TagNText: A parallel corpus for the induction of resource-specific non-taxonomical relations from tagged images

Theodosia Togia, Ann Copestake
Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue, Cambridge CB3 0FD
tt309@cam.ac.uk, aac10@cam.ac.uk

Abstract

When producing textual descriptions, humans express propositions regarding an object; but what do they express when annotating a document with simple tags? To answer this question, we have studied what users of tagging systems *would have said* if they were to describe a resource with fully fledged text. In particular, our work attempts to answer the following questions: if users were to use full descriptions, would their current tags be words present in these hypothetical sentences? If yes, what kind of language would connect these words? Such questions, although central to the problem of extracting binary relations between tags, have been sidestepped in the existing literature, which has focused on a small subset of possible inter-tag relations, namely hierarchical ones (e.g. "car" –is-a– "vehicle"), as opposed to non-taxonomical relations (e.g. "woman" –wears– "hat"). TagNText is the first attempt to construct a parallel corpus of tags and textual descriptions with respect to particular resources. The corpus provides enough data for the researcher to gain an insight into the nature of underlying relations, as well as the tools and methodology for constructing larger-scale parallel corpora that can aid non-taxonomical relation extraction.

Keywords: collaborative tagging, relation extraction, folksonomy

1. Introduction

Collaborative tagging is a cheap and easy way to produce information about digital documents. However, the resulting representation has low expressive power, since the inter-tag relations that users bring with them in the tagging process are not recorded. Tagging interfaces force – or give freedom to – users to provide just tags, omitting any relations that might connect the corresponding concepts in the users' minds. Put differently, tagging systems oblige users to exclude part of the language they might otherwise have used to describe resources. Inducing such relations could raise the semantic content of tag data, which could, in turn, facilitate search through query expansion or produce material useful for semantic applications.

It is safe to assume that most implicit statements made by users while tagging connect concepts with non-taxonomical relations. For example, while annotating a resource, such as an image, with "kids", "garden", "ice-cream", "people" and "flowers", users are more likely to be forming thoughts such as "The kids *are eating* ice-cream" and "The garden *contains* flowers" rather than "Kids *are* people". Nevertheless, research has concentrated on extracting hierarchical relations (subsumption, instantiation and equivalence) between folksonomy tags. Hierarchical relations are typically induced with graph-based techniques (Heymann and Garcia-Molina, 2006; Benz and Hotho, 2007) or Association Rule Mining (Schmitz et al., 2006; Lin et al., 2009) and attempt to reveal the taxonomy underlying an entire folksonomy. The few studies that have been conducted on extracting non-taxonomical relations have used some form of ontology as a corpus (Specia and Motta, 2007; Maala et al., 2007; Angeletou et al., 2008; Sordo et al., 2010), a process which suffers from data sparsity since structured data, in contrast to tagging data, are hard to acquire. A general-purpose text corpus has been used for a small part of the research by Trabelshi and his colleagues (2010). However, the relations induced are not resource-specific, and therefore make no prediction about the implicit statements made by users while tagging. Rather, such relations establish general facts such as "birds" –fly in– "sky".

To help overcome the above limitations, we have constructed TagNText, a corpus that contains both tags and text-like descriptions of five images, provided by 219 participants (i.e. 1095 tag-text parallel annotations). Although this parallel corpus has been compiled on a small scale, it provides a valuable language resource for theoretical research (e.g. on tagging behaviour and the nature of implicit relations) and lays the foundations for the construction of much larger corpora that can aid non-taxonomical relation extraction.

In this work we are exploring tagging and description patterns of *images*, as opposed to digital resources that often contain spoken or written language (e.g. bookmarks, videos, audio). The main reason is that, when labelling images, users perform tagging at its purest form, without the interference of language already existing in the resource. Moreover, the lack of words in images renders non-taxonomic relation extraction especially challenging, since the resource itself cannot provide us with relevant language that will assist in our task.

We also restrict our work to *resource-specific*, as opposed to folksonomy-wide, inter-tag relations, that is relations between the tags describing a particular document. This is what Peters and Weller (2008) refer to as 'syntagmatic relations'; ones which can hold between tags in the context of a resource.

