CHAPTER 11

A methodological toolbox for exploring collections of textually annotated georeferenced photographs

Ross S. Purves* and William A. Mackaness†

*Department of Geography, University of Zurich, Switzerland, ross.purves@geo.uzh.ch

†School of Geosciences, University of Edinburgh, UK

Abstract

This chapter provides a brief overview of some methodologies used to extract meaning from the analysis of geotagged images. Broadly they draw from research in natural language processing and statistical and exploratory techniques. The confidence we attach to outputs from such analysis depends upon the questions we ask, our ability to take account of both the behaviour and motivation of the users contributing to user generated content, and the close relationship between how the data are spatially aggregated and the meanings associated with descriptions of images.

Keywords

geo-tagged, images, user behaviour, spatial aggregation, spatial analysis, natural language processing

How to cite this book chapter:

Purves, R S and Mackaness, W A. 2016. A methodological toolbox for exploring collections of textually annotated georeferenced photographs. In: Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) *European Handbook of Crowdsourced Geographic Information*, Pp. 145–156. London: Ubiquity Press. DOI: http://dx.doi.org/10.5334/bax.k. License: CC-BY 4.0.

Introduction

We continue to witness phenomenal growth in the production of user generated content (UGC). Some of that content comes in the form of photographs. Many are either annotated or tagged in a manner that may reveal aspects of users' conceptual understanding of place. In this article we concern ourselves with methods to extract meaning from large collections of textually annotated georeferenced photographs. Such collections have been the subject of considerable attention over the last decade, for a number of reasons. Firstly, and perhaps most importantly, the data are accessible. For instance, both Flickr[1] and Panoramio[2] provide application programming interfaces which make it possible for researchers to scrape images and associated metadata, while Geograph[3] content is provided under a Creative Commons Attribution-ShareAlike licence. Secondly, unlike other social media, the link between position, annotation and content is often relatively direct and closely linked to people's sense of place. People take pictures of things and events that happen somewhere, at some time, and describe them accordingly.

A wide range of applications have been developed that variously utilise this data in order to extract information and meaning:

- Automatic generation of gazetteer data (Kessler et al. 2009)
- Extraction and delineation of vernacular place names (Hollenstein & Purves 2010)
- Tag recommendations for images based on location (Rattenbury & Naaman 2009)
- Adding information to existing spatial databases (Antoniou et al. 2010)
- Extraction of place semantics at a range of scales (Feick & Robertson 2014; Purves et al. 2011; Rattenbury & Naaman 2009)
- Summarising and aggregating properties of the semantics of space (Ahern et al. 2007; Dykes & Wood 2009; Purves et al. 2011)
- Exploring movement of groups of individuals in space (Girardin et al. 2008)
- Identification and prediction of locations in text (O'Hare & Murdock 2012)
- Extraction of events using space-time clustering (Andrienko et al. 2010)

All of these approaches require methods which go beyond analysing spatial patterns associated only with the locations of photographs. This is the province of an established toolbox of geostatistical techniques for point pattern analysis able to describe spatial distributions and multi-scale patterns (O'Sullivan & Unwin 2003: chap. 4). Additionally we may wish to infer place semantics from other metadata associated with images (e.g. user, annotation, and time as well

[1] www.flickr.com
[2] www.panoramio.com
[3] http://www.geograph.org.uk/

as location). In this article we are only concerned with annotations written by the user: the person who uploaded the photograph, and typically, but not always, took it. This user information allows association of a set of photographs with an, pseudo-anonymous, individual and annotations can take the form of a title, a narrative (often descriptive text), and a set of tags. Tags are lists of key words selected freely by a user (Rattenbury & Naaman 2009) and, like all annotation associated with photographs in user generated content, may have a number of different motivating factors, including organisation of content for personal reasons, providing informative descriptions and making photographs findable by others. Locations may reflect the scene photographed, but with the advent of smart phones capable of automatically annotating images with GPS coordinates, more commonly reflect the photographer's position. Finally, temporal information often reflects both time of upload to the database and the time at which a photograph was recorded by a camera as having been taken.

