

Social Tagging: Exploring the Image, the Tags, and the Game

Elena Konkova*, Ayşe Göker**, Richard Butterworth* and Andrew MacFarlane*

*Centre for Interactive Systems Research, City University London,
Northampton Square, London EC1V 0HB, England, <Konkova.elena87@yandex.ru>,
<richard@richardbutterworth.co.uk>, <andym@city.ac.uk>

**School of Computing Science and Digital Media, Robert Gordon University,
St Andrew St., Aberdeen, AB25 1HG, Scotland, <a.s.goker@rgu.ac.uk>



Elena Konkova is a former research assistant at City University London. She has an MSc Degree in information systems and technology with a dissertation project on social participation in image collection management. Elena worked for a number of research projects in areas including image retrieval, social web analysis, and user and context modelling.



Ayşe Göker is a professor based at Robert Gordon University in Aberdeen. She has over 20 years of research experience in areas including context-learning algorithms, web user logs, personalization and mobile information systems. Her work involves a strong user-centred approach to algorithm and search-system design, development and evaluation. Prof. Göker also holds a lifelong Enterprise Fellowship from the Royal Society of Edinburgh and Scottish Enterprise for previous commercialisation work.



Richard Butterworth has a background in both information science and human-computer interaction, and worked as a research fellow in the Centre for Interactive Systems Research at City University London until recently. He gained a Ph.D. from Loughborough University studying formal models of interactive systems in 1997, and worked on several HCI academic projects subsequently. He also investigated digital library systems within both academic and commercial contexts. He has developed digital library systems for The Bridgeman Art Library and The English Folk Dance and Song Society. He now works as an independent consultant and developer.



Andrew MacFarlane is a Reader in information retrieval at City University London, and currently co-directs the Centre of Interactive Systems Research with Stephen Robertson of Microsoft Research Cambridge. He got his Ph.D. in Information Science from the same institution. His research interests currently focus on a number of areas including image retrieval, disabilities and information retrieval (dyslexia in particular), AI techniques for information retrieval and filtering, and open source software development.

Konkova, Elena , Göker, Ayşe, Butterworth, Richard, and MacFarlane, Andrew. **Social Tagging: Exploring the Image, the Tags, and the Game.** *Knowledge Organization.* 41(1), 57-65. 21 references.

Abstract: Large image collections on the Web need to be organized for effective retrieval. Metadata has a key role in image retrieval but rely on professionally assigned tags which is not a viable option. Current content-based image retrieval systems have not demonstrated sufficient utility on large-scale image sources on the web, and are usually used as a supplement to existing text-based image retrieval systems. We present two social tagging alternatives in the form of photo-sharing networks and image labeling games. Here we analyze these applications to evaluate their usefulness from the semantic point of view, investigating the management of social tagging for indexing. The findings of the study have shown that social tagging can generate a sizeable number of tags that can be classified as interpretive for an image, and that tagging behaviour has a manageable and adjustable nature depending on tagging guidelines.

Received 31 July 2013; Revised 3 September 2013; Accepted 3 September 2013

Keywords: tags, image, tagging, semantic, Flickr, games

1.0 Introduction

A large quantity of social media data (text, audio, video, images, etc.) is uploaded to the web constantly, due to the popularity of digital photo cameras, mobile phones (with cameras) and social networks. Images used to be managed and categorised by librarians and archivists, amongst others. However, professional keyword assignment is too time-consuming to be used effectively on large image collections available on the web. Although a number of content-based image retrieval systems have been launched, they have not demonstrated sufficient utility on large-scale collections like the web. These systems are usually used as a supplement to existing context-based (or metadata-based) image retrieval systems using text, with additional functionality (e.g. search of similar images, search of specific colour scheme, etc).

The main aim of this work is to investigate whether an alternative, social tagging, can efficiently provide images with semantic descriptions, and how the social tagging behaviour can be managed. The work focuses on the following research questions:

- What are the facets of image tags in a popular photo-sharing social network?
- How do these tag facets change in a gaming environment? and,
- Can imposing restrictions on a game along with the provision of guidelines improve the semantic description of images?

To address these questions, a multi-faceted methodology was used. We tackle these research questions by analysing existing tagging behaviour on a popular photo sharing website (Flickr) and investigate the use of games with a purpose (GWAP), widely used to support image indexing. This work aims to provide a clearer picture of tagging-generation environments and their outcomes.

The paper is organized as follows. Related work and research context is presented in section 2. Section 3 describes our methodology based on a modified image attributes classification system. In section 4 we discuss the main results of applying the classification system and an experiment using Games With A Purpose (GWAP). This is followed by a discussion in section 5. Lastly, section 6 presents our conclusions.

