Leveraging Knowledge from the Linked Open Data Cloud in the task of Reverse Geo-tagging

Victor Charpenay [1,2], **Előd Egyed-Zsigmond** [1], **Harald Kosch** [2]

1. *Université de Lyon*
   INSA Lyon, LIRIS, 20 Avenue Albert Einstein
   69621 Villeurbanne Cedex, France
   victor.charpenay@insalien.org,elod.egyed-zsigmond@insa-lyon.fr
2. *Universität Passau*
   Innstraße 41
   94032 Passau, Germany
   harald.kosch@uni-passau.de

ABSTRACT. *Currently, Image Reverse Geo-tagging use probabilistic algorithms that reliy on the keywords describing a picture to guess the localization of the depicted scene. However, such algorithms still perform poorly and show clear limitations. Notably, the location estimation only occurs at the landmark level; regions or countries are reduced to their centroid.*
In this paper, we address this particular issue by exploring a semantic approach, which identifies geographical entities among the keywords to localize the picture (being a landmark or a country). We leverage the Linked Open Data cloud to find possible entities. The benefits of our approach, as opposed to numerical approaches, include an in-depth study of the "geo-relevance" of an image.

RÉSUMÉ. *Actuellement, la géo-localisation d'une image consiste à appliquer des algorithmes probabilistes sur les mots-clés la décrivant pour estimer la position de la scène qu'elle représente. Cependant, de tels algorithmes montrent des limites clairement identifiables. En particulier, l'estimation se fait toujours à l'échelle d'un point, les régions et pays étant réduits à leur barycentre. Dans cet article, nous nous concentrons sur ce problème en explorant une méthode sémantique qui identifie des entités géographique (issues du Linked Open Data) pour localiser une photo (qu'il s'agisse d'un point sur une carte ou un pays). L'avantage d'une telle approche vis-à-vis des méthodes numériques est notamment la possibilité d'étudier la pertinence géographique d'une image.*

KEYWORDS: *information retrieval, Semantic Web, Linked Data, multimedia.*

MOTS-CLÉS : *recherche d'informations, Web sémantique, metadonnées, multimedia.*

1. Introduction

The number of photos available on the Web is constantly growing, due to our storage, sharing and editing activity on platforms like Flickr[1]. Gathering these photos has been made easy by Flickr, whose database of several millions of photos is publicly accessible via an API. However, processing them is a harder task. It has motivated more than two decades of research in the field of Multimedia Information Retrieval (MIR). Reverse Geo-tagging is one of the application fields of MIR.

Reverse geo-tagging consists in estimating the location of a picture without contextual information (such as surrounding text or a caption). For instance, it might be easy to recognize that an image was taken in the Sahara if it depicts red sand dunes and a camel.

The openness of Flickr API has facilitated the creation of large data sets, shared among researchers. MediaEval[2] is an annual benchmarking initiative where research teams compete against each other on the same data. Since 2010, MediaEval contains a placing task, addressing the problem of reverse geo-tagging. Since 2013, a placeability prediction sub-task was added after performances of the state-of-the-art repeatedly proved to be poor compared to human performances. It is often claimed that many photos are actually not placeable. However, there is another key problem, which is the following: in our Sahara picture, despite its discrimative attributes, it does not illustrate a precise location on Earth with fixed latitude/longitude but a region, which covers no less than eleven countries. A reverse geo-tagging application that only outputs coordinates is then likely to make an error of hundreds of kilometers (given that the Sahara is 1,800 km wide and 4,800 km long).

That is why we explore the use of semantics in this paper, in contrast to existing methods that are mostly numerical and probabilistic, to localize a collection of pictures. With the help of a knowledge base, we hope to be able to give an extended output in the form of a geographical object containing geo-coordinates along with other information, so as to increase reverse geo-tagging expressiveness and investigate further the notion of "placeabilty" –what we thereafter also call geo-relevance.

Next section summarizes the state-of-the-art in reverse geo-tagging and points out the main limitations that motivated us to take a new semantic-oriented direction. We called the resulting method "geo-semantification". Section 3 details our implementation of a geo-semantifier and section 4 gives the results of some experiments and observations around geo-semantification. The paper ends with an open conclusion about the added-value of semantics in reverse geo-tagging.

1. https://www.flickr.com/
2. http://multimediaeval.org/

2. Related Work

Our goal is to merge two very distinct research topics, namely reverse geo-tagging, as part of information retrieval, and knowledge representation and semantics. We review first the state-of-the-art of reverse geo-tagging. As it leads to the conclusion that the most effective method to localize image is to take advantage of the keyword distribution of a collection of human annotated images, we review next existing methods to capture knowledge from those keywords. In the context of semantics, such keyword structure is called "folksonomy", as it corresponds to an attempt to make a crowd agree on a common classification (or taxonomy) of a resource set in an unstructured fashion.