This set of properties allows us to formulate a set of basic questions which can be asked of a collection of annotated, georeferenced photographs:

1) What language is used to describe photographs?
2) How can structured knowledge be extracted from annotations?
3) What influence do users have on information extracted from annotated georeferenced photographs?
4) How can we capture the relationship between language and location?
5) How do descriptions extracted from annotations vary according to scale and region definitions?

In the following, we introduce a methodological toolbox, drawn from a representative set of literature working on georeferenced annotated images, which allows us to explore these questions. As argued above, our focus goes beyond purely spatial analysis, and in particular focuses on textual annotations. In fact, many of the methods applied come from the domains of statistical natural language processing and information retrieval and focus on extracting information from a corpus (Manning & Schütze 1999). Common to all corpora are the basic notions of documents (in our case represented by annotations related to an individual photograph). Information about authorship (in our case in the form of unique users) is somewhat less common, and explicit links to spatial locations are what make our collections of georeferenced photographs particularly interesting. Thus, in the following, we will firstly introduce some **global analysis methods** – and ignore potential stratifications of the data by user or location (Questions 1 & 2). We will then discuss the link between **user behaviour and language** (Question 3) before finally looking at the explicit link between **language, location and scale** (Questions 4 & 5).

In this paper we use as exemplary data two examples of UGC: Geograph and Flickr. Our analyses are based on previous work reported in Purves et al. (2011). We focus on two forms of text input associated with georeferenced images in the British Isles: firstly short descriptive texts from Geograph, and secondly, tags associated with images in Flickr.

Global analysis methods

The first question that we can ask of any corpus concerns its composition. These are simple questions of frequency – what words occur and how often, and how are frequencies distributed in a corpus. A second, often neglected question is to ask, are the answers to the former in any way surprising?

For narrative text, function words (prepositions, conjunctions, pronouns) will typically be most frequent in any corpus and only by filtering out such terms (often called stop words) or exploring specific parts of speech (for example the use of nouns and proper nouns) can peculiarities of a collection with respect to general language be explored (Manning & Schütze 1999; Purves et al. 2011) (Table 1). Word frequencies in a corpus typically broadly follow Zipf's law – frequency is inversely proportional to rank. This implies in turn that a small number of different words account for a large proportion of the total word count in any corpus, and many words occur rarely in a given corpus.

It is important to note that tag lists are typically shorn of much the accoutrement of narrative text, and consist of relatively informative, freestanding terms (O'Hare & Murdock, 2012; Purves et al. 2011; Rattenbury & Naaman 2009). Thus, frequency counts of tags may already be informative with respect to semantic content, with for example around 80% of the Flickr tags analysed by Purves et al. (2011) taking the form of generic nouns (e.g. **church**[4], **hill**, **wedding**) or proper nouns (e.g. **tom, monday, nikon, edinburgh**) (Table 1). Hollenstein and Purves (2010) reported an average of 25% of tags as referring to locations and Rattenbury and Naaman (2009) identified some 12-16% of tags as being 'place tags'. Place tags still typically show Zipfian distributions.

In the above we effectively ignore the semantics or meaning of individual terms or tags. Thus, **forest** and **woods** are treated as entirely independent terms, as are **New York** and **Big Apple**, despite their obvious overlapping meanings. The first step in dealing with this problem is tokenisation – that is parsing some given input text to a set of meaningful units. This, at first glance, trivial problem is anything but. Approaches to tokenisation can have significant impacts on results (for example, is **New York** one token or two?) (Manning & Schütze 1999: chap. 4). The second step typically involves the use of more advanced methods such as lemmatisation and tagging of parts of speech, which fall firmly into the domain of natural language processing. Once again, the popularity of tags can be attributed to their simple structure, but it is important to note that