2. Related work

2.1 Image retrieval systems

According to Ferecatu et al. (2008), the value of interpretive (defined below) and semantically rich keywords for image retrieval is undeniable. However, these keywords can-

not be derived automatically from image content, as there is a need for an association between content low-level features (defined below) and the high-level semantic concepts behind them. This kind of reasoning can only be done by a human either through professional description of images or through image tagging in various social applications.

Image retrieval systems can be broadly categorized into two main categories: context-based and content-based (Westman, 2009). Context-based (also known as metadata, piggy-back) text-based or concept-based) image retrieval systems use text to describe the image, whereas, content-based image retrieval (CBIR) systems employ visual features such as colour, shape, texture, object position for image description. Context-based image retrieval systems have been used since late 1970s, and are still the predominant method used for image search. They are known to be more efficient and accurate, and are based on assigning metadata to images. Content-based image retrieval (CBIR) is an alternative to a context-based approach, as it does not involve text to describe images. It focuses on low-level features (colour, texture, and shape) in an image. However, they are unable to retrieve high-level features such as subject and meaning, which are of primary importance in image search. The discrepancy between low-level visual features and high-level semantic concepts is often referred to as the problem of the semantic gap (Eakins and Graham 1999; Sawant et al. 2010).

2.2 The problem: the known semantic gap

Semantics, with respect to images, is an association between low-level features, such as shapes, colours, textures, and high-level concepts that could be presented by words (Sawant et al, 2010). Smeulders et al (2000) define the semantic gap as the “lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.” In other words, it is the difference between the way a human perceives the image and the actual image content. Hare et al. (2006) differentiate between “the gap between the descriptors and object labels” and “the gap between the labelled objects and the full semantics.” We can characterize the semantic gap in two ways therefore. The first gap (the one that lies between feature-vectors of the image and generic objects) is covered by CBIR algorithmic work, whereas the second gap (the one which is between object labelling and high-level reasoning) still needs human intellect as an essential component.

2.3 A solution: the social approach

A method by which web users could add their own searchable keywords to bookmarks, photos, videos, etc.

for future retrieval is known as social tagging (collaboratively creating and managing tags,) and these descriptive keywords are known as tags (Motive 2005). Several studies (Rorissa 2010; Rafferty and Hilderley 2007) focus solely on tagging behaviour within social networks like Flickr, while other means of contribution, such as crowdsourcing and social games with a purpose, have been less investigated.

2.3.1 Social tagging in photo-sharing networks

An online object can have multiple tags, and objects with the same tags can be grouped together, with the tags themselves being used to create a folksonomy (Gordon-Murnane 2006). The term folksonomy was coined in 2005 by information architect Thomas Vander Wal by combining the words “taxonomy” and “folk” (Dye 2006). Folksonomies can be of two types: the first is a broad folksonomy, which is created by assigning various tags to the same content by different users; the second type is called a narrow folksonomy, where users tag their own content for future retrieval and sharing (Dye 2006). Probably the best known example of a photo-sharing environment is Flickr. Tagging, comments and rating used in this and other systems have a huge impact on image description. Flickr predominantly addresses ‘findability’ within personal content (Dye 2006). Although Flickr is more about narrow folksonomy, where creation of metadata is the business of the person who posts the image, it also has social groups collecting tag specific photos. This is called “tagography.”

2.3.2 Games with a purpose (GWAP)

Sawant et al. (2010) state that along with photo-sharing services, collaborative gaming has significantly influenced the area of image retrieval and interpretation. While tagging in photo-sharing websites is known to be subjective and contains a lot of unidentified and misspelled words, guidelines could be designed to create social games for given tagging behaviour. GWAP or “games with a purpose” are computer games that are designed to use human’s cognitive abilities as a side effect of the playing process. They are used to get people involved in performing tasks that cannot be performed automatically. However, people usually play not because they want to solve “an instant computational problem” (von Ahn and Dabish 2008), but because they want to be entertained with a fast-paced and enjoyable game. The computation is just a side-effect of a game. Players are motivated to score as many points as possible within some time limit. A variety of online Games are available include image description games such as TagATune, PopVideo and ESP game.

3.0 Methodology

3.1 Overall Approach

Our work addresses image tagging habits and how we can specify and analyse the means of reaching a semantic description of an image through social tagging applications such as a photo-sharing network and gaming environments.

3.1.1 Tags in photo-sharing networks

Flickr is an online photo-sharing web site which was launched in 2004. It serves as an online storage with sharing facility. The application allows its users to annotate uploaded images with titles, descriptions or tags. Users could also set privacy settings both for visibility and for tagging and commenting activities. Flickr has already been used in previous research (e.g. Van Zwol et al. 2010). The application shows real-world use, storage and classification of images in contrast to laboratory-constructed experiments, and its images are not limited to particular subject domains.