Since the vast majority of studies on relation extraction between tags focus on taxonomical relations, resource-specificity has, so far, been irrelevant. For instance, inducing the relation ‘is equivalent to’ between the tags ‘car’ and ‘automobile’ neither profits from nor requires a focus to a particular resource, since **i)** it makes a statement that can be true of every resource that has been annotated with both tags and **ii)** it does not necessitate that the two tags have been drawn from the same resource. Nonetheless, when attempting to extract generic, non-taxonomical relations, the resource tagged (the image in our case) is an essential point of reference. For example, the relation between tags ‘boy’ and ‘tree’ can be ‘climbing’ for one image and ‘sits under’ for another image. Additionally, the co-existence of two tags within an image’s annotations provides some evidence for the existence of a possible connection between the tags and, thus, legitimises the search for a relation.

TagNText was compiled not only for the purpose of investigating the nature of tag-tag relations, but also as a precursor to larger parallel corpora, since, ultimately, image-specific relation extraction may require the use of large corpora; ones that, for any unseen tagged image that we need to extract relations for, will contain similarly annotated images with corresponding text. Such a corpus can be created by various means, including crowd-sourcing. One might think that an alternative to creating a large parallel corpus is image processing, which could allow us to ‘see’ the relations between the objects denoted by the tags. However, what we see with this method is limited to attributes and (relative) positions of objects, which is only a small subset of the relations that seem to lie beneath a collection of tags for an image. At the moment we are also investigating the extent to which text corpora alone can be used instead of large parallel corpora, but this is outside the scope of this paper.

In the next section we provide an overview of tagging as a way to ‘talk about’ an image. In section 3. we present some previous work on creating corpora of tags (from folksonomy and image or video annotation experiments) with or without corresponding text, while in section 4. we outline the methodology and process of constructing TagNText. Section 5. describes the data acquired, explaining how the corpus has been processed so far and how it can be used in the future. Finally, section 6. concludes the paper.

2. Tagging

Tagging is the practice of annotating digital resources with uncontrolled vocabulary. It can be seen as the evolution of subject cataloguing for libraries as it produces metadata whose implied relationship with the document it is attached to is nothing more than ‘about’. Tagging is a form of *descriptive*, as opposed to prescriptive, document indexing (Tennis, 2006).

An analysis of two tagging systems Delicious[1] and Flickr[2] (Tonkin and Guy, 2006) revealed that 90% of the tags submitted on the websites are nouns. Spiteri (2007) found that nouns account for 88% of the tags used, while 6% are adjectives. Her conclusion was that the tags people use in order to annotate resources tend to represent things.

But why do people tag? Strohmaier and his colleagues (2010) distinguish between two types of activities that users perform on a tagging system, *categorising* resources and *describing* resources. These activities aim to facilitate later browsing and later retrieval respectively. When categorising a resource, tags can be approximated by categories of Flickr images; when describing it, tagging behaviour can be simulated using data from ESP games (von Ahn, 2006), in which users try to guess each other’s tags for an image and end up labelling the image with highly descriptive tags. Strohmaier and his team show that different tagging systems motivate users differently (see (Trant, 2009a) for a summary of other reasons for tagging). In the tagging system that we have used for our experiments (Steve folksonomy; see section 4.), the motivation is stated as describing (“user-contributed descriptions”) resources for facilitating later retrieval (“improving on-line access to works of art”) (Trant, 2009b).

Tags can be assigned to a particular resource by either its author or by everyone (Vander-Wal, 2005). The latter case is more interesting since it allows us to construct the ‘social meaning’ of a resource, created by means of collective intelligence. When different users annotate the same resource, the result is a multi-set of tags, which can be visually represented through a tag cloud. Such a representation informally resembles a fragmented piece of text, which invites us to find the ‘missing bits’, that is the inter-tag relations.

According to Peters and Weller (2008), there are two types of inter-tag relations, *paradigmatic* ones and *syntagmatic* ones. The former are context-independent and can hold between tags from across resources. The latter are context-specific and hold between tags within a particular resource.

It's a landscape of a cottage in the country side. The rectangular cottage is set in a desolate setting, with many dead trees surrounding it (although there are also many live trees and bushes, and plenty of greenery). There is a mountain-range, or set of hills in the background.	***TEXT***
house, landscape, painting, countryside, cottage, mountains, hills, desolate, isolated	***TAGS***

Figure 1: Resource “house” and responses of Participant 33

[1] http://www.delicious.com/
[2] https://www.flickr.com/

To enhance the tags with induced syntagmatic relations for images, we need to gain a deeper understanding of resource-specific text and identify patterns which help us decide on the missing relations between tags (see Figure 1 for instance).