[4] We refer to tags in the text thus: **tag**

	Geograph (Top 10)		Geograph (Top 10 nouns)			Flickr (Top 10)		
Rank	Count	Word	Rank	Count	Word	Rank	Count	Tag
1	426936	the	13	45768	road	1	187605	london
2	275878	of	21	24085	view	2	97696	england
3	189089	to	24	21119	farm	3	96622	uk
4	184705	a	32	17242	lane	4	40528	2007
5	179553	in	36	16232	hill	5	34032	scotland
6	171429	and	37	16157	church	6	29654	unitedkingdom
7	153707	on	38	15815	bridge	7	24525	2006
8	152091	is	43	14737	river	8	21535	edinburgh
9	141579	from	45	14150	square	9	20215	ireland
10	132451	this	48	13690	house	10	17596	dublin

Table 1: Most frequent terms from narratives of 912874 Geograph photographs and tags of 759638 Flickr photographs for data collected in a bounding box corresponding to the British Isles in April 2008 (more details in Purves et al. (2011)).

this does not remove problems of, for example, ambiguity (e.g. does the tag **bath** refer to a town in England or a place to wash oneself?).

One approach taken to explore in more detail how words or tags are semantically related to one another is the use of co-occurrence to identify meaningful collocations – an 'expression consisting of two or more words that correspond to some conventional way of saying things"' (Manning & Schütze 1999: 151). The key task here is to disentangle expressions which co-occur by chance from those whose co-occurrence is statistically and semantically meaningful.

A surprisingly effective and efficient approach to this is adding some form of structure to words or tags found in a collection through annotation. Such annotation tasks often take the form of the formulation of a set of rules, applied independently by a group of annotators, in which final decisions about class membership is based on some majority decision (e.g. Purves et al. 2011; Rattenbury & Naaman 2009). Thus, for example, Purves et al. (2011) generated a simple taxonomy classifying words or tags as elements (things that are visible in an image), qualities (properties which might modify an element or suggest feelings or moods) and activities. Using this taxonomy it was then possible to explore co-occurrence, and identify both meaningful collocations or co-occurrences (e.g. **steep hill** or **city park**). Annotation tasks such as those described here can be seen as substituting specialised task-defined term dictionaries for more commonly available, but less specific, semantic resources such as WordNet (Miller 1995).

User behaviour and language

Other chapters in this book concern themselves with issues of participation inequality – the basic notion that a small number of users contribute much of the content to most examples of user generated content. The importance of this observation in analysis of georeferenced annotated photographs is straightforward – are we analysing the way in which many people have described a particular type of photograph (and their locations) – or the behaviour of only a few? Thus, for example, tags describing **trucks** and **lorries** were the 21st and 22nd most frequent in a collection of 450,272 photographs contributed by a total of 12,682 users, but only used by 15 and 7 users respectively. By contrast, the most frequent tag, **edinburgh**, was used by a total of 7,427 users, and the 20 most frequent tags were all used by more than 300 users. However, simply being used rarely does not *per se* indicate that a tag is not meaningful. In this particular case **trucks** and **lorries** are presumably the subject of interest of a small group, but this does not mean that the locations where they were photographed are unrepresentative. Considering the influence of individual users on tag semantics is therefore an important, and ongoing research challenge, in the analysis of annotated georeferenced photographs.

Purves et al. (2011) explored tagging behaviour by binning all photographs contributed to a collection, sorted by user prolificness. Histrograms of individual tag usage then showed the proportion of tags contributed by more or less prolific users, along with z-scores provided a summative value indicating whether a tag was used in similar ways by all contributors to a collection. This approach has the advantage of allowing exploration of individual tags, rather than contributions, and their influence through user behaviour. Furthermore, it provides a way of dealing with bias caused by, for example, bulk uploads, at the level of individual tags, rather than users.

Language, location and scale

In a book on Volunteered Geographic Information it is of course the location of information which is of primary interest. Georeferenced images were adopted very rapidly by researchers in this area because not only were locations explicitly recorded, but the assumption that the content was linked to a location is more immediate and seems more realistic in describing images taken *somewhere*. However, issues of granularity quickly become apparent, with for example the most frequent three tags in a collection of 1,520,212 images captured within the bounding box of Scotland being **scotland**, **edinburgh,** and **glasgow** respectively (Figure 1). Clearly **scotland** is not wrong, but neither is it informative. This problem is identical to that illustrated by the top ten words from Geograph in Table 1 – **the** is indeed a very frequent word, but it isn't terribly interesting!