For the evaluation of photo-sharing tags, 130 top tags and 500 random tags from the Flickr collection were selected. Most of the existing research in the area is based on randomly retrieved tags and queries. We inspected the most popular tags contained in the Flickr collection in the form of a tag cloud in order to show the overall trend. We then examined five hundred tags from the Flickr-based CoPhIR collection (Bolettieri 2009) which were randomly selected and analyzed in order to understand the nature of average tagging behaviour in a photo-sharing environment.

3.1.2 Tags in image-labelling games

The aim of the next theme in our research was to analyse the influence of collaboration and predefined tagging guidelines for conceptual tagging improvement. We use two games to do this. The first game (*Image-Labelling Game 1*) was designed based on a Google Image Labeller mechanism. It was used to analyze the default image-tagging behaviour during a game. The second game (*Image-Labelling Game with Guidelines 2*) was a modification of the first game through changing the rules by assigning each image a list of taboo/forbidden words to prevent players from describing an image with visual entities like colours (red, blue) and explicitly-presented objects (girl, house). This encouraged a more semantic-oriented approach in image description in comparison with the first game, and motivated players to tag images more conceptually (happiness, joy). These ‘taboo’ words were defined by the first author. As with the majority of existing image-labelling games, both our games are collaborative in nature (Goh et al. 2010b).

For each game two players were randomly chosen from all potential players. In each round, both players were given the same image as an input. Within a time limit players had to produce and match on as many descriptive keywords—tags—as possible based on the given image. For each match the players obtained 50 points and were notified of the outcome. The final score was a sum of match points, therefore players were rewarded for agreement on the number of tag matches with other players. They did not have to produce the tags at the same time. There was no “correct” tag. The main aim was for his/her partner to guess the same tag, helping to avoid biased image description. Although participants were co-located in the same lab, they did not know who their partners were, and direct communication among participants was prohibited.

This approach was used to cover the main characteristics of social image-tagging behaviour and to analyze the usefulness and success factors of social input for semantic image tagging. We used this approach to investigate the output in different social-based image environments and to provide an indication of how human knowledge can be used to bridge the semantic gap between image objects and high-level reasoning, which cannot be achieved automatically (see above).

3.2 Classification for tag analysis

In order to analyse tags, it is necessary to understand image attributes—features that can include visual, as well as spatial, semantic or emotional characteristics (Jorgensen 1996). There are many frameworks for classification of image attributes. Some of them are oriented towards indexing (Jaimés and Change 2000), some towards searching (Chung and Yoon 2011), and some combine both, concentrating on image descriptions which can be both search terms and indexing terms (Jorgensen 1996; Westman 2009).

For tag analysis we have chosen the following classification method. The coding of tags was done in two steps. First of all, tags were assigned to the following levels of image attributes: 1) metadata features, 2) primitive features, 3) visible general objects, and 4) semantic features (see table 1). Then, as the Level 4 (semantic tags) is of primary interest for this work, tags which fell into this category were analysed according to further facets: who, what, where and when. The coding system was initially tested on a sample set of descriptive words.

The chosen classification scheme is derived from the literature and corresponds to existing frameworks (Jaimés and Change 2000; Jorgensen 1996). Table 1 compares it to the classification used with Jorgensen’s framework. It contains levels of non-visual, visual and conceptual information. The main difference is that this classification consists of four levels, splitting visible objects from the interpreta-

Employed classification		Jorgensen	
Non-visual features (Meta-Data)		Art historical information	Interpretive
Primitive syntactic features, which include colours, shapes, textures, orientation and arrangement		Color Visual Elements Descriptions Location	
Visible objects/people in the image, as well as generic spatial features, which could be recognized by global colour analysis		Literal objects People	Perceptual
Semantic (conceptual) features involving interpretation of the meaning and purpose of the visual features	Who?	People qualities	
	What?	Content/Story Abstract concepts	
	Where?	Content/Story	
	When?	Content/Story	

Table 1. Comparison of our employed classification with Jorgensen’s framework

tions of visible objects (people vs. family or friends). It is similar to Jorgensen’s (1996) division of image attributes into perceptual and interpretive groups. The term “perceptual” refers to things in the image e.g. person, ship, beach, whereas the ‘interpretive’ term refers to a subjective view of what is happening in the image e.g. person laughing, having a good time, being sad, etc. This differentiation will help to evaluate the significance of interpretive attributes for image description in contrast to perceptually visible objects that could be indexed by automatic indexing algorithms. The derived image attributes’ levels are listed below:

Level 1 – non-visual metadata features: contain information about the author of the image, creation/upload date, photo camera characteristics, etc.

