual query analysis .**

The proposed treatments within textual analysis inherit from natural language processing standard. First we extract the representative information by the elimination of non informative terms called stoplist [2]. Keywords extracted from the initial query are divided into positive and negative terms by following a linguistic study: the positive keywords represent what is required in the query and the negative ones represent some details that should be avoided. Fig3 shows an example for a query keywords classification into positive and negative.
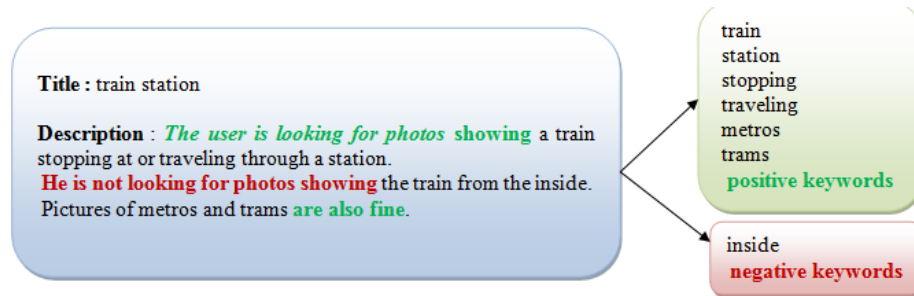
**Fig. 3.** Positive and negative keywords classification

Each retained keyword in the query representation is projected on the concept space; a query to concept mapping. Using WordNet [3] as taxonomy, we evaluate the similarity values between a keyword and the 94 proposed concepts. We retain the most related concept to each query term.

**Visual query analysis .**

Regarding the initial query textual description, we distinguish between positive and negative images. Image is considered as positive if it translates the textual description and negative if there are some particularities which are not congruent to the textual description.
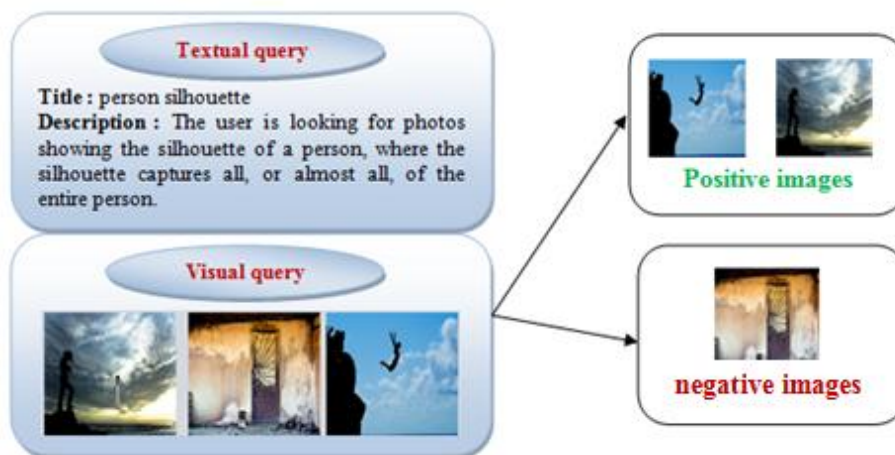


**Fig. 4.** Positive and negative image classification

Alike the textual treatment, we deduce a set of positive and negative concepts. The related concepts per image are extracted from the annotation file [4].

In some query, the same concept appears in positive and negative images. In this particular case, we retain this concept as positive.

**Textual and visual concepts fusion .**

Textual and visual analyses bring out a set of positive and negative concept per each modality. The obtained sets are firstly merged. We compute a weight value for each positive concept.

Regarding the concept frequency in textual and visual component, the associated weight is computed as follows:

$$\text{Weight(C)} = \frac{\text{Concept Frequency (C)}}{\text{Max\_Concept\_Frequency}} \qquad (2)$$

When the same concept appears in the positive and negative sets, we decide to maintain it as positive or negative as follows.

Let's note:

- w: the weight of the redundant concept
- moy: the average of all weights of positive concepts
- nb: the number of positive concept
- σ: the standard deviation between weight of positive concept.

According to the following formula we decide if either the redundant concept is positive or negative.

If $w \in [\text{moy-} \sigma/\sqrt{nb} , \text{moy+} \sigma/\sqrt{nb}]$, the concept is considered as positive and keeps the same weight else, we cancel its weight value.

**Query refinement.**

The final stage of the proposed process is an inter-concept relation study which goal is to enhance the query interpretation. Two possible scenarios can occur:

- Query expansion: when adding new concepts to the previous set.
- Concepts reweighting: when modifying existing concepts weights.

We built an inter-concepts semantic graph where concepts are inter-connected between them. Weights within edges represent the semantic similarity value computed according to the WordNet hierarchy.