One approach to identifying more interesting terms is to home in on those which more effectively characterise a document by comparing frequency of a chosen term in a given document to frequency across a corpus as a whole. This approach is known as term frequency- inverse document frequency (tf-idf) and is a baseline ranking method in information retrieval. It can be applied in a geographical context by counting the number of images with a particular tag within a prescribed region (or cell) and comparing this with frequency over a larger geographic region (Ahern et al. 2007; Rattenbury & Naaman 2009). The basic effect of geographical applications of tf-idf is to privilege locally common, but globally rare tags over globally common tags. Recognising the nature of user generated content and the issues relating to user behaviour described above, many researchers have added a term to capture user frequency in this characterisation, typically ranking tags used by many higher within in a region (Ahern et al. 2007; Feick & Robertson 2014; O'Hare & Murdock 2012; Rattenbury & Naaman 2009).

Obviously the size and form of the regions within which frequencies are calculated will have an influence on the results. The former property, size effectively captures notions of scale, while the latter, form, is closely related to the classical Modifiable Areal Unit Problem (MAUP). To capture notions of scale it is important to characterise tag semantics at multiple scales (c.f. Ahern et al. 2007; Feick & Robertson 2014; Rattenbury & Naaman 2009). Dealing with MAUP has led to a number of approaches. Rattenbury and Naaman (2009) and Ahern et al. (2007) generated regions bottom up, by clustering on photograph positions themselves using K-means. Feick and Robertson (2014) imposed a multi-scale hexagonal tessellation, which they is argued is better able to capture the complex geometries of real world regions. They explored similarity between tag characterisation of connected hexagons to identify larger semantic regions.

Figure 1 illustrates some of these notions for a dataset consisting of 1,520,212 photographs, containing a total of 53,842 unique tags and captured by 31,292 unique users. The ten most common tags were: **scotland, edinburgh, glasgow, uk, united kingdom, geotagged, england, music, uploaded:by=flickr_mobile,** and **highlands**. Seven of these are toponyms, but contain little or no useful information (the images were all from within Scotland's bounding box, and Edinburgh and Glasgow are simply the two most populous cities). Two (**geotagged** and **uploaded:by=flickr_mobile**) refer to properties of the data which are self-evident in the first case and refer to an application used to deliver data in the second. Finally, **music** reflects Flickr's popularity as a platform for describing leisure activities (Antoniou et al. 2010). Figure 1 ranks tags using three methods discussed above for a square grid. Firstly, tags are ranked using only frequency and, as was the case in Table 1, simply reflect characteristics of the collection as a whole (note the predominance of **scotland**). Secondly, tf-idf, filtered for multiple users gives back a much more local picture, and is dominated by more local toponyms, with the exception of larger cities, where activities and their locations (e.g. **fringe festival**, **murrayfield** (rugby), and **bongo**

