Image Search Results Diversification

Mihai Lupu**, João Palotti**, Navid Rekabsaz**, Adrian Popescu*, Adrian Iftene[†], Pinar Duygulu Sahin[‡]

*CEA, LIST, LVIC

** Vienna University of Technology, ISIS, IMP

[†] "Al. I. Cuza" University

[‡] Bilkent University

Contacts: lupu@ifs.tuwien.ac.at, adrian.popescu@cea.fr

MUCKE Project, Deliverable 5.1

01/03/2015

Contents

Abstract

This deliverable covers both text-based and image-based diversification methods. For image-based we show the method we have developed for our participation in MediaEval 2014-2015 Retrieving Diverse Social Images task. It uses an array of clustering methods to identify optimal diversification results. For text-based, we show an ontology-aware method tested on publicly available image sets.

1 INTRODUCTION

While an important amount of work was dedicated to the subject, result diversity in multimedia information retrieval remains an open problem. A vast majority of current approaches are based on result clustering (applied to textual, visual content or to both types), in order to diversify results. While clustering approaches are well adapted to visually coherent queries (which are often used as example in related literature), they are less likely to work well for more general queries, for which a semantic decomposition of the query is difficult to obtain using clustering. In these latter cases, diversification is usually accompanied by important result quality degradation. As an alternative, we will introduce a more flexible diversification method, which adapts the diversification strategy to the given query. Each query is split into concepts and its generality is evaluated according to the visualness of query components and of their conceptual neighbourhoods obtained with multimodal concept similarity. Results for queries that include visually coherent concepts (i.e. wild turkey) will be diversified using multimedia clustering whereas more general queries (which cover a larger conceptual space – i.e. bird) will be first split into more specific concepts and results clustering will be used only for these specific concepts. For complex queries, which generate a large number of possible reformulations, the multimedia concept similarity module will be used in order to determine which should be the focus of the reformulation (via an analysis of the relation of top related concepts to concepts in the initial query). The semantic similarity based reformulation will produce intuitive query decompositions and will also boost the results of multimedia clustering, which will be applied in visually coherent spaces. Given that the query space is potentially infinite, one important feature of the system will be to label queries as possible to be process with the resources created in the project, or not. When queries fall outside of the area covered by our resources, appropriate action will be taken to inform the user.

1.1 RELATED WORK

Over time, various theories involving search results diversification have been developed, theories that have been taken into consideration [3]: (i) content [5], i.e. how different are the results to each other, (ii) novelty [1], [2], i.e. what does the new result offer in addition to the previous ones, and (iii) semantic coverage [13], i.e. how well covered are the different interpretations of the user query. In the MUCKE project, we work with a collection of approximately 80 million images and their associated metadata that have been downloaded mainly from the Flickr database. Over this collection, we perform several processing tasks at both textual (on associated metadata) and image level and retrieve the results in a diversified way.

2 IMAGE DIVERSIFICATION

2.1 CUSTERING

We worked on three methods for clustering, all based on similarity measures. They share the idea of creating a similarity graph (potentially complete) in which each vertex represents an image for one point of interest, and each edge represents the similarity between two images. Different similarity metrics and different set of features can be used. Next, we explain each algorithm and how we combined them.

2.1.1 METIS

The first approach, called Metis [9], tries to collapse similar and neighbor vertices, reducing the initial graph to a smaller one (known as coarsening step). Then, it divides the coarsest graph into a pre-defined number of graphs, generating the clusters.

2.1.2 SPECTRAL

Spectral clustering [10] can also be seen as a graph partitioning method, which measures both the total dissimilarity between groups as well as the total similarity within a group. We used the Scikit-learn[1] implementation of this method.

2.1.3 HIERARCHICAL

Hierarchical clustering [12] is based on the idea of a hierarchy of clusters. A tree is built in a way that the root gathers all the samples and the leaves are clusters with only one sample. This tree can be built bottom-up or top-down. We used the bottom-up implementation from Scikit-learn.

To find diverse images we experiment with different clustering methods. From our experiments we learned that an approach based on ensemble of clusters can perform better than using only one single clustering method. We also learned that a pre-filtering step can potentially remove irrelevant images that harm the process of clustering creation. Here we briefly comment on these two aspects:

2.2 PRE-FILTERING

We use hand-coded rules previously shown to perform well in this task, to exclude probably irrelevant pictures [8]. We exclude pictures based on three rules: without any views, geo-tagged 8km away from the POI, or with description length greater than 2000 characters.

[1] `http://scikit-learn.org/`

2.3 CLUSTERING SOLUTION

The basic idea is that, given a clustering algorithm A, a feature set F that describes an image and a distance measure Di, we can create a cluster set $C = (A, F, Di)$. For example, C_1 can be the result of applying K-Means (A) using the Color Histogram of the images (F), based on the cosine distance (Di): $C_1 =$ (K-Means, ColorHistogram, Cosine).

A common strategy used by a number of teams in the 2013 MediaEval Retrieving Diverse Social Images task was to go one by one of the clusters made in C_1 and pick the "best" image from each cluster to form the final ranked list. We noticed that small differences, for example having $C_2 =$ (K-means, NeuralNetworkFeatures, Cosine), could have a large impact in the clusters formed, consequently strongly influencing the final ranked list. As described below, our solution is to use the development set to learn what are the best clustering algorithm, features sets, and distance measures. After that, we combine the results of different Cs and count the frequency that any two images end up in the same cluster.