Level 2 – primitive syntactic features: are the basis for CBIR systems and include colours (yellow, green, hue, saturation, brightness), shapes (round, triangle) and textures (a texture of a tissue, bricks, orange peel).

Level 3 – visible objects/people on the image: are usually generic in nature (ball, chair, child).

Level 4 – semantic (conceptual) features: involve interpretation of the meaning and the purpose of the visual features (see below).

As the primary interest of this work is the influence of social tagging on bridging the semantic gap, Level 4 tags are analysed in more detail. Based on a combination of Enser et al. (2007) and Sawant et al.'s (2010) definitions of semantic levels, Level 4 tags are divided into four groups:

- Who: Who is portrayed on the image? The facet includes specific naming of people (John, Michael Jackson), general naming of professions (lawyer, businessman) and the naming of people's groups (family, couple, crew).
- What: What does the picture portray? The facet deals with visual semantic interpretation (gift, education, football, etc), aesthetical and emotive features (cute, sexy, happy, etc).
- When: When is the picture taken? This facet identifies person-specific (birthday), community-specific (New Year, Second World War), global events (swimming, skiing, cooking, etc.), time with no direct visual presence presented as natural values (night, autumn, etc.), artificial values (year, week, era), and specific values (1st January, 2011, 8:15 am, etc).
- Where: Where is the picture taken? This facet is associated with “geographically-grounded places” (London, Brazil, etc) and “non-grounded” entities (restaurant, museum, etc.) (Enser et al., 2007; Sawant, 2010).

4.0 Results

Here we describe the results of analysis of the photo-sharing network (Flickr) tags and the two games that followed.

4.1 Tagging behaviour in a photo-sharing network

The following analysis is based on information publicly available on Flickr. The information about popular tags was retrieved on 07/05/2011 from the Flickr tag cloud. After data analysis, the plural forms of nouns and spelling variations were eliminated leaving 134 tags for further classification. We aimed to analyze the current state of tagging behaviour on Flickr and the nature of tags based on the chosen classification scheme. Figure 1 shows the distribution of the top Flickr tags. Level 4 (semantic) tags remained the most popular (63.4%) and this percentage was considerably higher than those of subsequent categories. The next most popular category was Level 3 (visible objects) tags with 20.1%. Level 2 (primitive features) comprised 10.5% and Level 1 (metadata) comprised 6% of tags. Using Jorgensen's (1996) classification, the number of interpretive tags (69.4%) was considerably higher than that of perceptual tags (30.6%). These figures mean that tagging in a photo-sharing environment heavily

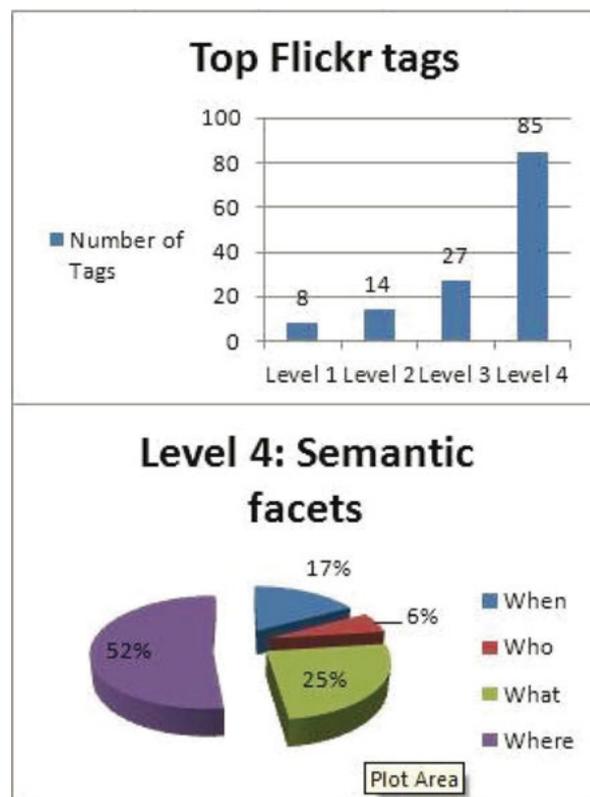


Figure 1. Top Flickr tags' distribution.

depends on human interpretative abilities and preliminary knowledge about the photograph subject and the history of creation. The location facet (52%) dominated among semantic tags. This could be explained by the fact that the majority of images on Flickr are people's vacation and travelling photographs and are tagged with visited geographical places. The “who” facet remains the least popular (6%) among tags, while “what” and “when” facets share the 2nd (25%) and the 3rd (17%) places respectively.