2.1. Reverse Geo-tagging

Research about geo-localizing images has been made possible by the trend of tagging photos with geo-coordinates, in the first place on Flickr. It first triggered a renewal in MIR, many research projects took the direction of Geo-based Multimedia Retrieval. Among them, one can find a precursory work of reverse geo-tagging (Crandall *et al.*, 2009). From their 35 million images data set crawled from Flickr, they were able to detect popular landmarks by clustering photos. The results of the geo-localization task were promising: for a given city, about 65% of the test photos were given the appropriate landmark among the city's ten principal landmarks (however, in contrast, reverse geo-tagging aims at localizing arbitrary images all around the globe without restriction).

2.1.1. Content-based methods

As geo-based multimedia retrieval gained interest, a recurring problem was observed: even though the number of geo-tagged photos is constantly growing, most of the photos still lack such data; would it be possible to automatically recover it? This problem was first addressed by researchers working in the field of computer vision, who proposed content-based methods. True reverse geo-tagging started with a project called IM2GPS (Hays, Efros, 2008). The authors exploited Scene Recognition descriptors (color histograms, GIST, tiny images...) to perform location estimation. For a training set of 6 million photos, they reached an accuracy of 25% in a range of about 750 km, which they compare to the size of a small country.

Parallel to IM2GPS, another content-based location estimator was developed, reusing a technique from classical information retrieval: Latent Semantic Analysis (LSA) (Cristani *et al.*, 2008). The article mentions that 40% of their test images was located correctly. However, such results are hardly comparable to others since their data set only contains around 3,000 photos of a finite number of places in south-eastern France. Other works centered on geo-localization with prior assumptions include recognizing famous landmarks on Earth (Li *et al.*, 2009) or a city among a list (Fang *et al.*, 2013).

However, location estimation from the content only quickly proved limited. Crandall *et al.* stated that "at a landmark scale (100 m)," visual content is "very effective in estimating location" while "at a metropolitan scale (100 km), text tags are (...) highly effective for estimating location, but the image features are no longer useful" (Crandall *et al.*, 2009). Overall, if one uses keywords (or tags) associated to an image instead of its content, results are significantly improved. On Flickr, most of the photos are tagged (especially because photographers's visibility is increased this way). Therefore, tag-based methods currently dominate research around reverse geo-tagging.

2.1.2. Tag-based methods

The first tag-based proposal was made by (Serdyukov *et al.*, 2009). Since then, their method was re-used and extended several times. The underlying framework of their approach is called Language Model. They splitted the world in small regions and set up a probability distribution –the "language" that is spoken– in each region. Then, for an input image, its tag set is compared to that of every region. By integrating geographical specificities in the process, the authors achieved a precision of 29.6% with 100×100 km regions and a precision of 47% for regions three times bigger. The last result is almost two times better than what IM2GPS showed, with region dimensions that are two times smaller.

During the five years following (Serdyukov *et al.*, 2009), the language model has been added pre-processing steps (Joshi *et al.*, 2010) or post-processing refinement (Van Laere *et al.*, 2011; Trevisiol *et al.*, 2013). In an application paper, Joshi *et al.* propose a way to filter out all tags that may be irrelevant to set up the language model (Joshi *et al.*, 2010). To do so, conditional entropy from Information Theory is used: tags that are spread around the globe are not likely to give information about where the photo was taken (that is, their entropy is high with regards to the regions) and thus could be skipped.

Along with trying to improve the language model offline, it is also possible to refine the query online, as post-processing. After a region is identified as the most probable location according to the tag set of the input image, one can either find the closest image in terms of tags instead of computing a medoid (as in (Van Laere *et al.*, 2011)) or re-apply a more precise language model in the selected region (in a divide-and-conquer fashion, as in (Trevisiol *et al.*, 2013)). The latter achieves best results with an accuracy of 50% within 100 km. It has been hardly outperformed until now (without external resources) and can be seen as current state-of-the-art. It results from a participation in the MediaEval Placing Task 2012.

As a language model discards the fact that our resources are photos, it is also possible to apply it to tweets, which is getting more and more popular (Cheng *et al.*, 2010; Hauff, Houben, 2012; Roller *et al.*, 2012), or to short news (O'Hare, Murdock, 2012).

2.1.3. MediaEval 2013

MediaEval, as a benchamrking initiative, gives a good overview of the state-of-the-art in reverse geo-tagging and helps identify emerging trends in the field. First, the two best performing participants used a language model (Popescu, 2013; Cao, 2013). The third best participant used a slightly different approach but still based on tag probability distribution (Davies *et al.*, 2013). They all proposed diverse ad-hoc improvements to the process but the most interesting one is maybe the definition of the "geographicity" of a tag, so that photos whose tags have low geographicity are considered as unplaceable (Popescu, 2013). They could then be substituted with similar but placeable photos in the training set, if available. The "geographicity" of a tag equals its probability to appear around its most probable regions from the language model.

