A Simple Tags Categorization Framework Using Spatial Coverage to Discover Geospatial Semantics

Camille Tardy
Computer Science Centre, University of Geneva
Battelle, 7 Route de Drize
CH-1227 Carouge, Switzerland
Camille.Tardy@unige.ch

Laurent Moccozet
Computer Science Centre, University of Geneva
Battelle, 7 Route de Drize
CH-1227 Carouge, Switzerland
Laurent.Moccozet@unige.ch

Gilles Falquet
Computer Science Centre, University of Geneva
Battelle, 7 Route de Drize
CH-1227 Carouge, Switzerland
Gilles.Falquet@unige.ch

ABSTRACT
There exist many popular crowdsourcing and social services (Volunteered Geographic Information (VGI)) to share information and documents such as Flickr, Foursquare, Twitter, Facebook, etc. They all use metadata, folksonomy and more importantly a geographic axis with GPS coordinates and/or geographic tags. Using this available folksonomy in VGI services we propose a logical approach to highlight and possibly discover the characteristics of geographic places. The approach is based on the notion of spatial coverage and a model of tags categorization and on their semantic identification, using semantic services such as GeoNames, OpenStreetMap or WordNet. We illustrate our model with Flickr to retrieve the characteristics (function, usage…) of places even if those places have a small number of related photos. Those found characteristics allow tag disambiguation and can be use to complete the semantic gap on places and POIs such as the function of buildings, which can exist in geographic services.

Keywords
Semantic; Folksonomy; Geographic Information; Coverage; Tag; VGI;

1. INTRODUCTION
The advent of social multimedia repositories such as Flickr or YouTube has raised a lot of interest for mining these sources in order to extract and discover knowledge for many purposes such as geographic information [8] [24]. These implicit VGI data sources offer promising opportunities to discover geospatial knowledge. Unfortunately the data cannot be directly exploited and need to be processed [20].

Our objectives here are to propose a simple logical framework, using existing Volunteered Geographic Information (VGI) sources, to find the characteristics of geographic places, and improve the formulation and the precision of search queries.

Most of the social repositories rely on folksonomies to annotate the multimedia contents. Folksonomies offer a very simple and attractive framework to annotate multimedia contents.

However, they are greatly lacking of semantics in order to exploit them appropriately. Various approaches can be applied in order to discover this semantics from many sources [1]. The analysis of folksonomies attached to social image repositories has been proposed to uncover geospatial knowledge [7, 14, 18, 19, 21].

It is important to differentiate spatially explicit vs. implicit (such as Flickr) social sharing sources [2]. With implicit sources, the distribution of geographic information is clustered in popular locations (related to entertainment and tourism). It suggests that traditional clustering and aggregation methods used to discover spatial semantics and knowledge are not appropriate in other places, where the density of images is low. In [4] Flickr tags are aggregated at multiple scales to study their spatial and thematic properties. The study confirms that for sources like Flickr, tags are clustered in touristic and entertainment areas. It also identifies a strong interaction between tag spatial semantics and the associated spatial scale.

A three place–related facets classification is proposed in [16] that includes: elements (objects and people that can be identified in photos), qualities (modify elements or suggest feelings) and activities. In the paper, annotators achieve the classification manually. A similar approach is adopted in [17], which suggests to use the Pansofsky-Shatford facet matrix to provide a formal universal image description model. The matrix is based on two aspects including facets. The first aspect is organized into the Who? What? Where and When? facets and the second one according to the Specific of, Generic of and About facets. Based on this formal framework, the authors show that it is possible to improve image descriptions (and therefore image annotation analysis) particularly with the proposed notion of place.

In this paper, we propose a multi-facets model for categorizing text content description such as tags for documents in general and images in particular. The model is based on a spatial coverage. We show how it can be applied to individually process images from Flickr in order to extract spatial knowledge using simple logical filtering rules with semantics resources such as GeoNames [6], OpenStreetMap [13] or WordNet [22].