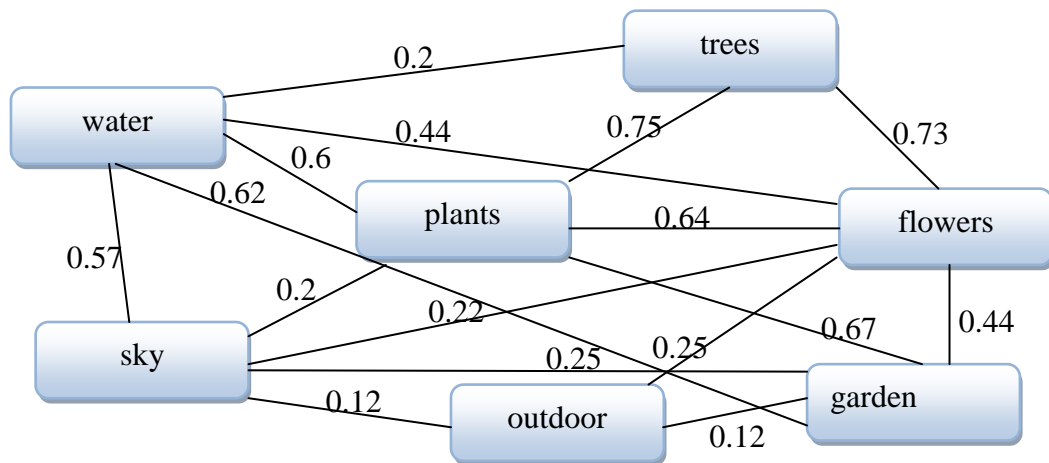Fig 5 shows a partial view of the semantic graph:



**Fig. 5.** inter-concepts semantic graph

We perform a random walk over the semantic graph to refine our concepts' list either by reweighting concepts or by adding new concepts according to the semantic similarity measure between concepts. We attempt to use another type of graph called contextual graph [5] to bring the best refinement.

We also use another graph called contextual one [5] to bring the best refinement. This last, is built based on the contextual information inter-concepts dealing with the co-occurrence by exploring Flicker resources as a largest public available multimedia corpus.

To more refine our query after random walk step and the addition of new relevant concepts, we try to manage the semantic conflict [6] between the selected concepts: exclusion and implication. An exclusion conflict traduces a contradiction between selected concepts, i.e concepts "indoor" and "outdoor" cannot simultaneously occur in the same query. For such case, we have to decide which one to eliminate. The implication case is implied when two concepts are closely related (according to the semantic similarity value). In such case, concepts in relation with theses in the query are added to this last.

## 2.2    Query-Image similarity

Based on experimental semantic similarity measures study [7], we decide to adopt an approach for semantic similarity between a given query and an image that is analogous to Cosine similarity measure. The semantic similarity between $C_q$ and $C_i$ , which are respectively the sets of query q concepts and image $x_i$ concepts, is defined as:

$$similarity(C_q, C_i) = \frac{\sum_{t=1}^{|C_i|} c_t^q \times c_t^i}{\sqrt{\sum_{t=1}^{|C_q|} c_t^{q^2}} \times \sqrt{\sum_{t=1}^{|C_i|} c_t^{i^2}}} \qquad (3)$$

This semantic similarity is computed between the query concepts and each concept sets of images that belong to the sub-collection $D_q$ relative to this query. For other images belonging to $D$, their similarity scores are equal to zero.

Instead of searching through a large collection, a user query is concerned in only a part of selected image results from the whole collection. An inverted file is a data structure where images are indexed by (concept × image) structure instead of the conventional (image × concept) structure. It allows saving much time since the algorithm search concerns only the sub-collections.

We denote by $v$ the vector of semantic similarity between a query and the collection $D$. It is defined as follows:

$$\vec{v} = \left[ similarity(C_q, C_1), similarity(C_q, C_2), \ldots, similarity(C_q, C_{|D|}) \right]^T \quad (4)$$

This vector will be used as an input for refinement phase, which is provided thanks to a random walk with restart.

## 2.3    Result refinement

We try to achieve a high-quality relevance-based result not only with choosing an adequate similarity measure but also with improving obtained scores. Therefore, elaborating a random walk [8] is an attempt for enhancing results.

In this section, we denote by $W$ a similarity matrix whose element $W_{ij}$ indicates the similarity between images $x_i$ and $x_j$ and the elements on the diagonal are null.

Semantic similarity scores vector according to a given query and the inter-images semantic similarity graph $W$ are the input of random walk process [9]. It is defined as: consider a random image that starts from node i, the image iteratively transmits to its neighborhood with the probability that is proportional to their edge weights. Also at each step, it has some probability c to return to the node i.

$$\vec{r} = c \times \widetilde{W} \times \vec{r} + (1 - c) \times \vec{v} \tag{5}$$

After some iteration we have an optimal scores vector $\vec{r}$.