The results given by (Popescu, 2013) seem impressive: within 1 km, the estimation accuracy is 43% and within 100 km, it reaches 72%. These are the best results until now[3]. However, there are several observations showing that the proposed method has strong limitations. The most important one is the following: the introduction of a "geographicity" was not the main reason why it performed efficiently. It is mostly due to a larger training set. Starting from a size of 8.5 million images, they increased it to 90 million. Ten times more images were required for an improvement of about 17% at 100 km. Further increase would mean scalability issues. Furthermore, as the size of the training set gets larger, if the test set remains unchanged, there is an overfitting effect; the probability that very similar pictures exist between training set and test set rises, which weakens the role the language model.

What is more, the "geographicity" filtering, although it contributes to improve the overall accuracy, does not fit tags designating countries. It detects mostly landmarks and cities (an empirical radius of 15 km was used to characterize the proximity with a region). Previous attempts to define "geographicity" already existed in the literature. For instance, (Trevisiol *et al.*, 2013) defined a "geo-relevance" function, based on tag frequency and haversine distance, that suffers the same problem. Its parameters have to be hand-tuned so that it eliminates the right tags.

If we put apart results from (Popescu, 2013) for the reasons mentioned above, other methods perform too poorly to be used at a large scale. Only half of the photos could be localized correctly. The challenging issue is about finding a geo-relevance mechanism that could help determine a priori if a photo can be considered as placeable or not. As we saw, language model in itself cannot discriminate tags, hence the search for an alternative function. Our contribution to the field of reverse geo-tagging is the exploration of a new kind of geo-relevance function, based on semantic information instead of numerical information.

3. Results of MediaEval 2014 are not known at the time of writing.

2.2. Semantics in Folksonomies

Flickr and its millions of tagged photos is not only of interest in multimedia information retrieval (then seen as a database of enriched visual content) but also in the study of collaborative systems on the Internet (then seen as a social platform where users share knowledge). It makes available a large "folksonomy".

Folksonomies, expressing unstructured and consensual knowledge, may be opposed to structured taxonomies and ontologies. The latter are models considered as exhaustive that define relations between given concepts. They may cover a limited domain or intend to be high-level. A taxonomy only defines hierarchical relations and thus is less expressive than an ontology. From this perspective, knowledge can be seen as trees or graphs.

A very popular framework that embodies the idea of graph knowledge is W3C's Resource Description Framework (RDF), which is at the basis of the Semantic Web[4]. It has given birth to a huge graph of interconnected knowledge bases: the Linked Open Data (LOD) cloud[5]. Among others, the LOD cloud includes knowledge derived from the whole of Wikipedia in a semantic repository called DBpedia (Lehmann *et al.*, 2014).

If the tags attached to an image were associated to a concept defined in the LOD cloud, one could determine if a tag is geographically relevant or not, thanks to an analysis of its semantic neighborhood.

Nonetheless, Flickr and other folksonomies on the Web are not related to the Semantic Web (yet) and are not integrated into the LOD cloud. The need to perform automatically or semi-automatically a semantic association between tags and concepts emerged. It was addressed in the last few years by several research teams, reviewed in (García-Silva *et al.*, 2012). Most methods first disambiguate a tag thanks to lexical resources (e.g. WordNet[6]) or ad-hoc measures. Then, they use a semantic search engine (e.g. Swoogle[7]) or other semantic repositories to retrieve matching Semantic Web entities. The most interesting methods achieve a precision of about 90% in the test cases they published (Angeletou *et al.*, 2008; Tesconi *et al.*, 2008). However, their tests were not performed with tag sets as generic as needed in our current focus, as explained in section 4.

To finish, it is worth mentioning DBpedia Spotlight[8] as another tool to perform semantic association. Nonetheless, this tool was thought to be used for full text instead of tags only and, as such, it uses methods that slightly differ from those reviewed by García-Silva *et al.*

---

4. http://www.w3.org/standards/techs/rdf
5. http://linkeddata.org/
6. http://wordnet.princeton.edu/
7. http://swoogle.umbc.edu/
8. http://spotlight.dbpedia.org/

3. Implementation

With the idea of introducing semantics in reverse geo-tagging in mind, we previously identified that tags in a folksonomy could be "augmented" with a semantic context, referred to as semantic association. We believe that such a process could help define a more expressive geo-relevance function, which is a major issue in reverse geo-tagging.

We designed a system we called geo-semantification that contains the two following steps:

- Semantic association (as defined by García-Silva *et al.*)
- Geographic knowledge extraction (further discussed in section 3.2)

For the semantic association, we implemented several methods described in (García-Silva *et al.*, 2012) so as to find the best performing one in our context. Next, we evaluated knowledge bases available in the LOD cloud that provide geographic information and selected the most suitable one for our application.