Alternative to traditional clustering methods, we developed a new method that clusters the data into sub-categories while eliminating the outliers. The details are described in Deliverable 4.2 and could be found in Golge and Duygulu [?]. In summary, the method revisits the well known Self Organizing Maps idea and extends it in order to clean the data to eliminate outliers. There could be some clusters coherent in themselves but different than the rest of the data, and there could be outlier elements inside the relevant clusters. The proposed methods handles both in order to capture the representative characteristics of a category, while at the same time organizing the data into sub-categories. These sub-clusters pruned from outliers are the representatives of the collection. In order to provide a diverse set of representatives we could look at the similarities of the clusters, and choose the ones which are the farthest from each other, which is left as a future work.

3 TEXT DIVERSIFICATION

3.1 TEXT PROCESSING MODULE

The text processing module is used to process on one hand, the images associated metadata and, on the other hand, the user queries. For the text processing tasks, standard tools are used for POS-tagging [11], lemma identification [11] and named entity identification [4]. After the images associated metadata is processed, the image collection is indexed with Lucene[2]. In order to achieve diversification in the results set, the system incorporates a query expansion module that makes use of the Yago[3] ontology [6].

[2] Lucene: http://lucene.apache.org/

[3] Yago: https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago/

Yago ontology comprises well known knowledge about the world (Hoffart et al, 2013). It contains information extracted from Wikipedia[4] and other sources like WordNet[5] and GeoNames[6] and it is structured in elements called entities (*persons*, *cities*, etc.) and facts about these entities (which *person* worked in which *domain*, etc.). For example, with Yago we are able to replace in a query like "*tennis player on court*", the entity "*tennis player*" with instances like "*Roger Federer*", "*Rafael Nadal*", etc. Thus, instead of performing a single search with the initial query, we perform several searches with the new queries, and in the end we combine the obtained partial results in a final result set. Because of its structure, YAGO will be used only when the text queries will match WordNet concepts that are linked by a hypernymy relationship to other Wikipedia entities, such as, person, location or organization.

Wikipedia: to decide when to use Yago, we created a resource based on hierarchies of Wikipedia categories. For this, we started with Romanian Wikipedia which has 8 groups of categories: culture, geography, history, mathematics, society, science, technology, privacy. In turn, these categories have subcategories or links to pages directly, as follows: Culture (30) (among which we mention *photo, architecture, art, sports, tourism*, etc.) Geography (15) (among which mention *Romania, Africa, Europe Countries, maps,* etc.), History (6) (among which mention *After the recall, By region,* etc.), Mathematics (11) (among which mention *Algebra, Arithmetic, Economics, Geometry, Logic,* etc.), Company (22) (among which mention *Anthropology, Archaeology, Business, Communications, Philosophy , Politics,* etc.), Science (23) (among which mention *Anthropology, Archaeology, Astronomy, Biology,* etc.), Technology (19) (among which mention *Agriculture, Architecture, Biotechnology, Computer,* etc.), Private life (8) (among which mention the *Fireplace, Fun, People, Health,* etc.). In the end, we obtained 8 big groups with 134 categories, which are subdivided into several subcategories and pages (hierarchical depth depends on each category and subcategory). In general, this hierarchy covers most of the concepts available for Romanian. For example, for Sport, we obtained 70 subcategories containing other subcategories and 9 pages. Going through these categories and subcategories, we built specific resources with words that signal concepts of type *person*, *location* and *organization*.

Some examples of signal words from these categories are:

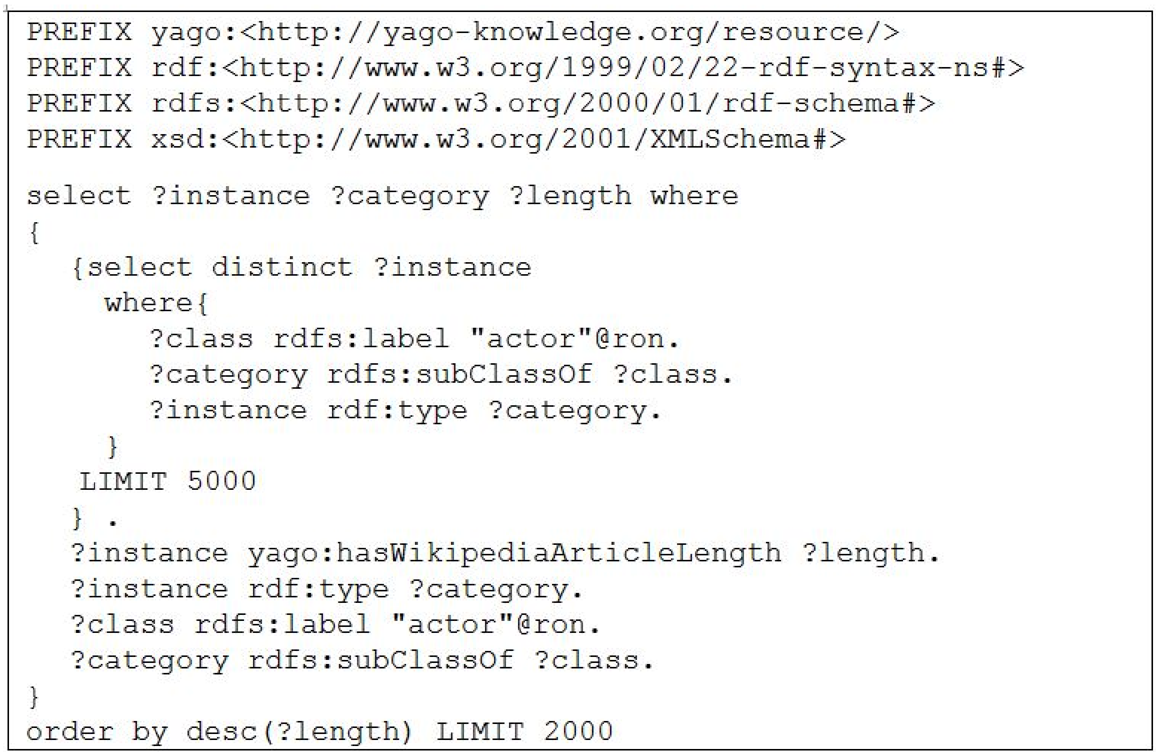
- For Person: "*acordeonist, actor, inginer, antropologist, arheolog, arhitect, femeie, arhivist, asasin, astronaut, astronom, astrofizician*, etc." (En: accordionist, actor, engineer, anthropologist, archaeologist, architect, woman, archivist, assassin, astronaut, astronomer, astrophysicist.) This is the biggest resource with over 391 signal words.

- For Location: "*continent, țară, oraș, comună, sat, regiune, munte, râu, fluviu, piață, stradă, bulevard, târg, instituție, universitate, spital, teatru*, etc." (En: continent, country, city, township, village, region, mountain, river, market, street, avenue, fair, institution, University, hos-

[4] Wikipedia: http://en.wikipedia.org/

[5] WordNet: http://wordnet.princeton.edu/

[6] GeoNames: http://www.geonames.org/



```
PREFIX yago:<http://yago-knowledge.org/resource/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

select ?instance ?category ?length where
{
   {select distinct ?instance
     where{
        ?class rdfs:label "actor"@ron.
        ?category rdfs:subClassOf ?class.
        ?instance rdf:type ?category.
     }
    LIMIT 5000
   } .
   ?instance yago:hasWikipediaArticleLength ?length.
   ?instance rdf:type ?category.
   ?class rdfs:label "actor"@ron.
   ?category rdfs:subClassOf ?class.
}
order by desc(?length) LIMIT 2000
```

FIGURE 1: SPARQL QUERY TO RETRIEVE *ACTOR* ENTITIES

pital, theatre).

- For Organization: "*companie, SRL, partid, grupare,* etc." (En: company, LLC, party, group).

Examples:

1. starting from a query that includes the word *actor* (En: actor), it decides to use YAGO because our system identifies this word in the list with signal words for type *person*, and it calls a Sparql query with the following form: (see Figure 1):

   The results retrieved by YAGO are ordered by their article length and include entities like: *Ronald Reagan, Jennifer Lopez, Elvis Presley, Madonna, Hulk Hogan, Clint Eastwood, Linda Ronstadt, Steven Spielberg, Orson Welles, Britney Spears, Eminem, Paul Robeson, John Cena, Lindsay Lohan, Cher*, etc. It is noted that not all entities are of type actor (for example: *Steven Spielberg*), but most are. After performing a search on Google with the word *actor* (En: *actor*) we obtain the results from Figure 2. After performing the same search in our application with the word *actor* (En: *actor*), we obtain the results from Figure 3.

2. starting from a query that includes the word *companie* (En: *company*), it decides to use YAGO because our system identifies this word in the list with signal words for type organization, and it calls a Sparql query with the following form: (see Figure 4) The results retrieved by YAGO include entities like: *Cirque du Soleil, English National Opera, Théâtre Lyrique, American Ballet Theatre, The Royal Ballet, New York City Opera, Tulsa Ballet, San Francisco Opera, Pacific Northwest Ballet, The Second City,* etc. It is noted that the majority are operas