Along with the most popular tags, it was useful to analyse average tagging behaviour. Five hundred distinct tags were selected from the CoPhIR database for further analysis. The Flickr collection has many less ‘meaningful’ tags that could only be understood by people knowing the employed abbreviation or term. This was the reason behind creating a “Meaningless” category in the following analysis. Although there is functionality in Flickr to enter a phrase tag in a form of separate words, many users prefer to type in phrases as one word e.g., “summervacation.” For the purpose of tag content analysis phrases were treated separately. However, it should be noted that in practice, the majority of the genuine tags of this type will not support image retrieval with the usual queries. Another peculiarity of Flickr tags is the presence of a tag category naming Flickr group names. Although these tags are difficult to interpret, those that were found were inserted into the metadata class (Level 1). In order to preserve as much information as

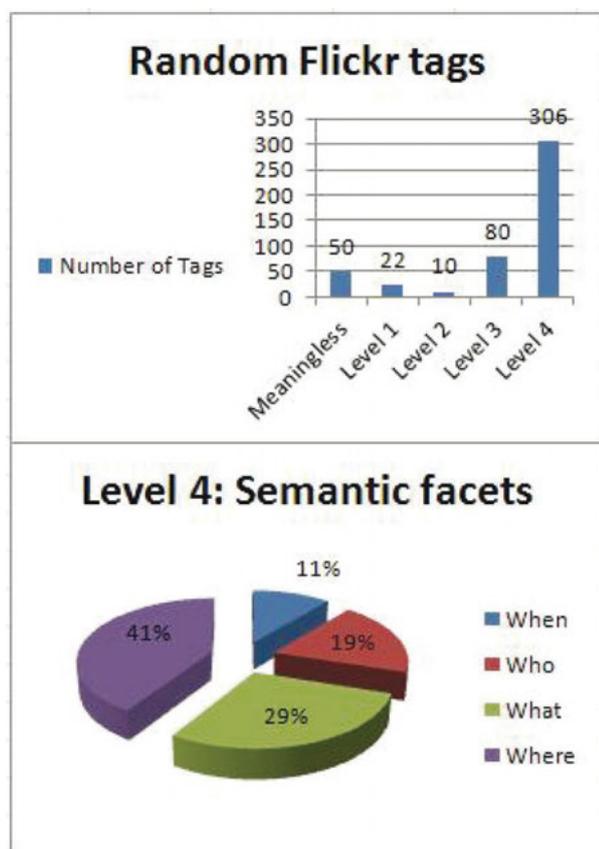


Figure 2. Random Flickr tags distribution

possible the non-English words (Spanish, French, etc) were translated with the help of Google Translator and the meaning of a number of words was checked with Wikipedia.

Tags were analyzed, and plural forms and spelling variations were reduced, leaving 468 tags for further analysis. Figure 2 shows the random Flickr tag distributions. About 11% of the remaining tags were coded as meaningless, including examples such as numbers (6, 17, 812), not generally-accepted abbreviations (co, kma, haas, etc), website names (httpwwwflickrcomphotosliyin), and symbols ($\frac{1}{2}$, $\frac{1}{2}$, ä, æµ). The majority (65.4%) of the rest of the tags fell into the Level 4 group. Using the Jorgensen categorization, perceptual tags (Level 2 and Level 3) were 19% of the total. The distribution of levels is similar to popular Flickr tag distribution, with the only difference being that Level 1 (metadata) tags are more popular in a random sample set compared with a top sample set. These Level 1 tags comprise about 5% of all tags and mostly include camera and lense information (fuji film pro 400h, 45mm), as well as names of the groups, creators, and genres (anime, self-portrait, etc). Level 2 (primitive features) tags were the least popular (2.1%) and were predominantly composed of colour names (amber, catchy colours, grey) and image orientation (landscape, portrait).

The next step included coding and analysis of Level 4 tags. The tags were analyzed without inspecting the image they were assigned to, and therefore a number of ambiguous and polysemous words were assigned to several semantic facets. Most of the tags (41%) represented the location facet, which is similar to the top Flickr tag distribution. In contrast to the most popular tags, random tags more often belonged to the “who” facet (19%). The reason behind this difference is the diversity of names. Although photos are quite often tagged with people’s names, none of these is widely used in the top tag set.

Interestingly, in contrast to a traditional filing system of image storage, where people tend to organize their collection chronologically, Flickr users are more location-oriented. However the second refinement in both systems is event information, which in the classification system employed was assigned to the ‘what’ facet and partially, if it was a community seasonal event like Christmas or Halloween, to the ‘when’ facet. It could be argued, that online systems like Flickr or Facebook provide easier access to tagging functionality for users. However, PC applications like Picasa also offer its users the functionality to identify people.