2. CATEGORIZATION MODEL
Our model is firstly based on the distinction between the spatial coverage and the spatial references of a document. We define the spatial coverage for photos as a combination of semantic and geographic notions that describes precisely the saptio-temporal region that is shown in the picture. We make a distinction between the geographic and spatial content of a resource and its actual spatial coverage. The goal is to

differentiate the spatial content, which is composed of places that are not part of the coverage, and the actual spatial context of the resource.

The annotations of Flickr photos consist of freeform text tags created by the owner of the photo and if possible of a location tag and GPS coordinates. We propose to semantically identify each tag of the set and organize them according to the categories described below. The categories are used to identify the unnecessary information in non-geographic tags such as colors or weather indications. Those are not important information in the task of identifying the geographic places characteristics or function.

First we need to sort the tags in two sets the geographic and the non-geographic ones. The geographic tags are the geographic features and classes and their translation. Their identification is done using geographic services such as OpenStreetMap and GeoNames. The geographic features are the instances such as countries, cities, POIs, etc. The geographic classes are the main types categories of the geographic features, for example "tunnel", "city", "university", "lake", "forest", "restaurant", etc. [5, 9]. Those tags can be part of the picture coverage or spatial references.

For example this photo [https://flic.kr/p/m9ZBPB] (as shown in Fig. 1) of a concert of a Canadian artist at the Théâtre Pitoëff in Geneva is tagged with both "Geneva" and "Canada". Both are geographic tags but one is part of the spatial coverage of the picture and the other is a spatial reference. The concert took
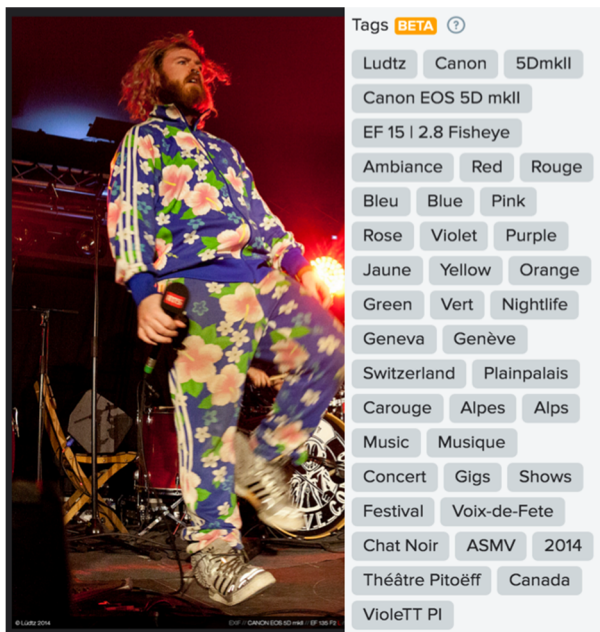

Figure 1. "VioleTT Pi" from Ludtz

place in Geneva so this is where the photo is true; "Geneva" is then part of the photo spatial coverage. The artist in the picture comes from Quebec, so the tag "Canada" is a spatial reference and part of the content of the photo.

Once every geographic tag has been identified the rest of the tags are all non-geographic tags. In order to filter the unwanted information, we have identified the following categories: Event, Temporal, Weather, Actors, Meta and Color. Those categories have been inspired by [16] and [17].

The **Event** category regroups all the tags that refer to events, or punctual things. In our example the tag "Voix-de-fête", as the name of a festival is part of the Event category.

The **Temporal** category regroups time related tags such as "afternoon", "night", "day", etc. It can be considered as a subcategory of the Event category.

The **Weather** category regroups weather tags such as "cloud", "sun", "rain", "storm", etc. Like the Temporal category the Weather one can also be considered a subcategory of Event.

The **Actor** category regroups people or in movement objects references that are not always at the place the photo was taken, such as "jogger", "tram", "car", "birds", etc.

The **Meta** category regroups tags that don't refer to content or coverage information: any photographic tags such as the photographer nickname, camera brand, camera model number and photo awards. In our example it would be: "Ludtz", "Canon", "5DmkII", "Canon EOS 5D mkII", "EF 15 | 2.8 Fisheye".