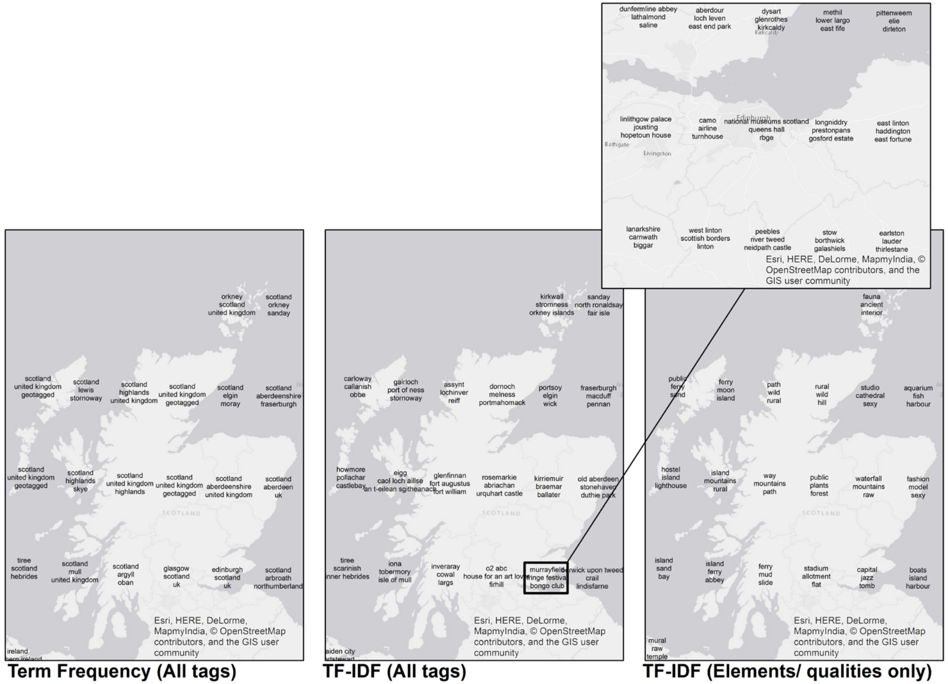

Figure 1: Top ranked tags for Flickr images in Scotland's bounding box according to term frequency, TF-IDF and TF-IDF filtered using list of elements and qualities according to [ref]. Only terms used by a minimum of 3 users are shown. Regions are defined as 1 degree x 1 degree.

Inset map shows top three terms calculated by TF-IDF for region around Edinburgh (Region defined as 0.2 degree x 0.2 degree).

club (Nightclub, gig, and events venue) for Edinburgh) become visible. Zooming in to a more detailed grid using tf-idf reveals finer granularity toponyms. To start to explore not only the names of the locations in grid cells, but what sorts of places these might be, tags are filtered according to a structured list from Purves et al. (2011). The resulting tf-idf values show locations associated with, for example, outdoor activities (**rural, wild, hill**) or more urban locations and activities (**stadium, allotment, flat**).

The techniques described so far focus on tags independent of one another. But, as discussed above co-occurrence can reveal more semantically rich information (e.g. **castle ruin** or **tall building**) and by using (most profitably) interactive visualisations such co-occurrence can be geographically located (Dykes

Figure 2: Top ten Geograph terms describing elements and their co-occurrence with one another presented as a spatial treemap Dykes & Wood (2009). The size of a rectangle indicates the overall count of co-occurrences for a particular term, while the nested rectangles indicate the relative predominance of individual collocates, and the colours link these to location – thus, for example, the most common terms used with **farm** are **hill**, **house**, **lane** and **road**. Figure adapted from data published in Purves, Edwardes & Wood (2011).

and Wood 2009), for example by using spatial treemaps (Dykes & Wood 2009). Spatial treemaps are hierarchical structures which can show 1) the overall occurrence of an individual term, 2) the most commonly co-occurring terms associated with each term, and, when linked to a key using colour 3) the locations of the co-occurrences. Figure 2 shows co-occurrences of the top ten most frequent elements, together with a colour legend linking the distribution for co-occurring terms across the British Isles in the Geograph dataset. Such visualisations allow us to start to explore the link between particular sorts of locations and their properties, for example the relative importance of **river** and **road** with respect to **bridge** compared to the importance of **land**, **house, road,** and **hill** with respect to **farm**.

Recommendations

The motivations for analysing geotagged imagery are as varied as its contributors. Thus the challenge lies not in the analysis per se, but in the initial processing of the data and in the interpretation of the results. Consensus need not be a prerequisite in extracting semantics; just because a prolific user contributes images of a highly thematic form does not make that contribution biased. However, some basic understanding of what properties in a collection might be surprising and a related awareness for the spectrum of existing approaches are both indispensable. In this short chapter we have scratched the surface of available methods – however we hope this material and the related references will prove a useful starting point for researchers new to the area.