To conclude, it should be said that most of the user-assigned tags are by nature interpretive. In social networks and photo-sharing websites it is more evident, as the main purpose of these online communities is story-telling by means of pictures – hence the dominance of the interpretive category. This explains why images were described with information like place name and history, event and event participants.

4.2 Experimental gaming environment for image tagging

For each game similar sets of 20 images were selected from the CoPhIR image database, each of which contain information on the same concept. In each game 10 post-graduates with no or partial preliminary knowledge of the topic participated. Out of ten participants seven were female and three were male, all in the 20-39 years age range. Over half of the participants had an IT background. Other professional areas presented were law, finance, journalism and social science. Participants’ tagging experience mostly comes from the tagging of friends on Facebook photos; however, two participants were regular Flickr/Picasa users and one had no tagging experience at all. None of the participants had played online games on a regular basis. The majority had no or only a vague idea about games with a purpose. Each game was conducted for 20 minutes, collecting 590 and 342 tags for Game (1) and Game (2) respectively. After the first analysis stage, all duplications and spelling variations for each image were excluded, leaving 354 and 250 tags for further analysis.

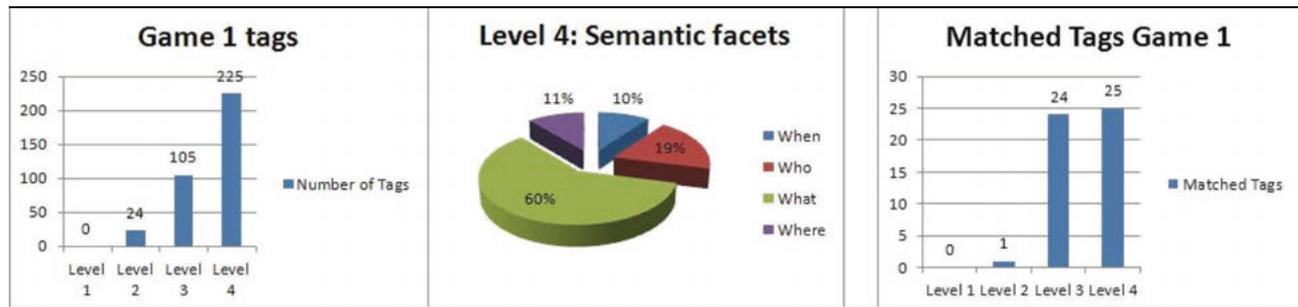


Figure 3. Game 1 tags' distribution

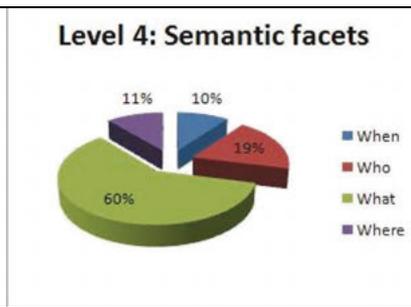


Figure 4. Game 1 matched tags

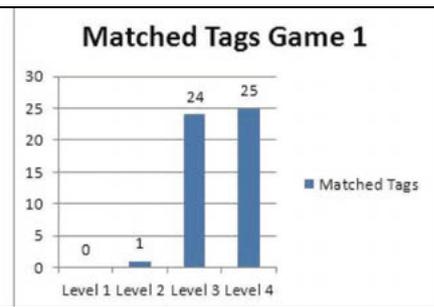


Figure 5. Game 2 tag distribution

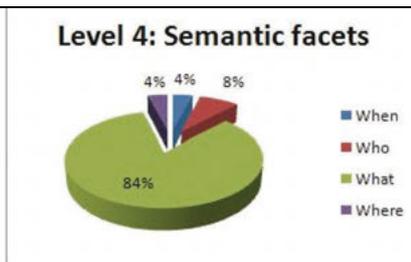
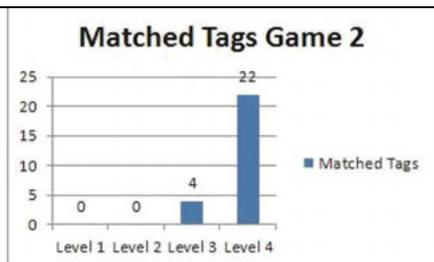


Figure 6. Game 2 matched tags



The facet analysis and distribution of these tags are shown in Figures 3-6.