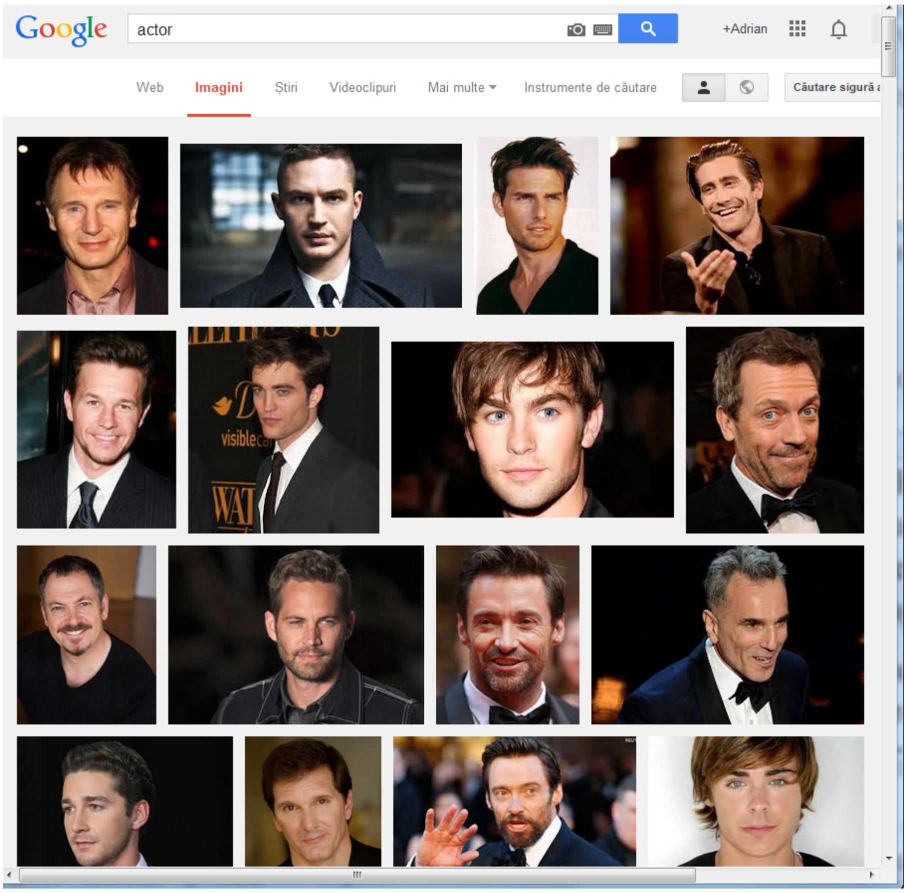


FIGURE 2: RESULTS OFFERED BY GOOGLE IMAGE SEARCH FOR THE QUERY *ACTOR*

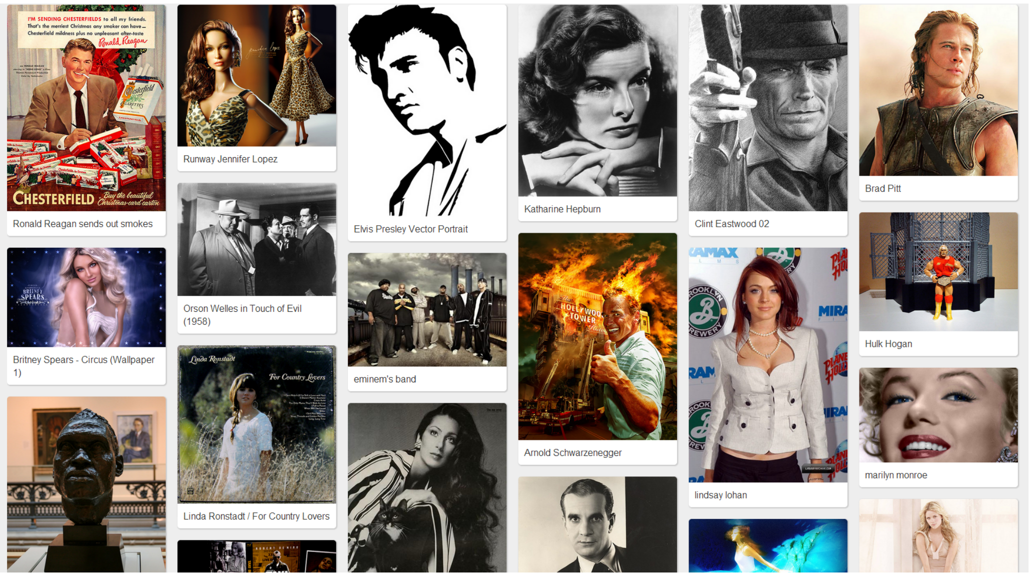
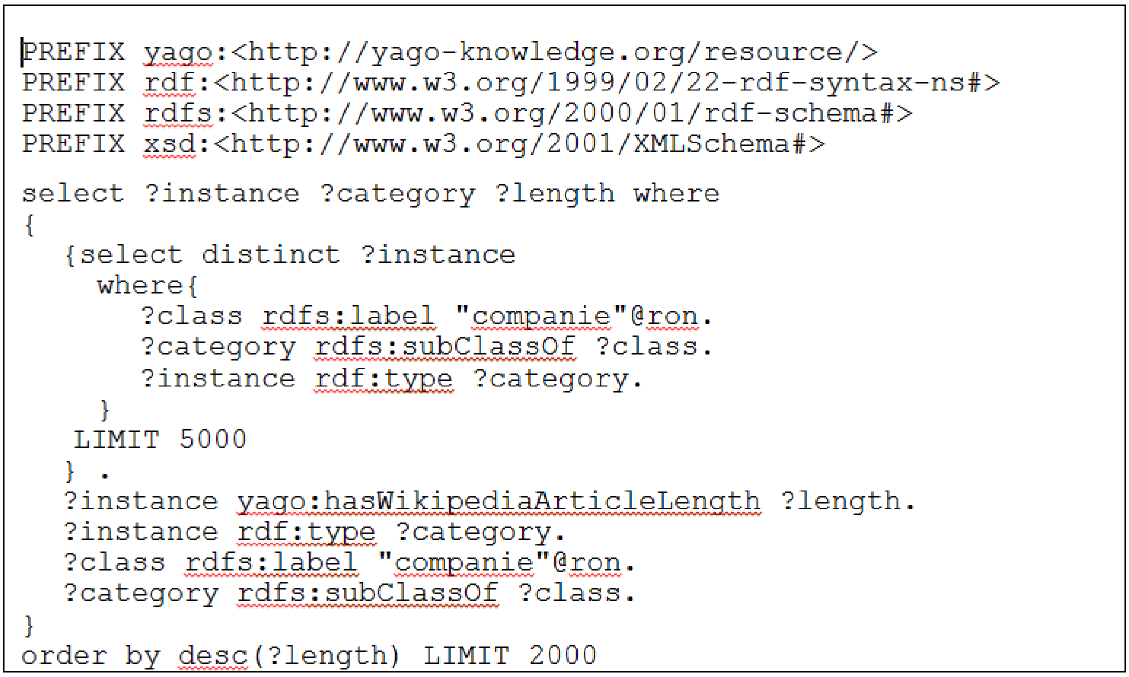