The **Color** category regroups all the tags that reference colors. They are often found in Flickr tags but don't refer to meaningful information in our use case.

The rest of the tags are the probable invariant characteristics that are recurrent in similar pictures and might give pertinent information on the places in the picture. For example the photo shown in Fig.1 have those tags left after categorization: "Ambiance", "Music", "Musique", "Nightlife". We can deduce from this, that the "Théâtre Pitoëff" geographic feature is related to music, nightlife and ambiance. We remark that the rest of the tags are those interesting in the context of finding the characteristics and functions of geographic features.

3. THE IMPLEMENTATION
3.1 Model implementation

As previously explained, our goal is to find place characteristics, such as building function or usage, and the spatial coverage of a picture. To do so, we need to identify the tags that represent geographic features (geo tags) and more particularly the coverage tags that represent features that appear in the picture. Secondly, we need to identify non-geographic tags that can hold information on the place characteristics.

The tag categorization algorithm uses external resources (Geonames, Openstreetmap, Wordnet) and the WOEID (Where On Earth Id) of the photo. Each Flickr photo has a WOEID that refers to a spatial entity of Yahoo! GeoPlanet [23]. The WOEID is determined by the GPS information contained in the EXIF file. Each WOEID belongs to a hierarchy of other geographic feature. The hierarchy goes from the country down to the neighborhood.

The algorithm is comprised of the following steps

1. Retrieve coverage geo tags in the WOEID area
2. Retrieve geo tags in the extended area
3. Retrieve geographic classes
4. Disambiguate and classify non-geo tags

(1) This step retrieves the tags that are most probably part of the spatial coverage of the picture. The identification is done by matching the label of the tag with the name or alternate names of the geographic features that lie within a bounding box. This step uses the GeoNames API. The bounding box is provided by the picture's location tag composed of the WOEID related to

the picture. This increases the rightness of the tags identification in case of homonyms. There is more probability that the term used to tag a picture is one referencing a closed by geographic feature. For example for a picture located in Europe, the tag "Paris" has greater probability to refer to the French capital than to the Texas town.

(2) The bounding box is extended to retrieve additional geo tags that are not geographically close to the picture location. The extension is done following WOEID hierarchy. This usually identifies features related to the content of the resource and not its coverage. However, some distant features may appear in the picture if they are tall or massive or wide, such as mountains, lakes, etc. A feature is considered as potentially part of the coverage if the ratio *feature size*/*distance* is greater than a threshold (typically 0.04). Since the feature sizes are generally not known, the algorithm uses a rough estimate that depends on the feature class (e.g. 3000m for a mountain, 20m for a building)

(3) This step retrieves tags that correspond to classes of geographic features (mountain, city, building, *etc.*). The algorithm tries to find matches between the tags and the classes and their closed hyponyms and meronyms. For example, the main class code S "*spot, building, farm*" is matched with the *facility* entity in WordNet. And a more precise example: the SCH (school) GeoNames class is matched with *school, schoolhouse (a building where young people receive education)* synset form WordNet.

(4) A word sense disambiguation (WSD) algorithm, such as [15] or [3, 10, 11] is applied to disambiguate the remaining tags. The system uses Babelfy [12] API as WSD tool. The algorithm receives as input the "sentence" formed by all the photo tags (even the previously identified geo tag). The output of the algorithm is that association of a WordNet synset (concept), or Wikipedia named entity, to every tag. This synset or named entity is then used to classify the tag in one of the categories described in Section 2 (Event, Temporal, Weather, Actors, Meta and Color) or in none of them.

Tags that have not been disambiguated by the WSD algorithm because, for example, there is no corresponding word in WordNet or Wikipedia and for which there exist a GeoNames entity are classified as geo tags. On the same level, if tags have been identified with a faraway geo entity during the first steps of the algorithm, and if they are also identified as non-geo entities by step 4, then the non-geo concept is kept as the tag disambiguation. This is induced by the fact that many common words are also names of geographic entities in the world. For example "Pub" is the name of a mountain in Pakistan, but it also means a bar, a tavern, in English.