3.1. Semantic Association

The review from García-Silva *et al.* lists eight possible semantic association candidates. Though, they do not all fit our specific criteria, which are the following: we are looking for a method that (1) is fully automatic, given that reverse geo-tagging deals with millions of pictures, (2) does not require any prior restriction about the foklsonomy (any topic, any lexical form. . .), (3) provides a Semantic Web entity as an output, with existing semantic neighborhood. Among those described, two methods involve the tagger or require human intervention, one method only handles verbs and two methods do not rely on fully-defined ontologies. Thus, we only re-implemented the three remaining methods presented by (Angeletou *et al.*, 2008), (García-Silva *et al.*, 2009) and (Tesconi *et al.*, 2008) and compared them.

Those three methods show similar procedures, as identified by (García-Silva *et al.*, 2012). Given an input tag, they first include a context identification phase, where words semantically related to the tag are gathered. Then, a second phase consists in a context-based disambiguation (only needed if the tag is found several meanings). At last, during a semantic identification phase, the meaning found previously is matched with an entity of the Semantic Web. As the input tag is associated to a resource (e.g. a photo) and a tagger, its context is either the remaining tags of the resource or all other tags used by the tagger on other resources.

Because each method has its specificities, we proceeded to some adaptations. First, as Flickr tags are especially noisy, tag cleaning is required. We defined a common tag filtering strategy that aims at combining their respective proposals. The following tags are removed as pre-processing: (1) words known as photography-specific concepts, like `nikon`, `blackwhite` or `iphonography` (list of 26 words); (2) tags contain-

ing digits, since they are most likely camera names (e.g. `d700`, `d80`) or technical details (e.g. `50mm`); (3) Flickr machine tags[9].

Moreover, as presented next, we used DBpedia as a knowledge base. Unfortunately, as certain resources used by the authors were not publicly available, we could not perfectly re-implement their work. Notably, (García-Silva *et al.*, 2009) and (Tesconi *et al.*, 2008) use an index of Wikipedia pages that takes into account disambiguation pages and redirect pages (TAGora Sense Repository and Tagpedia, respectively). Although both claim to make it available, none is still accessible on the Web and because we did not have enough resource to re-process all Wikipedia pages, we designed a simplified index. Instead of using the whole text of an article, our index detects disambiguation pages from the single title, thanks to Wikipedia guidelines for titles[10] (for instance, possibly ambiguous titles should contain additional information in parenthesis or after a comma). Of course, our index could not equal performances of original ones and introduces a certain amount of false positives. Redirect pages could not be processed at all.

So as to be able to compare results of the different methods, we discard Semantic Web entities that are not from DBpedia. As a consequence, for Angeletou *et al.*'s method, we did not use Watson semantic search engine[11] as they do, we query a DBpedia semantic repository directly on entities' label.

We also implemented two simple baselines as further comparison point. They provide naive disambiguation processes. The first one queries the Wikipedia index and randomly selects one of the results as the most probable meaning for the query tag. The purpose of this baseline was to ensure (or to put in doubt) that a further selection strategy is needed. The second baseline takes into account the context of a tag, that is, its co-occuring tags on one resource (a Flickr image in our case). For each result of the index query, it computes a weight using the TF-IDF formula. This baseline was used to discuss the impact of a context in the results.

3.2. Geographic Knowledge Extraction

Semantic assocation is not strongly tied to a specific semantic repository and theoretically applies to any LOD knowledge. However, we have to ensure that the concepts we retrieve contain geographic information, if any. For instance, if it designates a city, the most important information is its coordinates or related geograhic entities (e.g. landmarks in the city or the country to which it belongs). The relevant geographic knowledge is extracted by performing SPARQL queries on the knowledge base. There exists 31 geographic knowledge bases in the LOD cloud (out of 295) and 19.43% of the total number of atomic statements deal with geography[12]. We can men-

9. https://www.flickr.com/groups/api/discuss/72157594497877875
10. http://en.wikipedia.org/wiki/Wikipedia:Article_titles#Disambiguation
11. http://watson.kmi.open.ac.uk/WatsonWUI/
12. http://lod-cloud.net/state/

tion GeoNames[13], Linked GeoData[14] or Yahoo! GeoPlanet[15]. However none of them is exhaustive: GeoNames does not provide a hierarchy between geographical entities, Linked GeoData collection is incomplete for some regions and Yahoo! GeoPlanet does not contain GPS coordinates.