Ronald Reagan sends out smokes
Britney Spears - Circus (Wallpaper 1)
Runway Jennifer Lopez
Orson Welles in Touch of Evil (1958)
Linda Ronstadt / For Country Lovers
Elvis Presley Vector Portrait
eminem's band
Katharine Hepburn
Arnold Schwarzenegger
Clint Eastwood 02
lindsay lohan
Brad Pitt
Hulk Hogan
marilyn monroe

FIGURE 3: RESULTS OFFERED BY OUR APPLICATION FOR THE QUERY *ACTOR*



```
PREFIX yago:<http://yago-knowledge.org/resource/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

select ?instance ?category ?length where
{
  {select distinct ?instance
    where{
      ?class rdfs:label "companie"@ron.
      ?category rdfs:subClassOf ?class.
      ?instance rdf:type ?category.
    }
   LIMIT 5000
  } .
  ?instance yago:hasWikipediaArticleLength ?length.
  ?instance rdf:type ?category.
  ?class rdfs:label "companie"@ron.
  ?category rdfs:subClassOf ?class.
}
order by desc(?length) LIMIT 2000
```

FIGURE 4: SPARQL QUERY TO RETRIEVE *COMPANIE* ENTITIES

and ballet companies. After performing a search on Google with the word *companie* (En: *company*) we obtain the results from Figure 5. After performing the same search in our application with the word *companie* (En: *company*), we obtain the results from Figure 6.

3. starting from a query that includes the word *munte* (En: mountain), it decides to use YAGO because our system identifies this word in the list with signal words for type *location*, and it calls a Sparql query with the following form: (see Figure 7)

   The results retrieved by YAGO include entities, such as, *Rogue River (Oregon), Aliso Creek (Orange County), Ore Mountain passes, Santa Ana River, Matterhorn, Klamath River, Loi'hi Seamount, Mount St. Helens, Mount Pinatubo, Mount Edziza volcanic complex, Mount Garibaldi, Metacomet Ridge, San Juan Creek, Mount Rainier, Mount Baker*, etc. It is noted that many of the entities are of type creek or river, but in this case this kind of entities can be easily eliminated with simple rules from our list.

In all three cases results offered by Google are similar from the point of view of concepts presented in images returned. In the case of our application are more "colours" and more concepts in comparison with results offered by Google.

Query reformulation module provides a technique of processing a given query by obtaining new concepts that are both efficient and relevant in the context of information retrieval operations. This module is very similar to the module responsible with question analysis in a question answering system [7]. In this case, we face two major issues that occur when an end user entered a query: *it is not precise enough*, meaning that there are too many results returned, most of them being irrelevant or *it is not abstract enough*, meaning that the search does not return any results at all. Here, we