The inter-images semantic similarity graph $W$ illustrates the semantic relationship between images in the collection. It informs on the semantic similarity degree between each pair of images. We adopt an approach for semantic similarity between tow images that is analogous to Jaccard similarity. The similarity between $C_i$ and $C_j$ , which are respectively the sets of image $x_i$ concepts and image  $x_j$ concepts, is defined as:

$$similarity\left(C_i\,,C_j\right) = \frac{\sum_{t=1}^{|C_i|} c_t^i \times c_t^j}{\sum_{t=1}^{|C_i|} c_t^{i\,2} + \sum_{t=1}^{|C_j|} c_t^{j\,2} - \sum_{t=1}^{|C_i|} c_t^i \times c_t^j} \tag{6}$$

The semantic similarity graph consists in the matrix W, which must be normalized. In fact, there are several ways to normalize the weighted matrix W.

The most natural way might be the row normalization. Complementarily, we normalize W using the normalized graph Lapalician, as follows:

$$\widetilde{W} = D^{-1/2} \times W \times D^{-1/2} \tag{7}$$

With the vector D is defined as:

$$D_i = \sum_{j=1}^{|D|} W_{ij} \tag{8}$$

The refinement method consists in a random walk process, and it will converge to a fixed point. We use the graph that illustrates relations between images for the random walk since the relevance scores of semantic similar images should be close. The convergence condition consists in comparing the difference between two successive scores vectors resulting by the Random Walk with a fixed parameter ε.

## 3    Experimental results

We describe the experimental study conducted to evaluate the proposed approach within the relevance computing. The experiments were performed on ImageCLEF12 benchmark.

The relevance-based approach for image retrieval is experimented with the Visual concept retrieval using Flickr photos1 task. It consists in analyzing a Flickr photos collection. We look for extracting concepts from the visual and textual features. These concepts are used in image retrieval.

The global results of our system are relatively bad. This fact is explained by the huge number of selected concepts related to each query. This flow of information can deviate the real meaning of the initial query.

For example the treatment of the query n°29 "sleeping baby" with the description "*The user is looking for photos showing a sleeping (calm, quiet) ba-*

---

[1]http://imageclef.org/2012/photo-flickr

*by"* , provides a list of concepts composed from 11 items which are:  *one, baby, partial_blur, day, no_blur ,portrait, indoor, happy, calm, overlay, smoke*.

We notice that the appearance of wrong concepts is the consequence of the query to concept mapping shortage. For example, the keywords  *"sleeping"* and *"quiet"* appearing in the previous textual query are respectively projected to *"smoke"* and *"day"* after  performing the step of query to concept mapping.

For the query n°5 "hot air ballon", selected concepts after query analysis process are far from the user request. In fact, these concepts are textually close to the query (air ballon, airplane) but semantically different.

The deep study of queries individually shows that our approach can perform well in some cases. for expamle , the query  n°25 "grass field recreation" previously mentioned in Fig2. In effect, the majority of extracted keywords from the textual query belong to the concepts set. (grass is a concept)

Finally, we notice that the setting of the parameter c in the random walk has relatively minor impact on the ordering of the images in the result list.

## 4      Conclusion

In this work, we present our proposed approach that deals with two important steps in the retrieval process, which are query analysis and relevance-based ranking. Query analysis takes into account two forms of queries: textual and visual which present the form of queries provided by ImageCLEF. Relevance-based result is based on choosing an adequate similarity measure and enhancing obtained scores by elaborating a random walk with restart, which is performed over inter-images semantic similarity graph.

## References

1.  Amel Ksibi, Anis Ben Ammar and Chokri Ben AmarREGIMvid at ImageCLEF2011: Integrating Contextual Information to Enhance Photo Annotation and Concept-based Retrieval,2011

2.  G. Salton, M. Mcgill,  Introduction to Modern Information  Retrieval. McGraw-Hill, 1983

3.  Christiane Fellbaum WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press, 1998.

4.  M. Zarka, N.Elleuch, A.B. Ammar, A.M. Alimi, A Fuzzy Ontology-Based Framework for rea-soning in Visual Video Content Analysis and Indexing, In the intl Workshop on Multimedia Data Mining, 2011

5.  Amel Ksibi, Anis Ben Ammar, Chokri Ben Amar, Integrating Contextual Information to En-hance Photo Annotation and Concept-based Retrieval, 2011.

6.  Sabrina Tollari, Marcin Detyniecki, Christophe Marsala, Ali Fakeri Tabrizi, Massih-Reza Amini, Patrick Gallinari, "Exploiting Visual Concepts to Improve Text-Based Image Retriev-al", European Conference on Information Retrieval (ECIR), 2009.

7.  S.-H. Cha, Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions, International Journal Of Mathematical Models And Methods In Ap-plied Sciences,Issue 4, Vol 1, 2007.

8.  H. Tong, C. Faloutsos, J.-Yu Pan, Random walk with restart: fast solutions and applica-tions, KnowlInfSyst (2008) 14, 327-346.

9.  K. Yang, M. Wang, X.-Sheng Hua and H.-J. Zhang, Social Image Search with Diverse Rele-vance Ranking, S. Boll et al. (Eds.): MMM 2010, LNCS 5916, 174–184, Springer-Verlag Ber-lin Heidelberg 2010.