That is why we concentrated on generic knowledge bases, DBpedia being the biggest one. Such repositories, although their structure is sometimes heterogeneous, also provide geographical data. Table 1 illustrates that places (all geo-localized concepts) represent almost 20% of the whole DBpedia data. Places have geo-coordinates, they are typed according to a place ontology[16] (that distinguishes populated places from natural regions, for instance) and may contain additional information like area or related landmarks. Moreover, DBpedia states that it contains GPS coordinates for 1,094,000 concepts, which is more than the number of places. Indeed, some entities are not directly landmarks but are strongly related to a place. For example, one could think of a festival (an event, according to DBpedia classes) or a sculpture that is exposed in a square (which is classified as a work in DBpedia).

Table 1. The most frequent high-level classes in DBpedia and their frequency (source: http://wiki.dbpedia.org/Ontology39)

Class	Instances
Resource (overall)	$3,220,000$
Place	$639,000$
Person	$832,000$
Work	$372,000$
Species	$226,000$
Organization	$209,000$

All this made us base our geo-semantifier on DBpedia only. All concepts retrieved from the semantic association are defined within DBpedia. To extract geographical knowledge from a given concept, we simply check its type. If it is of type "Place" (defined in DBpedia's ontology), we further query its GPS properties, expressed as properties from the WGS 84 ontology[17].

4. Results

Previous section presented our two-steps geo-semantification process. In the remainder of this paper, we illustrate its benefits over numerical methods and discuss its shortcomings.

13. http://www.geonames.org/
14. http://linkedgeodata.org/About
15. https://developer.yahoo.com/geo/geoplanet/
16. http://mappings.dbpedia.org/server/ontology/classes/
17. http://www.w3.org/2003/01/geo/wgs84_pos

To perform several tests, we randomly selected 1,000 images from a database of 14 million elements, whose acquisition is described in (Mousselly-Sergieh *et al.*, 2014). All images have unique photographers. In the next section, we selected one tag from each photo (as the evaluation is per tag instead of per image). We took care that the 1,000 tags we then obtain are unique.

4.1. Semantic Association Bencharmking

As mentioned earlier, the survey about semantic assocation methods highlighted three suitable ones for our geo-semantifier. We carried out a quick benchmarking to determine which performs best. It is worth noting that we used a simplified Wikipedia index for two of the three methods, which may have worsen some results.

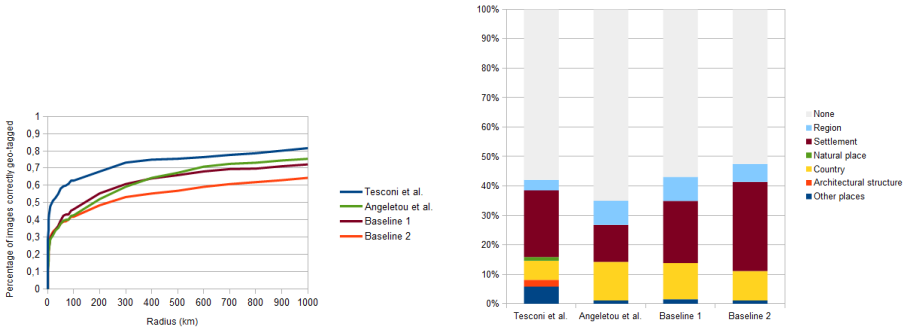
We found that the method proposed by (García-Silva *et al.*, 2009) performed poorly. It only reaches an accuracy of 26.0%, i.e. only one fourth of the tags were associated the right Semantic Web entity, if at least one was available (all tags were manually reviewed). As no result was published by the original authors, we decided to put this method aside and concentrate on the two others: (Angeletou *et al.*, 2008; Tesconi *et al.*, 2008). We found an accuracy of 62.2% and 57.7% respectively. At first sight, the method from Angeletou *et al.* performs better. However, it has a poorer coverage, i.e. less tags could be given a meaning since no concept was found during the disambiguation process. Their respective coverage is 66.0% and 79.0%. With such results, we decide to conduct experiments with both methods.

One could note that the accuracy results we give are far below those claimed by original authors (that both reach at least 90%). We argue that it is mostly due to very different tag sets. In the case of (Angeletou *et al.*, 2008), they find a coverage of 26.6%, which is very low. As their selection method is not clear, they may have applied much stricter criteria than we did to decide whether a tag can be "semantified" or not, leading to uncomparable results. About (Tesconi *et al.*, 2008), there are several observations that make us think that their test set was "easier" than ours. First, while we have 1,000 taggers for 1,000 tags, they only selected 9 users who contributed a total of 3,500 tags. Moreover, the average number of tags per resource is 3.38 in their case and 5.8 in ours. It let us think that the users they selected had well-behaved tagging habits, thus obtaining a clean test set, contrasting with Flickr recurring noise.