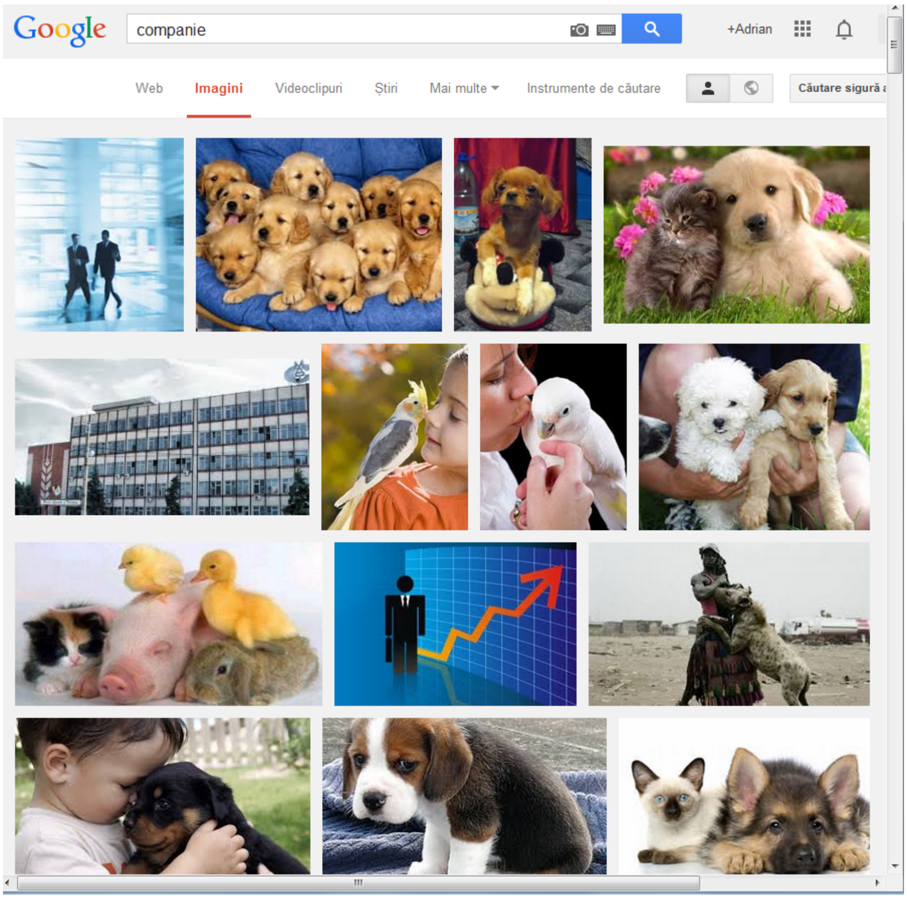


FIGURE 5: RESULTS OFFERED BY GOOGLE IMAGE SEARCH FOR THE QUERY *COMPANIE*

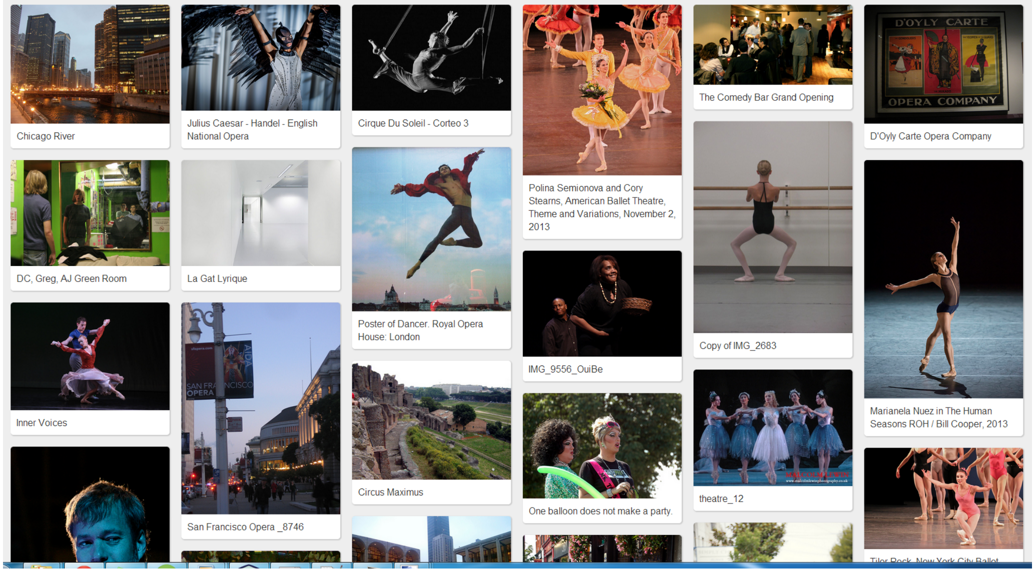
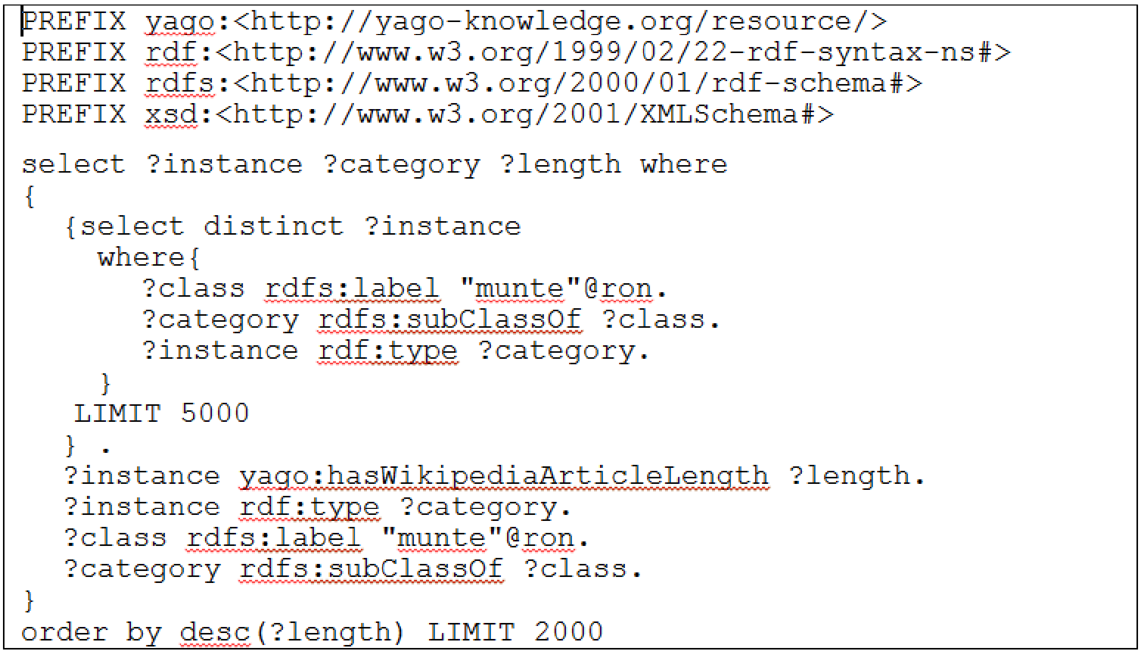