4.2.1 Image-labelling game (1)

The main outcome of this collaborative game was that most of the tags were interpretive (63.6%); however, the percentage of perceptual descriptions (Level 2 and Level 3) was also quite high (36.4%). The majority of the interpretive tags included semantic interpretation of visual objects/scenes (football, kitchen, tombs, etc), aesthetic and emotive features (sadness, peace, cute, etc), and activities (cooking, sleeping, etc). The absence of metadata (Level 1) tags is explained by the lack of knowledge about the images' background information. Matched tags made up 14% of the game's outcome. The amount of perceptual and interpretive matched tags were spread equally. The majority of perceptual tags were general objects, scenes and people (man, umbrella, sky, etc). The distribution of semantic (Level 4) matched tags was similar to the distribution of all semantic Game1 tags, with a prevalence of the concept semantic facet (i.e. the 'what'), followed by person, location and time facets.

According to a number of studies, image-labelling games are recognised as a good source of image tags (von Ahn and Dabbish 2008). This study indicates that the game's outcome within an unrestricted game scenario has provided evidence for a balanced image description with general and interpretive words. However, due to the CBIR systems development and enhancement of object descrip-

tion algorithms, the need for object naming could be less important than image semantic interpretation which cannot be achieved through computer-based algorithms. Thus, in order to benefit from human input, there is a need for image tagging guidelines which prompt for more semantic, interpretive tagging.

4.2.2 Image-labelling game with guidelines (2)

The second collaborative game imposed restrictions on players, forbidding the use of words representing visual entities e.g. colour and explicitly-presented objects. The major outcome of this experiment was that the large majority of the tags (90.8%) were semantic interpretive words with a prevalence of "what" tags, with many fewer tags representing "who," "when" and "where." For example, one image was described with the following words: fans, victory, sport, cheering, team, happiness, support, fun, friendship, passion, game, exciting, pleasure, football. Matched tags made up 10.4% of the game's outcome, which is slightly less than in the first game (14%). The taboo word list reduced the number of matched perceptual words which made up only 15.4% whilst eliminating primitive feature (Level 2) tags—colours, shapes, etc. The majority of Level 4 tags are "what" concept words with "who" and "where" concepts used with much less frequency. The absence of "when" facet tags in a matched group could be mostly explained by spelling variations/errors/typos such as: "Halloween/Holloween/Halloween," etc.

5.0 Discussion

Although previous research e.g. (Rorissa 2010) showed that more perceptual attributes (colour, shapes, objects) were used for image descriptions of Flickr images, the results of this work show that tags can also be interpretive. Flickr users tend to assign specific names and geographical locations, as well as generally describe images by naming the general events and concepts presented. However, the number of tags for perceptual visual features tends to be lower than for conceptual features. These findings correspond with previous research of search image attributes e.g. (Chung and Yoon 2011), which found that semantic (conceptual) category of image attributes is the most popular among users' queries.

On this evidence games with a purpose (GWAP) are a useful application for image tagging, and could be used for various purposes depending on the game's rules and winning conditions. Within unconditional gaming environments, players tend to use a balance of perceptual and interpretive image attributes. However, the limitation on words that could be used for tagging may stimulate players' interpretive descriptions. This helps to beneficially employ human abilities – without having duplicate data that can be extracted by CBIR or automatic indexing systems. According to the results of this study, the variety of social tagging applications could satisfactorily generate semantic descriptions of images. Although photo-sharing networks support more balance in terms of semantic facets tagging, games with a purpose can be used to augment the tagging process. However the design of the game needs to be very clearly thought out (Goh and Lee 2011) and there is some evidence that tagging images normally may outperform either collaborative or competitive games. Different types of noise (Wang et al. 2012) may be generated than with standard tagging (e.g. bias of the participants). Goh et al. (2010a) provide some evidence which conflicts with the earlier study i.e. that competitive games produced the best result. Designers therefore need to be clear about how to engage players and reward them for providing high quality tags in order to obtain the best possible outcome.

6.0 Conclusion and future work

The aim of this exploratory analysis was to examine the value of social tagging for image description by investigating facets of tags in two different social-based tagging applications: a photo-sharing social network and an image-labelling game-based experiment. The tags were coded and evaluated according to a classification of image attributes based on a combination of established image attribute frameworks.

The results of the research showed that social tagging is predominantly an interpretive activity. However, the number of perceptual tags depends on the context of image use. Photo-sharing communities mostly use images for story-telling and/or as an event diary; therefore, there is more semantic information associated with images with a prominent amount of people and location recognition, and event and activities tags. The gaming application has shown to be slightly more perceptual oriented, as visual features (colours, shapes, and distinct objects) are easier to spot and to match. However, specific guidelines can influence the game's outcome in order to obtain a given result (or more particular types of tags). This shows that social tagging is a manageable process, but this does to some extent depend on the taggers' understanding of the image use and on the nature of the tagging environment. It is also seen from the study that games are more oriented towards describing 'what' in an image, while photo-sharing social networks present a more balanced picture of semantic facets (what/where/when/who). It would be useful to analyse whether person, place and time recognition is needed and achievable through GWAP.