3. Related work

Some notable attempts to create tagging corpora for research use were made within the TAGora project (2006-2009)[3], although none of the datasets available are tags-text parallel corpora. The reason is that such information, that is text explaining the use of tags, is simply not available on the web, with the exception of some sparse and mostly laconic captions of tagged images (e.g. on Flickr[4]) or notes on tagged bookmarks (e.g. on Delicious[5]).

A large-scale corpus that contains annotations of both tags and text is Social-ODP-2k9[6] (Zubiaga et al., 2009). The corpus consists of 12,616 unique URLs (bookmarks) from Delicious, all of which are annotated with, among other things, the number of users tagging the bookmark, its top 10 tags, user notes and reviews from StumbleUpon[7]. Despite its large coverage and tag-text parallel data, the corpus is not ideal for our task since **i)** Delicious notes do not amount to descriptions; they simply constitute complementary information, **ii)** the resources are typically textual.

Two datasets from the Referring Expression Generation field (the GRE3D7 corpus (Viethen and Dale, 2011) and the Wally Referring Expression Corpus (WREC)[8] (Clarke et al., 2013)) contain textual descriptions of entities, used for modelling how humans select attributes that distinguish entities in an image from 'distractors'. GRE3D7 is a collection of 4480 descriptions of geometrical objects in different shapes, sizes, colours and positions. As evident from our Introduction, such a dataset would be too limited for the scope of relationships we wish to cover. WREC, on the other hand, contains 4256 descriptions of individuals in 28 different real and visually complex scenes. It encodes wide-ranging relations between objects (e.g. 'a man with lots of light blue hair and a grin') but still lacks the tag-text duality that we are after (as it contains only text). Given that such descriptions are similar to our intended end-product (assuming, for instance, that 'man' and 'hair' are tags submitted for the folksonomy image in question), they might be useful as a gold-standard for the evaluation of our resource-specific non-taxonomical relation extraction system. However, as a corpus, it still cannot inform us of the nature of relations implicit between tags, like TagNText does.

Finally, an interesting attempt at creating a tag and text parallel corpus is the work of Khan and his colleagues (Khan et al., 2012), from the field of video annotation. In their experiment, 140 videos from TREC data were annotated by 13 people with a title, keywords and a textual description (1820 annotations in total), that can be used for improving video search. This corpus is the closest to TagNText with respect to its structure. It would be interesting to explore whether the inter-keyword relations revealed from the text in the videos are of a similar nature to the inter-tag relations uncovered through TagNText.

[3]http://www.tagora-project.eu/data/

[4]http://www.flickr.com/

[5]http://www.delicious.com/

[6]http://nlp.uned.es/social-tagging/socialodp2k9/

[7]https://www.stumbleupon.com/

[8]http://datashare.is.ed.ac.uk/handle/10283/337

4. Corpus construction

In order to construct the parallel corpus, we designed an experiment in which participants perform *tagging* under conditions similar to the ones in tagging websites, and also *describe* the images in a way that is as free of bias as possible. The main hypothesis of the experiment was that the tags used by a participant for a given picture resemble the words that the same participant uses while coherently describing the same image. If this is true, then the text should provide useful information as to the relations underlying particular tag pairs. Below is a step-by-step description of the corpus construction process[9].

4.1. Determining experiment objectives

The main goals of the experiment can be summarised as follows:

- to acquire tagging data and natural language descriptions in a controlled environment.
- to build a folksonomy which is enriched with fully fledged natural language descriptions provided by each user for each picture along with the respective tags.
- to measure the extent to which tags in a particular image annotation appear as words in the description of the same image.
- to measure the extent to which pairs of tags can be found as words in the same sentence within a description and to examine the nature of the natural language stretch that connects such pairs of words in the descriptions (i.e. the *relations* to be extracted). This 'nature' can be described in many ways, for example, in grammatical dependencies holding within the tag-language-tag sequences found in the descriptions.
- to test whether this interaction between descriptions and tags is the same for different groups of users (e.g. native speakers vs. non-native speakers of English).
- to test how this tag-description interaction differs if tags and descriptions are examined *collectively* (i.e. from all users) as well as within *individual* user annotations.