Chicago River
Julius Caesar - Handel - English National Opera
Cirque Du Soleil - Corteo 3
Polina Semionova and Cory Stearns, American Ballet Theatre, Theme and Variations, November 2, 2013
The Comedy Bar Grand Opening
D'OYLY CARTE OPERA COMPANY
D'Oyly Carte Opera Company
DC, Greg, AJ Green Room
La Gat Lyrique
Poster of Dancer. Royal Opera House: London
IMG_9556_OuiBe
Copy of IMG_2683
Marianela Nuez in The Human Seasons ROH / Bill Cooper, 2013
Inner Voices
SAN FRANCISCO OPERA
San Francisco Opera _8746
Circus Maximus
One balloon does not make a party.
theatre_12
Tiler Peck, New York City Ballet

FIGURE 6: RESULTS OFFERED BY OUR APPLICATION FOR THE QUERY *COMPANIE*



```
PREFIX yago:<http://yago-knowledge.org/resource/>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>

select ?instance ?category ?length where
{
  {select distinct ?instance
    where{
       ?class rdfs:label "munte"@ron.
       ?category rdfs:subClassOf ?class.
       ?instance rdf:type ?category.
    }
   LIMIT 5000
  } .
  ?instance yago:hasWikipediaArticleLength ?length.
  ?instance rdf:type ?category.
  ?class rdfs:label "munte"@ron.
  ?category rdfs:subClassOf ?class.
}
order by desc(?length) LIMIT 2000
```