3.2 Examples

3.2.1 Place de Sardaigne
The Table 1 regroups three photos of the same square in a city. The WOEID associated with all the photos is Carouge city. We can make the first assumption that the photos have all been taken in Carouge. After the identification process we know that "Carouge" is a geographic feature of type *city*, and "Genève" is a geographic feature of type *county* in the hierarchy of the "Carouge" WOEID. "Sardaigne" as a label is most commonly referencing the Italian island. But if we add to our query the bounding box of Carouge city, then "Sardaigne" becomes clearly the square or pedestrian way in Carouge. This identification is confirmed with the tag "Place" that is identified

as the geographic class SQR (square) in GeoNames. We can also affirm that Sardaigne (the square) is part of the spatial coverage of the photo.

We can also deduce from the tags in photo 1 that there might be a statue in the square, with the identification of "Statue" with the geographic class MNMT form GeoNames. The second photo non-geographic tags: "Bus" and "TPG", inform us on a bus line visible form the square. "TPG" is the name of the public transport in the canton of Geneva. They are identified as a pair in combination with the geographic constraint of "Geneva" using Wikipedia.

Table 1. "Sardaigne" disambiguation example

	photo url	geo tags	non-geo tags (category)
1	https://flic.kr/p/cGQS4d	Genève Carouge Sardaigne --*geo Class* -- Statue	Quiétude Nu
2	https://flic.kr/p/6JAMD1	Sardaigne Geneva Carouge --*geo Class*-- Place	~~de~~ Bus (Actor) TPG (Actor)
3	https://flic.kr/p/8YqFc	Genève Carouge Sardaigne	Nocturne (Temporal)

Figure 2. "Place de Sardaigne (Carouge / Geneva)" from X.Com - photo 2

In Fig. 3 we can see the result of the disambiguation from Babelfy as cited in section 3.1 for the picture 3. We can see the wrong identification (in our context) of the "Sardaigne" tag. In this case the algorithm will select the geo tag identification previously done through GeoNames using the bounding box.

Figure 3. Babelfy disambiguation for photo 3

3.2.2 Café de la Gymnastique
The photo [https://flic.kr/p/qppuvM] shown in Fig. 4, has three tags identified as geographic features: "carouge", "Geneva" and "switzerland". Their identification is done as described in the

previous example. The tag "restaurant", is identified as a geographic class REST (restaurant) in GeoNames. The rest of the tags are identified as a group using WordNet and in pair with "restaurant": "bar", "café", "bistrot". We can deduce the function of the building in the photo as a restaurant and bar establishment.
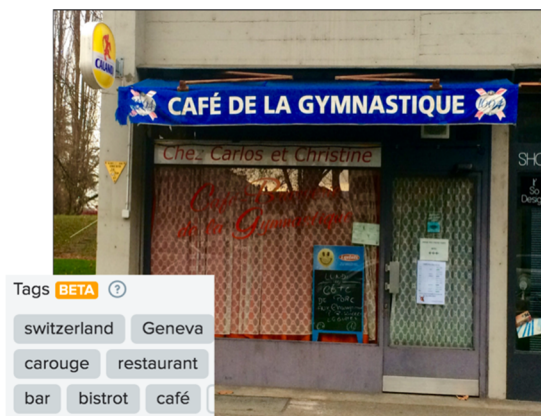

Figure 4. "Café de la Gymnastique" from Eric

4. CONCLUSION AND FUTURE WORK

We have developed a simple multi facet categorization framework based on spatial coverage. This framework helps to find the characteristics of a photo spatial definition. It is based on simple filtering rules to categorize and find the semantic of tags from geospatial resources open and available like OpenStreetMap, GeoNames, WordNet or Wikipedia.