4.2. Choosing stimuli

Five images were chosen from the Steve folksonomy[10] (Trant and Wyman, 2006) to be presented to participants. The folksonomy was a result of an ongoing project in which

[9]Additional information can be found on http://www.cl.cam.ac.uk/~tt309/TagNText.html

[10]http://tagger.steve.museum/

museum professionals and everyday people annotate images of art objects from museums across the United States, so as to enrich official metadata. The images selected for our experiment were drawn from a subset (33,948 images) of the Steve folksonomy that we collected by crawling the website during a three-week period in October and November 2011. Highly tagged images were preferred because this allows the original tags to be later compared to the tags acquired from our experiment. Among the top 70 most tagged images[11], five were hand-picked according to the following criteria:

1. balance of multi-word to single-word tags: Images with more single-word tags were preferred as more representative of the whole folksonomy and as a way to avoid complication when later comparing the original tags with the ones obtained through our experiment.

2. complexity of image: Simplistic images (e.g. a single sculpture against a white background) were dispreferred because of the limited possibilities they offer in term of possible tags and underlying relations.

3. diversity: The five images chosen were reasonably different to each other in terms of themes, style and clarity of the messages conveyed.

During our experiment, the original tags, official metadata and previous participants' tags were not available to the subjects.

4.3. Outlining experiment structure

The experiment was designed to involve two tasks, a *tagging* task and a *description* task in two separate phases. We decided that half of the participants should start (Phase One) with the tagging task and proceed (Phase Two) with the description task, while the other half should describe first and then tag.

4.4. Establishing relevant participant groups

We decide that each participant should be assigned to three groups (one from each of the three categories below), after ticking the relevant boxes before the start of the experiment. The categories were:

- native speakers of English vs. non-native speakers of English

- with tagging experience vs. without tagging experience

- starting Phase One with the tagging task vs. starting Phase One with the description task

The reason was to determine if there are any significant differences between the responses of the above groups and, if not, to rest assured that there are no interactions or bias involved.

[11]Repeated tags were taken into consideration. For example, an image with tags 'fruit' occurring three times and 'yellow' occurring twice has five tags in total.

4.5. Designing the interface

An online CGI script was written to collect data from participants. We decided that the call for participants should contain a web link to Phase One. Each participant was asked to provide either tags or descriptions; the script was designed to log responses and assign a tagging or a description task to participants interchangeably. This way each task is completed by an equal number (half +/- 1) of participants. For Phase Two, the script was designed to send a personalised link to those who completed Phase One, asking them to perform the task that they have not yet completed.

For the tagging task, tags were to be provided on separate boxes (fields), allowing for whitespaces to be used as word delimiters between multi-word tags. For the description task there was one big text field.

4.6. Deciding on instructions for participants

Both the tagging task and the description task were performed in the context of real-life scenarios, which were informative enough to be clearly understood and open-ended enough to avoid pre-disposing the users towards a particular tagging or describing behaviour. For the *tagging* task, participants were asked to imagine that they need to organise a personal collection of images on a website and that adding tags will help them retrieve the images in the future: "Imagine that there is an art website which contains images of artworks; let's call it *www.my-personal-gallery.com*. This website allows you to register, choose your favourite art images and create a personal collection. In order to organise your images and be able to find them in the future, the website allows you to label them with keywords (tags). Now you will be shown 5 pictures. Please provide tags for each one of them. You are free to type in anything as a tag as long as it helps you retrieve the picture from your collection later." Since personal retrieval is the main reason why people perform tagging (Vander-Wal, 2007), this scenario is meant to elicit tags similar to those found in existing tagging websites.

For the *description* task, participants were asked to imagine that they are at a book shop browsing a book of images, and a person with impaired vision is asking them to describe what the images are about. This scenario was presented to help users produce descriptions without forcing them to do too much guesswork. At the same time, it avoids prescribing a particular format (e.g. "write a paragraph") or a particular information content (e.g. "describe what you see", "describe how the image makes you feel"):

"Imagine that you are in a bookshop holding a book in your hands. The book contains art images. Next to you there is a person with impaired vision and they are asking you to describe and explain to them what the pictures are about. Now you will be shown 5 pictures. Please describe them to this person."

4.7. Piloting

Before the final experiment, three pilots were conducted (with three, two and two participants respectively), that tested various parameters and possible complications of the experiment process. Feedback was received on the usability and technical aspects of the interface as well as the clarity of the instructions.