FIGURE 7: SPARQL QUERY TO RETRIEVE *MUNTE* ENTITIES

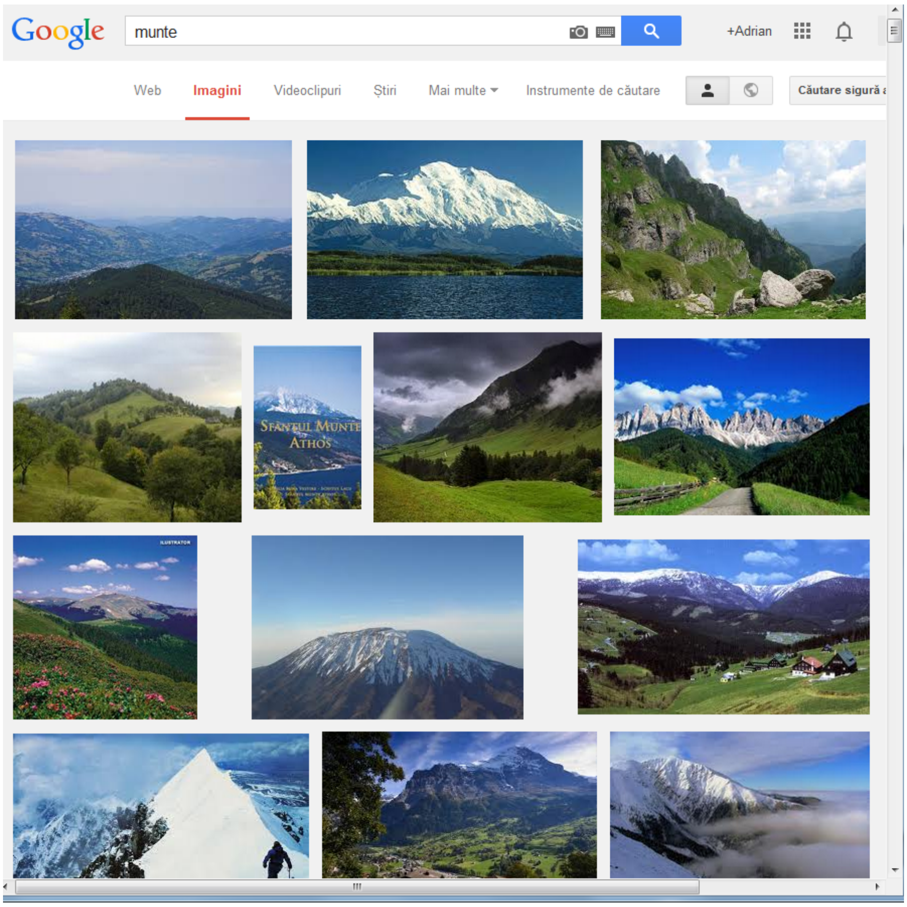


FIGURE 8: RESULTS OFFERED BY GOOGLE IMAGE SEARCH FOR THE QUERY *MUNTE*

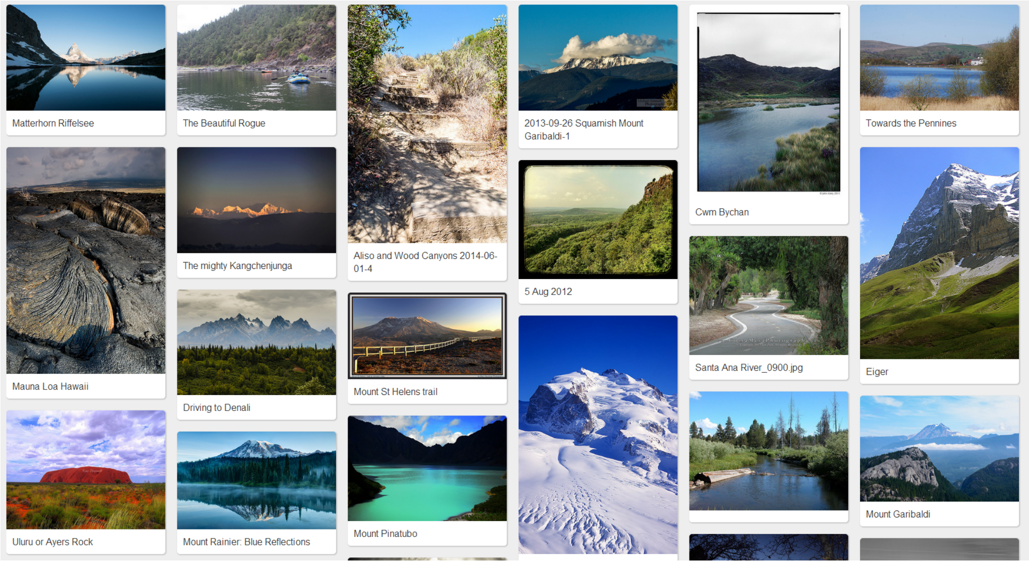


FIGURE 9: RESULTS OFFERED BY OUR APPLICATION FOR THE QUERY *MUNTE*

apply two approaches: (1) *a global technique*, which analyses the body of the query in order to discover word relationships (synonyms, homonyms or other morphological forms from WordNet), to remove stop words ("a", "un", "la", "pentru", (English: the, a, at, for), etc.), to remove wh-words ("cine", "ce", "de ce", "unde", (English: who, what, why, where), etc.) and to correct any spelling errors; (2) *local feedback* which implies the analysis of the results returned by the initial query, leading to re-weighting the terms of the query and relating it with entities and relationships originating from the target ontology.

4 CONCLUSION

We perform several text processing tasks on user queries and we identify signal words related to entities of type *person, location or organization*. If such words are identified, we use Yago to expand the user query and we perform several searches with the new queries in our image collection. Finally, we perform image processing tasks in order to create clusters of similar images. From what we have seen so far, the results are promising, and as future work we want to develop a module that allows us to evaluate the created system.

References

[1] Jaime Carbonell and Jade Goldstein. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, pages 335–336, New York, NY, USA, 1998. ACM.

[2] C. L. A. Clarke, M. Kolla, G. V. Cormack, O. Vechtomova, A. Ashkan, S. Bttcher, and I. MacKinnon. Novelty and diversity in information retrieval evaluation. In *SIGIR (2008)*, pages 659–666, 2008.

[3] M. Drosou and A Pitoura. Search result diversification. In *SIGMOD (2010)*, pages 41–47, 2010.

[4] A. L. Ginsca, E. Boros, A. Iftene, D. Trandabat, M. Toader, M. Corici, C. A. Perez, and D. Cristea. Sentimatrix - multilingual sentiment analysis service. In *In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (ACL-WASSA2011), Portland, Oregon, USA*, pages 189–195, 2011.

[5] S. Gollapudi and A. Sharma. An axiomatic approach for result diversification. In *WWW*, pages 381–390, 2009.

[6] A. Iftene, A. Siriteanu, and M. Petic. How to do diversification in an image retrieval system. In *In Proceedings of the 10th International Conference "Linguistic Resources and Tools for Processing the Romanian Language", Craiova, 18-19 September 2014*, pages 153–162, 2014.

[7] A. Iftene, D. Trandabat, A. Moruz, I. Pistol, M. Husarciuc, and D. Cristea. Question answering on english and romanian languages. In *In C. Peters et al. (Eds.): CLEF 2009, LNCS 6241, Part I (Multilingual Information Access Evaluation Vol. I Text Retrieval Experiments), Springer, Heidelberg*, pages 229–236, 2010.

[8] Neha Jain, Jonathon Hare, Sina Samangooei, John Preston, Jamie Davies, David Dupplaw, and Paul H. Lewis. Experiments in diversifying flickr result sets. In *MediaEval 2013*, 2013.

[9] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM J. Sci. Comput.*, 20(1):359–392, December 1998.

[10] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, Aug 2000.

[11] R. Simionescu. Hybrid pos tagger. In *In Proceedings of Language Resources and Tools with Industrial Applications Workshop (Eurolan 2011 summerschool)*, pages 21–28, 2011.

[12] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

[13] W. Zheng, X. Wang, H. Fang, and H. Cheng. Coverage-based search result diversification. In *Journal IR (2012)*, pages 433–457, 2012.