4.2. Geo-semantification

4.2.1. Location Estimation

We first evaluate the performances of the overall geo-semantification process. Here, we found that Tesconi *et al.*'s approach clearly outperforms Angeletou *et al.*'s one, as seen in Figure 1a. The reasons of this phenomenon are discussed later, using Figure 1b. Given such results, we used the method from Teconi *et al.* as the default semantic association method in our implementation for comparison purpose with other



(a) Evolution of the location estimation accuracy within a given radius

(b) Distribution of the types of the selected places. 'None' means that no geographical information was found in the photo

Figure 1. Quantitative evaluation of our geo-semantifier. Both diagrams show results for the two semantic association methods benchmarked in previous section, along with results for the two baselines described in section 3.1

reverse geo-tagging methods. We kept Angeletou *et al.*'s method and our baselines to analyze newly introduced measures.

If we compare the best results achieved by our geo-semantifier (i.e. the blue curve in Figure 1a) with (Popescu, 2013), the latter still performs better: within 1 km, with Tesconi et al.'s method, we get an accuracy of 20.1% (while Popescu get 43%) and within 100 km, we obtain 62.9% (72% for Popescu). However, our method outperforms other results of the state-of-the-art, which reach around 50% within 100 km, as in (Trevisiol *et al.*, 2013). Moreover, if we look further the results, we observe that the gap between our method and Popescu's shrinks as error radius grows. Within 1,000 km, accuracies are almost equivalent (81.6% and 83%, respectively). If we refer to a scale introduced by Hays and Efros in IM2GPS that states that cities have an average radius of 25 km, regions of 200 km and countries of 750km, it confirms that the strength of geo-semantification lies in managing localization at the scale of a country as well as at a landmark scale. We already identified previously that Popescu's geographicity measure can hardly detect countries.

What is more, geo-semantification provides information that no numerical reverse geo-tagger could give. For instance, if we take advantage of the area data that comes along with place concepts in DBpedia, we are able to define a "geo-semantic accuracy". It gives a single value that computes the ratio of test images that were localized within a radius defined by their associated DBpedia concept: given a location estimation x of a photo located at x_p, given the area A of the place concept used to estimate the location, the estimation is considered as geo-semantically accurate if $dist(x, x_p) \leq \sqrt{\frac{A}{\pi}}$ (that is, we compare the euclidian distance between the true lo-

cation and its estimation with the radius of the shape of A, approximated as a circle). Errors due to the spherical shape of the Earth are ignored.

The geo-semantic accuracy for our implementation is given in table 2. We can see that, regardless of the scale (landmark or country), it always shows a poor accuracy. Moreover, compared to our two baselines, the performance gain of the non-trivial semantic association is poor (less than 10%). This observation confirms that the main improvements are to be found in the semantic association step and does not depend much on DBpedia. During our experiments, we figured out that one of the advantages of the method proposed by Tesconi *et al.* is that it is partially based on the "popularity" of a DBpedia concept (measured as the size of the corresponding Wikipedia page). For instance, there exists several places called Berlin in the world. However, most of the people ignore the existence of the 20 towns in the United States bearing the same name as the German capital. It is obvious that the popularity of a concept is highly correlated with the most probable meaning of a tag in a folksonomy; this aspect may be further investigated to improve semantic association.

Table 2. Number of location estimations within a radius defined by their associated place concept, used as an error interval. We called this measure geo-semantic accuracy

Method	Geo-sem. accuracy (%)
Tesconi *et al.*	66.3
Angeletou *et al.*	60.1
Baseline 1	54.1
Baseline 2	47.8

Along with the geo-semantic accuracy, it is possible to further analyze the results thanks to the DBpedia place ontology. As the latter formalizes several place classes, it is possible to visualize which kind of places help most localize a picture, as shown in Figure 1b. Since, the place types available in the DBpedia ontology are extremely diverse, we only show high-level classes ("Settlement" denotes things like cities, towns or villages). Here, we can put forward that the method of Angeletou *et al.* will more likely select countries and regions than settlements for localization, which leads to a lower overall accuracy, as stated before. The diagram also explains the appearance of the different curves in Figure 1a. For instance, the distance from which Angeletou *et al.* gets better than baseline 1 is roughly 300 km, which corresponds to an intermediary size between a region and a country. Before that point, the baseline detects more precise concepts while afterwards, it accumulates semantic association error (the detected places are wrong). Similar observation could be made to compare the two baselines with each other. Moreover, it is interesting to see that the method of Tesconi *et al.* favors a larger variety of places. This feature could have never been observed with classic reverse geo-tagging systems (we cannot tell if the selected places are correct though).

Figure 1b shows another important information: more than half of the photos cannot be localized in any case. It brings us back to the question of geo-relevance of a photo and placeability prediction, as formulated in MediaEval 2013.

4.2.2. Placeability Classification

Our assumption is that a photo can be considered as not geo-relevant if it was not tagged with geo-located concepts. To verify it, we made a simple photo classification: if a photo contains geo-located tags, it has the class "placeable", otherwise it has the class "not placeable". Figure 2 shows a sample for each category.