4.8. Recruitment

The participants were restricted to University of Cambridge members and were recruited through emails to particular departments, colleges and societies. The reason for limiting our sample to a university-internal audience was to avoid introducing too many demographic parameters that might inhibit our ability to make generalisations about the data collected. Moreover, we believed that methods like crowd-sourcing (e.g. with Amazon's Mechanical Turk[12]) would introduce unnecessary noise, for instance, dishonest responses that are hard to filter out (Ipeirotis et al., 2010). To motivate email recipients to engage in the task we offered every participant who completes both phases of the experiment participation in a draw for a £100 voucher. We also emphasised the fun aspect of the experiment and clearly stated the maximum amount of time that completion of each task was expected to take.

4.9. Data collection

Phase One was completed in mid-February 2013 by 267 people (134 providing tags and 133 providing descriptions). Phase Two took place 15 days later (to help minimise repetition bias) and was completed by 219 participants. Data from the 48 participants who did not continue to the second phase was ignored.

5. The dataset

In total 219 participants completed both phases of the experiment. Following the experimental design (Sections 4.3. and 4.4.), we collected four kinds of information from each participant: **i)** the task they performed in Phase One (tagging or describing), **ii)** their command of English (native vs. non-native speakers), **iii)** their familiarity with the tagging process (with vs. without previous experience), **iv)** descriptions and tags for each one of the five images.

The data was saved in a csv (comma separated value) file, where each line represents data from one participant at a time. The full corpus and the data collection interface will be made publicly available in the near future. A sample from the first three participants can be downloaded from `http://www.cl.cam.ac.uk/~tt309/TagNText.html`. Figure 2 shows the original data and part of the TagNText data for one of the images.

Half of the dataset (containing responses from 110 participants) is currently being used for processing (e.g. various measurements and relation extraction for the images available), while the rest is being kept aside for use as development and test data after the initial processing is completed. The following have been measured or performed:

- average **length of descriptions** per image (56.1 words)
- average **number of tags** per image (5.5 tags)
- **percentage of hapax legomena** in different participant groups (native speakers: 49.1%, non-native speakers: 33.3%, familiar with tagging: 37.5%, unfamiliar with tagging: 46.1%, performed tagging on Phase One: 44.2%, performed description on Phase One: 39.1%). Some of the hapaxes were due to spelling mistakes since spelling correction lowered the percentages of all groups by approximately 4%.
- **tag-word overlap**: Approximately 51.4% of the tags used by a participant (*individual overlap*) for a given image were found in his/her description of the same image[13]. After lemmatisation, the percentage increased to 53.4%. This is evidence that people tend to use the same words to tag and describe an image, therefore, image tagging can be seen as a form of fragmented textual description. Such evidence legitimises our search for relations between the 'fragments'. The overlap was also measured after dumping together all the tags for a given image (i.e. from all participants) and all the descriptions (*collective overlap*). On average 69.4% of the tags used by all participants for each picture appear as words in the collected descriptions for the same image. The percentage rises to 70.1% after lemmatisation. As can be seen, if we calculate overlap from all users collectively, the percentage is higher, which might suggest that, with collective consideration of tags and texts we are likely to extract (recall) more underlying inter-tag relations. Such collective relations will predict not what a particular user had in mind, but how the 'public opinion' or 'collective mind' would describe the image in question if they were to use language more complex than tags.
- **tag pair - word pair overlap**: Individual overlap for tag pairs into text was 16.7% (17.8% lemmatised). Collective overlap was 31.5% (33.4% after lemmatisation).
- **extraction of dependency patterns** between tag pairs found in the dependency-parsed descriptions provided by the users (using Briscoe and Carroll-style dependencies (Briscoe and Carroll, 1995; Briscoe, 2006) via the C&C parser (Curran et al., 2007)). Dependency patterns (paths) were extracted between tags in the dependency graphs of sentences. The paths were non-overlapping and had distance up to 3 (i.e. four nodes/words; three edges/dependencies). Among these, some filtered patterns are being used at the moment to extract relations from wikipedia.

  Extracting relations from a general-purpose corpus can help us determine the extent to which a text-only corpus can make predictions about resource-specific relations. Intuitively, tag-relation-tag triples extracted from a text corpus alone will be true of a particular image only if there is some degree of predictability (e.g. "house" + "countryside" are usually related with an "in the" relation in most contexts).