Of course, the astute reader is still waiting for a silver bullet – but the reality is that all techniques should be seen as exploratory, and that great care is required in the interpretation of these qualitative outputs. Nonetheless, we recommend the following basic considerations, which we link here to the questions set out in the introduction:

- Global views on datasets allow an initial quick view of datasets (Q1)
- Consideration of the meaning of tags, and an understanding of potential ambiguities can be aided by simple methods such as co-occurrence (Q2)
- User behaviours can lead to significant biases, for example through bulk uploads and users with particular thematic interests (Q3)
- Purely frequency-based methods are unlikely to reveal interesting spatial patterns – however, simple methods such as tf-idf can rapidly increase the amount of information available in collection (Q4)
- When analysing geographic data basic notions such as scale and MAUP cannot be forgotten (Q5)
- Novel visualisation techniques can provide useful insights and lead to the generation of new hypotheses (Q5)

Acknowledgements

Much of our contribution in this paper derives from work undertaken with a wide variety of colleagues over the years. We would like to in particular acknowledge the contributions of Alistair Edwardes, Livia Hollenstein and Jo Wood to the research reported in this paper, though all errors are of course our own. We would also like to acknowledge contributors to Geograph British Isles, see http://www.geograph.org.uk/credits/2008-04-16, whose work is made available under the following Creative Commons Attribution-ShareAlike 2.5 Licence (http://creativecommons.org/licenses/by-sa/2.5/). RSP gratefully acknowledges the support of the Swiss National Science Foundation project PlaceGen (200021_149823).

References

Ahern, S., Naaman, M., Nair, R., & Yang, J. H.-I. 2007. World explorer: visualizing aggregate data from unstructured text in geo-referenced collections. In: *JCDL'07: Proceedings of the 2007 Conference on Digital Libraries*. ACM, New York, NY, US, pp. 1–10.

Andrienko, G., Andrienko, N., Mladenov, M., Mock, M., & Pölitz, C. 2010. Discovering bits of place histories from people's activity traces. In: *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. pp. 59–66.

Antoniou, V., Morley, J., & Haklay, M. 2010. Web 2.0 geotagged photos: Assessing the spatial dimension of the phenomenon. *Geomatica, 64*: 99–110.

Dykes, J., & Wood, J. 2009. The geographic beauty of a photographic archive. In: Segaran, T., & Hammerbacher, J. (Eds.) *Beautiful Data*. O'Reilly, Sebastapol, CA, pp. 85–102.

Feick, R., & Robertson, C. 2014. A multi-scale approach to exploring urban places in geotagged photographs. *Comput. Environ. Urban Syst.*

Girardin, F., Calabrese, F., Fiore, F.D., Ratti, C., & Blat, J. 2008. Digital Footprinting: Uncovering Tourists with User-Generated Content. *Pervasive Comput. IEEE, 7*: 36–43.

Hollenstein, L., & Purves, R. 2010. Exploring place through user-generated content: using Flickr to describe city cores. *J. Spat. Inf. Sci, 1*: 21–48.

Kessler, C., Maué, P., Heuer, J.T., & Bartoschek, T. 2009. Bottom-Up Gazetteers: Learning from the Implicit Semantics of Geotags. In: *GeoS'09: Proceedings of the 3rd International Conference on GeoSpatial Semantics*. Springer-Verlag, Berlin, Heidelberg, pp. 83–102.

Manning, C. D., & Schütze, H. 1999. Foundations of statistical natural language processing. MIT press.

Miller, G. A. 1995. WordNet: a lexical database for English. *Commun. ACM, 38*: 39–41.

O'Hare, N., & Murdock, V. 2012. Modeling Locations with Social Media. *J. Inf. Retr., 16*: 13–62.

O'Sullivan, D., & Unwin, D. J. 2003. Geographic information analysis. Wiley, Hoboken, New Jersey.

Purves, R. S., Edwardes, A. J., & Wood, J. 2011. Describing place through user generated content. *First Monday, 16*.

Rattenbury, T., & Naaman, M. 2009. Methods for extracting place semantics from Flickr tags. T*WEB, 3*: 1–30.