Each sample illustrates the typical kind of image for each class. Placeable photos likely depict precise landmarks that are representative of the place where they were taken. Here, we have six photos representing buildings, whose name was input as a tag by the photographer, like the Bonn Minster (Germany), the oldest church of Malaysia, in Malacca or the city of Marseille (France), embodied by the CMA-CGM tower (the second highest French tower outside Paris). The two first photos (up left) do not represent landmarks. Though, as they symbolize events at known places, they are indirectly placeable. The first one is centered on a poster advertizing (in French) for an exhibition at the city hall of Paris. The second one is an image of a concert that took place at the Metropool concert hall, in Hengelo (Netherlands). The two next photos in the first row are exceptions, though. It is hard to guess that they were taken respectively in San Gabriel mountains (USA) and during the carnival of Barcelona (Spain).

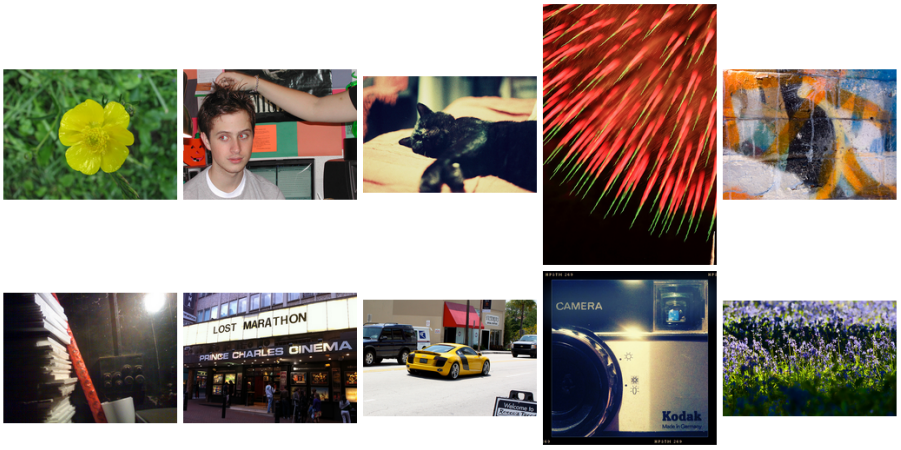
In contrast, not placeable photos, in the sample, are all close-ups or medium shots. They focus on an object (a flower, paintings on a wall, a car, books. . .) or people (here having unclear gestures since the frame is too narrow). They could have been taken anywhere –in the western world– and are geographically irrelevant.

The samples we presented here mostly echo our assumption. However, if one browses further all classified photos, one can notice that the class of placeable photos contains a certain number of false positives. Our geo-semantifier can localize some images even though their location is not clear or even impossible to guess for humans. It means that the photographers tagged them with places even if it was not the main purpose of the picture. That is, the presence of a geo-localized concept in the tags associated to the image is not a sufficient criterion to classify images as placeable. As a future work, we propose to investigate other possible criteria to improve the classification, such as distance measurement in the knowledge graph or machine learning on topological features, derived from the semantic neighborhood of the concepts that tag the photos.

5. Conclusion

In this paper, as we reviewed and pushed forward the use of semantics and the Linked Open Data in the task of reverse geo-tagging, we revealed several elements contributing to a future work.

(a) Placeable photos



(b) Not placeable photos

Figure 2. Samples of placeable and not placeable photos according to our geo-semantifier. The majority of photos in 2a are buildings, i.e. landmarks, whose name appears in their tags. In 2b, all the photos are close-ups or medium shots focusing on a given object

First, after we selected the best performing approach for semantic assocation in our study (from Tesconi *et al.*), geo-semantification was able to outperform the state-of-the-art, reaching a location estimation accuracy of about 63%. Along with competing with classical reverse geo-tagging, geo-semantification offers new possibilities in terms of geographical relevance. Among others, we defined a geo-semantic accuracy thanks to the knowledge obtained during the process of geo-localizing the image.

Moreover, we began to explore the relationship between image tags and the geo-relevance of an image. Although our classifier was trivial, we were already able to

conclude that if photos that do not contain geographic tags are likely *not* geo-relevant, photos with such tags are not always of interest for localization; Flickr users often tag photos with a city or a country even though it it is not related to its content. It made us think of a machine learning problem to build more evolved classifiers.

To finish, the location estimation accuracy we reached was lower than expected, partly due to the fact that the methods we surveyed had never been tested in a real-world application. We are convinced that it is worth investigating further tagging habits in folksonomies, especially the correlation between the probability of a given sense to match a tag and its popularity on Wikipedia.

Acknowledgements

This work was conducted within the frame of a double-degree between the French engineering school INSA and the university of Passau. It was grant-aided by the French region Rhône-Alpes and the Franco-German university (UFA-DFH).