[12] `https://www.mturk.com/mturk/welcome`

[13] The process involved splitting multi-word tags into separate tags

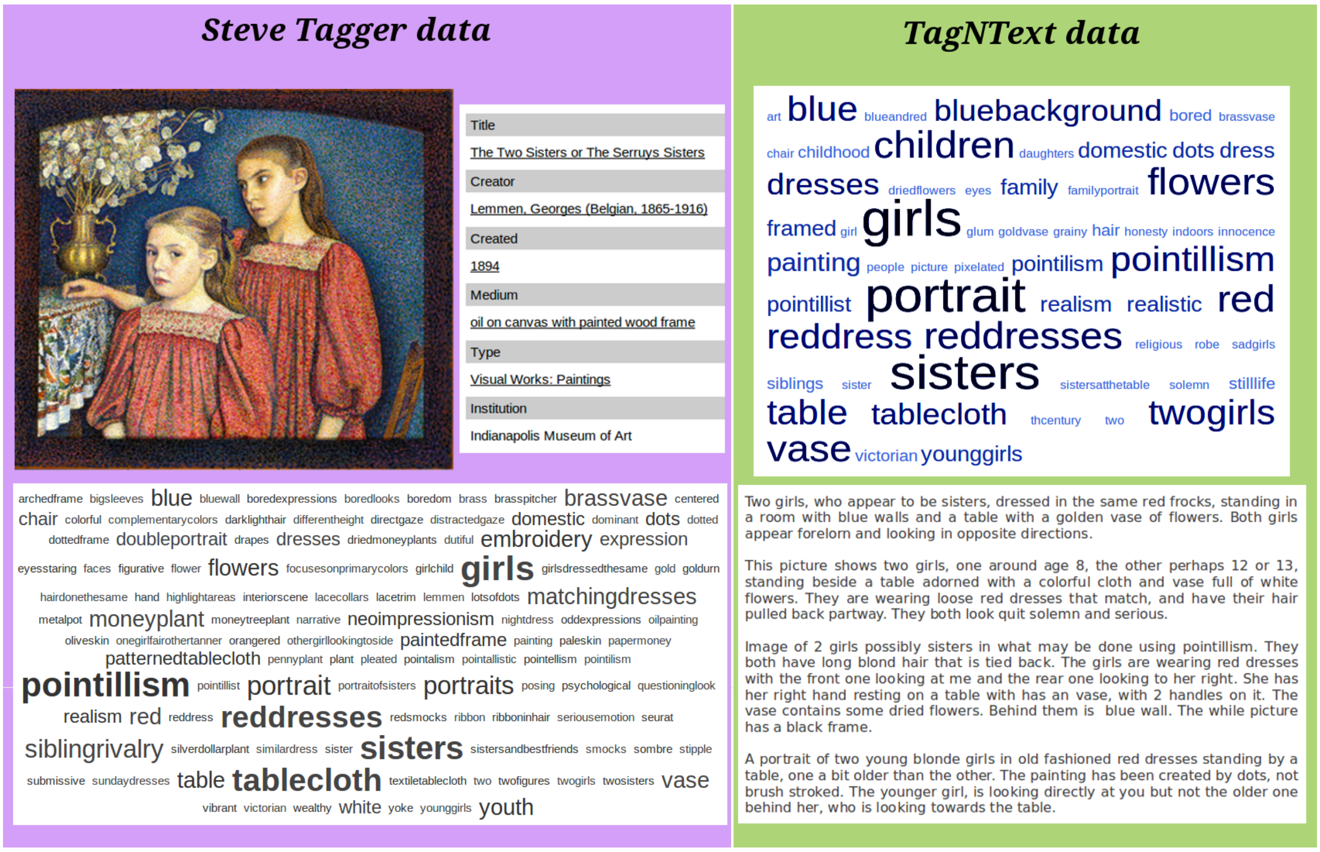

Figure 2: Data from the Steve project and our experiment. *Steve Tagger data* includes the image, the official metadata and a visualisation of relative popularities of tags as labels for the image on the Steve website (`http://tagger.steve.museum/steve/object/231`). *TagNText data* includes a tag cloud from the tags collected for the same image from the human experiment (with hapax legomena omitted because of space limitations) and a sample of descriptions (from four participants)