One of the main interests of this framework is its usability on geographic zones where the density of photos is relatively low and so to be able to process local zones. Finding the function or usage of geographic places in those cases cannot be easily done using statistical approaches. The examples detail in section 3 show the model potential. We are currently implementing the algorithm presented in this paper in order to calculate the precision and recall.

This approach can also be considered a pre-treatment to bring forward a certain amount of information that can then be used in clusters. It could be used to enable the enhancement of spatial descriptions in geospatial services like OpenStreetMap or help define precise queries in search engines.

5. REFERENCES

[1]   Abbasi, R. 2011. *Discovering and Exploiting Semantics in Folksonomies*.

[2]   Antoniou, V. et al. 2010. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica*. 64, 1 (2010), 99–110.

[3]   Bunescu, R.C. and Pasca, M. 2006. Using Encyclopedic Knowledge for Named entity Disambiguation. *EACL*. (2006).

[4]   Feick, R. and Robertson, C. 2015. A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*. 53, (Sep. 2015), 96–109.

[5]   GeoNames Feature codes: *http://www.geonames.org/export/codes.html*. Accessed: 2015-12-28.

[6]   geonames.org: *http://www.geonames.org*. Accessed: 2012-06-16.

[7]   Li, L. and Goodchild, M.F. 2012. Constructing Places from Spatial Footprints. (New York, NY, USA, 2012), 15–21.

[8]   Liu, Z. 2011. A Survey on Social Image Mining. *Intelligent Computing and Information Science*. R. Chen, ed. Springer Berlin Heidelberg. 662–667.

[9]   Map Features - OpenStreetMap Wiki: *http://wiki.openstreetmap.org/wiki/Map_Features*. Accessed: 2015-12-28.

[10] Medelyan, O. et al. 2009. Mining meaning from Wikipedia. *International Journal of Human-Computer Studies*. 67, 9 (Sep. 2009), 716–754.

[11] Mihalcea, R. 2007. Using Wikipedia for Automatic Word Sense Disambiguation. *HLT-NAACL*. (2007).

[12] Moro, A. et al. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics (TACL)*. 2, (2014), 231–244.

[13] OpenStreetMap: *https://www.openstreetmap.org/*. Accessed: 2015-12-28.

[14] Papadopoulou, C.-A. and Giaoutzi, M. 2014. Crowdsourcing as a Tool for Knowledge Acquisition in Spatial Planning. *Future Internet*. 6, 1 (2014), 109–125.

[15] Pedersen, T. et al. 2004. WordNet:: Similarity: measuring the relatedness of concepts. *Demonstration papers at HLT-NAACL*. (2004).

[16] Purves, R. et al. 2011. Describing place through user generated content. *First Monday*. 16, 9 (Sep. 2011).

[17] Purves, R.S. et al. 2008. *Describing the where – improving image annotation and search through geography*. http://eprints.whiterose.ac.uk/4566/.

[18] Rattenbury, T. and Naaman, M. 2009. Methods for Extracting Place Semantics from Flickr Tags. *ACM Transactions on the Web*. 3, 1 (Jan. 2009), 1–30.

[19] Rattenbury, T. et al. 2007. Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. (New York, NY, USA, 2007), 103–110.

[20] Sester, M. et al. 2014. Integrating and Generalising Volunteered Geographic Information. *Abstracting Geographic Information in a Data Rich World*. D. Burghardt et al., eds. Springer International Publishing. 119–155.

[21] Sizov, S. 2012. Latent Geospatial Semantics of Social Media. *ACM Transactions on Intelligent Systems and Technology*. 3, 4 (2012), 64:1–64:20.

[22] wordnet.princeton.edu: *http://wordnet.princeton.edu*. Accessed: 2014-01-20.

[23] Yahoo! GeoPlanet™ - YDN: *https://developer.yahoo.com/geo/geoplanet/*. Accessed: 2015-12-28.

[24] Zheng, Y.-T. et al. 2010. Research and applications on georeferenced multimedia: a survey. *Multimedia Tools and Applications*. 51, 1 (2010), 77–98.