References

Angeletou S., Sabou M., Motta E. (2008, November). *Semantically enriching folksonomies with FLOR.* Retrieved from http://oro.open.ac.uk/23497/1/10.1.1.142.7787[1].pdf

Cao J. (2013). Photo Set Refinement and Tag Segmentation in Georeferencing Flickr Photos. In *Working notes proceedings of the mediaeval 2013 multimedia benchmark workshop*, pp. 2–3. CEUR-WS.org.

Cheng Z., Caverlee J., Lee K. (2010, October). You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th acm international conference on information and knowledge management - cikm '10*, p. 759. New York, New York, USA, ACM Press.

Crandall D. J., Backstrom L., Huttenlocher D., Kleinberg J. (2009). Mapping the world's photos. *Proceedings of the 18th international conference on World wide web - WWW '09*, pp. 761.

Cristani M., Perina a., Castellani U., Murino V. (2008, June). Geo-located image analysis using latent representations. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.

Davies J., Preston J., Hare J., Jain N., Samangooei S., Dupplaw D. (2013). Identifying the Geographic Location of an Image with a Multimodal Probability Density Function. In *Working notes proceedings of the mediaeval 2013 multimedia benchmark workshop*, pp. 19–20. CEUR-WS.org.

Fang Q., Sang J., Xu C. (2013, October). GIANT: geo-informative attributes for location recognition and exploration. In *Proceedings of the 21st acm international conference on multimedia - mm '13*, pp. 13–22. New York, New York, USA, ACM Press.

García-Silva A., Corcho O., Alani H., Gómez-Pérez A. (2012). Review of the state of the art: Discovering and associating semantics to tags in folksonomies. *The Knowledge Engineering Review*, pp. 1–24.

García-Silva A., Szomszor M., Alani H., Corcho O. (2009, March). Preliminary Results in Tag Disambiguation using DBpedia. In *Proceedings of the fifth international conference on knowledge capture.* Facultad de Informática (UPM). Retrieved from http://oa.upm.es/6377/1/Preliminary_Results_in_Tag_D.pdf

Hauff C., Houben G.-J. (2012, August). Placing images on the world map: a microblog-based enrichment approach. In *Proceedings of the 35th international acm sigir conference on research and development in information retrieval - sigir '12*, p. 691. New York, New York, USA, ACM Press.

Hays J., Efros A. A. (2008, June). IM2GPS: estimating geographic information from a single image. In *2008 ieee conference on computer vision and pattern recognition*, Vol. 05, pp. 1–8. IEEE.

Joshi D., Gallagher A., Yu J., Luo J. (2010, July). Inferring photographic location using geotagged web images. *Multimedia Tools and Applications*, Vol. 56, No. 1, pp. 131–153.

Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., Mendes P. N. *et al. (2014). DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia.* Semantic Web Journal.

Li Y., Crandall D. J., Huttenlocher D. P. (2009, September). Landmark classification in large-scale image collections. In 2009 ieee 12th international conference on computer vision*, pp. 1957–1964. IEEE.*

Mousselly-Sergieh H., Döller M., Egyed-Zsigmond E., Watzinger D., Huber B., Kosch H. (2014, March). World-Wide Scale Geotagged Image Dataset for Automatic Image Annotation and Reverse Geotagging. In Multimedia systems conference 2014, mmsys '14. *Retrieved from http://liris.cnrs.fr/publis/?id=6425*

O'Hare N., Murdock V. (2012, April). Modeling locations with social media. Information Retrieval*, Vol. 16, No. 1, pp. 30–62.*

Popescu A. (2013). CEA LIST's Participation at MediaEval 2013 Placing Task. In Working notes proceedings of the mediaeval 2013 multimedia benchmark workshop*, pp. 1–2. CEUR-WS.org.*

Roller S., Speriosu M., Rallapalli S., Wing B., Baldridge J. (2012, July). Supervised text-based geolocation using language models on an adaptive grid. , pp. 1500–1510.

Serdyukov P., Murdock V., Zwol R. van. (2009, July). Placing flickr photos on a map. In Proceedings of the 32nd international acm sigir conference on research and development in information retrieval - sigir '09*, p. 484. New York, New York, USA, ACM Press.*

Tesconi M., Ronzano F., Marchetti A., Minutoli S. (2008). Semantify del.icio.us: automatically turn your tags into senses. In Social data on the web, workshop at the 7th international semantic web conference.

Trevisiol M., Jégou H., Delhumeau J., Gravier G. (2013, April). Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In Proceedings of the 3rd acm conference on international conference on multimedia retrieval - icmr '13*, p. 1. New York, New York, USA, ACM Press.*

Van Laere O., Schockaert S., Dhoedt B. (2011). Finding locations of flickr resources using language models and similarity search. Proceedings of the 1st ACM International Conference on Multimedia Retrieval - ICMR '11*, pp. 1–8.*