- **tag order in user annotations** and **word order in descriptions**: This experiment attempted to answer the question of whether it is conceivable to think of (unordered, syntax-free) tags as behaving compositionally. The fact that particular tags work in tandem to assign meaning to an image is evident not only from the high overlap between tag pairs in annotations and word pairs in descriptions, but also from the high number of word pairs that were connected with filtered dependency paths (i.e. 'relations'; see above) in description sentences, when these words are identical to tags. Simply put, there is enough evidence that compositionality occurs between some pairs of tags, however, the question is whether this composition happens *because of* underlying syntactic restrictions between the tags or *despite the absence of* such restrictions. To examine whether tags are as syntactically constrained as their twin words in text, we compared the flexibility of *tag order* within pairs of tags from all users' annotations for an image with the flexibility of *word order* within the corresponding pairs of words in all users' descriptions of the image, when the words were connected with an approved dependency pattern. It was found that in cases where the order of words in a two-word phrase from the descriptions is essentially *fixed* (e.g. words "young" and "girls" appeared only as "young girls" in textual descriptions and not as "girls young"), when tags are elicited, they may be given in *either order* with equal or similar frequency (e.g. "girls"-"young" 50%; "young"-"girls" 50%). Another example could be "vase" and "flowers" appearing as words in descriptions ("vase"-"flowers" 93%; "flowers"-"vase" 7%, in phrases such as 'vase of flowers", "vase full of white flowers", "vase with flowers") and as tags ("vase"-"flowers" 64%; "flowers"-"vase" 36%). Such an observation suggests that tags are largely unordered (or, more plausibly, tag order is more or less irrelevant to users), so the underlying composition between them occurs despite the absence of syntactic restrictions. An analysis of the entire tag and description datasets revealed that the order between two tags co-occurring in a users' annotations was at least 5 times more flexible than the order of the same pairs appearing as words within text descriptions (i.e. the difference between the probability of encountering the tag sequence 'a'-'b' and the probability of tags 'b'-'a' is 5 times smaller than the difference between the probability of words 'a'-'b' and the probability of words 'b'-'a' in descriptions). This finding might point in the direction of a more dynamic compositionality, as understood within the contextualist paradigm in Linguistics (Travis, 1997; Travis, 2000;

Recanati, 2004); a compositionality not strictly bound by syntax.

Our future plans include performing more measurements and comparisons between groups of users with significance testing, that will give us an insight into the tagging behaviour as well as the intended meaning of tag clusters.

6. Conclusion

Non-taxonomic resource-specific relations account for the largest part of the information omitted by users in the tagging process. However, recovering them has been attempted minimally before, largely due to the lack of parallel corpora that would guide relation extraction. TagNText is the first effort for the construction of such a corpus. Although it is a small-scale corpus, it provides **i)** data of interest to anyone researching tagging behaviour and the nature of implicit relations in folksonomy, as well as **ii)** the tools and methods that can pave the way for the construction of a large parallel corpus, which will, in turn, encourage more researchers to address this challenge in the area of inter-tag relation extraction.

7. References

Angeletou, S., Sabou, M., and Motta, E. (2008). Semantically enriching folksonomies with FLOR. In *1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008)*, Tenerife, Spain. 5th Annual European Semantic Web Conference (ESWC 2008).

Benz, D. and Hotho, A. (2007). Position Paper : Ontology Learning from Folksonomies. In Hinneburg, A., editor, *Proceedings of LWA (Lernen, Wissen, Adaption; Learning, Knowledge, Adaptation) Conference, September 2007*, pages 109–112.

Briscoe, T. and Carroll, J. (1995). Developing and evaluating a probabilistic LR parser of part-of-speech and punctuation labels. In *Proceedings of the ACL/SIGPARSE 4th International Workshop on Parsing Technologies*, pages 48–58, Prague/ Karlovy Vary, Czech Republic.

Briscoe, T. (2006). An introduction to tag sequence grammars and the RASP system parser. Technical Report 662, University of Cambridge, Cambridge.

Clarke, A., Elsner, M., and Rohde, H. (2013). Wheres wally: the influence of visual salience on referring expression generation. *Frontiers in Perception Science: Special Issue on Scene Understanding*, 4:392.

Curran, J. R., Clark, S., and Bos, J. (2007). Linguistically Motivated Large-Scale NLP with C & C and Boxer. In *Proceedings of the ACL 2007 Demonstrations Session (ACL-07 demo)*, number June, pages 33–36.

Heymann, P. and Garcia-Molina, H. (2006). Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. Technical report, Stanford University.

Ipeirotis, P., Tamir, D., and Kanth, P. (2010). Mechanical turk: Now with 40.92% spam. `http://www.behind-the-enemy-lines.com/2010/12/mechanical-turk-now-with-4092-spam.html`.

Khan, M. U. G., Nawab, R. M. A., and Gotoh, Y. (2012). Natural language descriptions of visual scenes corpus generation and analysis. In *Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*, pages 38–47, Avignon, France, April. Association for Computational Linguistics.

Lin, H., Davis, J., and Zhou, Y. (2009). An integrated approach to extracting ontological structures from folksonomies. In *6th Annual European Semantic Web Conference (ESWC2009)*, pages 654–668, June.

Maala, M. Z., Delteil, A., and Azough, A. (2007). A Conversion Process from Flickr Tags to RDF. In Flejter, D. and Kowalkiewicz, M., editors, *Business Information System. Workshop on Social Aspects of the Web*.

Peters, I. and Weller, K. (2008). Paradigmatic and syntagmatic relations in knowledge organization systems. *Information Wissenschaft und Praxis*, 59(2):100–107.

Recanati, F. (2004). *Literal Meaning*. Cambridge University Press, Cambridge.

Schmitz, C., Hotho, A., Jäschke, R., and Stumme, G. (2006). Mining Association Rules in Folksonomies. In *Proceedings of the 10th International Federation of Classification Societies (IFCS) Conference*, pages 1–9, Heidelberg, Germany.

Sordo, M., Gouyon, F., and Sarmento, L. (2010). A method for obtaining semantic facets of music tags. In *Workshop on Music Recommendation and Discovery, ACM Conference on Recommender Systems*, Barcelona.

Specia, L. and Motta, E. (2007). Integrating Folksonomies with the Semantic Web. In *Proceedings of the 4th European Web Semantic Conference (ESWC)*, number September 2006, pages 624–639.

Spiteri, L. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, 4(2).

Strohmaier, M., Körner, C., and Kern, R. (2010). Why do users tag? detecting users' motivation for tagging in social tagging systems. In *4th International AAAI Conference on Weblogs and Social Media*.

Tennis, J. (2006). Social tagging and the next steps for indexing. *Proceedings 17th SIG/CR Classification Research Workshop, Austin, Texas*.

Tonkin, E. and Guy, M. (2006). Folksonomies: Tidying up tags? *D-Lib*, volume 12(1).

Trabelsi, C., Jrad, A. B., and Yahia, S. B. (2010). Bridging Folksonomies and Domain Ontologies: Getting Out Non-taxonomic Relations. In *IEEE International Conference on Data Mining Workshops 2010*, pages 369–379. Ieee, December.

Trant, J. and Wyman, B. (2006). Investigating social tagging and folksonomy in art museums with steve.museum. *Web Tagging Workshop at WWW2006, Edinburgh*, pages 1–6.

Trant, J. (2009a). Studying social tagging and folksonomy: A review and framework.

Trant, J. (2009b). Studying Social Tagging and Folksonomy : A Review and Framework. *Journal of Digital Information*, 10:1–42.

Travis, C. (1997). Pragmatics. In Hale, B. and Wright, C., editors, *A Companion to the Philosophy of Language*, pages 87–107. Blackwell, Oxford.

Travis, C. (2000). *Unshadowed thought*. Harvard University Press, Cambridge, MA.

Vander-Wal, T. (2005). Explaining and showing broad and norrow folksonomies.

Vander-Wal, T. (2007). Folksonomy. http://vanderwal.net/folksonomy.html.

Viethen, J. and Dale, R. (2011). Gre3d7: A corpus of distinguishing descriptions for objects in visual scenes. In *Proceedings of the UCNLG+Eval: Language Generation and Evaluation Workshop*, pages 12–22, Edinburgh, Scotland, July. Association for Computational Linguistics.

von Ahn, L. (2006). Games with a Purpose. *IEEE Computer Magazine*, pages 96–98.

Zubiaga, A., Martnez, R., and Fresno, V. (2009). Getting the most out of social annotations for web page classification. In *Proceedings of DocEng 2009, the 9th ACM Symposium on Document Engineering*, pages 74–83, Munich, Germany.