Whilst our framework has been useful for the research carried out here, work on how we can use the various levels in conjunction with CBIR to improve image retrieval is worthwhile. Given the results presented here, it would be worth concentrating on interpretive tags initially in order to see what leverage can be gained from that part of the classification.

References

- Bolettieri, Paolo, Esuli, Andrea, Falchi, Fabrizio, Lucchese, Claudio, Perego, Raffaele, Piccioli, Tommaso and Rabbitti, Fausto. 2009. CoPhIR: A test collection for content-based image retrieval. *CoRR abs/0905.4627*. Available <http://arxiv.org/pdf/0905.4627v2.pdf>
- Chung, Eunkyung and Yoon, Jungwon. 2011. Image needs in context of image use: An exploratory study. *Journal of information science* 37: 163-77.
- Dye, Jessica. 2006. Folksonomy: A game of high-tech (and high-stakes) tag: Should a robot dictate the terms of your search? In an age when whole lives are lived online-via blogs, picture albums, dating, shopping lists-digital content users are not only creating their content, they're building their own infrastructure for making it easier to find.(navigating webs). *Econtent magazine* April. Available <http://www.econtentmag.com/Articles/Editorial/Feature/Folksonomy-A-Game-of-High-tech-%28and-High-stakes%29-Tag-15298.htm>.
- Eakins, John and Graham, Margaret. 1999. *Content-based image retrieval*. Available www.jisc.ac.uk/uploaded_documents/jtap-039.doc

- Enser, Peter G.B., Sandom, Christine J., Hare, Jonathan S. and Lewis, Paul H. 2007. Facing the reality of semantic image retrieval. *Journal of documentation* 63: 465-81.
- Ferecatu, Marin, Boujemaa, Nozha and Crucianu, Michel. 2008. Semantic interactive image retrieval combining visual and conceptual content description. *Multimedia systems* 13: 309-22.
- Goh, Dion. H. and Lee, Chei S. 2011. Perceptions, quality and motivational needs in image tagging human computational games. *Journal of information Science* 37: 515-31.
- Goh, Dion H., Ang, Rebecca P., Chua, Alton and Lee, Chei S. 2010a. Evaluating game genres for tagging images. In Blandford, Ann and Gulliksen, Jan, eds., *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10)*. New York: ACM Press, pp. 659-62.
- Goh, Dion H., Ang, Rebecca P., Lee, Chei S. and Chua, Alton. 2010b. Fight or unite: Investigating game genres for image tagging. *Journal of the American Society for Information Science and Technology* 62: 1311-24.
- Gordon-Murnane, Laura. 2006. Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher: the magazine for database professionals* 14: 26-38.
- Hare, Jonathan S., Lewis, Paul H., Enser, Peter G. B. and Sandom, Christine J. 2006. Mind the gap: Another look at the problem of the semantic gap in image retrieval. In Chang, Edward Y, Hanjalic, Alan and Sebe, Nicu, eds., *Multimedia content analysis, management, and retrieval 2006 : 17-19 January 2006, San Jose, California, USA*. Bellingham, Wash.: SPIE.
- Jaimes, Alejandro and Chang, Shih F. 2000. A conceptual framework for indexing visual information at multiple levels. In Beretta, Giordano B. and Schettini, Raimondo, eds., *Proceedings of IS&T/SPIE Internet Imaging 3964*, pp. 2-15.
- Jorgensen, Corinne. 1996. Indexing images: Testing an image description template. In *ASIS 1996 Annual Conference Proceedings, Baltimore, MD, October 19-24, 1996*, pp.209-13.
- Rafferty, Pauline and Hilderley, Rob. 2007. Flickr and democratic indexing: Dialogic approaches to indexing. *Aslib proceedings: New information perspectives* 59: 397-410.
- Rorissa, Abebe. 2010. A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology* 61: 2230-42.
- Sawant, Neela, Lee, Jia and Wang, James. 2010. Automatic image semantic interpretation using social action and tagging data. *Multimedia tools and applications* 51: 213-46.
- Smeulders, Arnold W.M., Worring, Marcel, Santini, Simone, Gupta, Amarnath and Jain, Ramesh. 2000. Content-based image retrieval at the end of the early years. *IEEE Transactions on pattern analysis and machine intelligence* 22: 1349-80.
- Van Zwol, Roelef, Sigurbjornsson, Börkur, Adapla, Ramu., Pueyo, Lluís G., Katiyah, Abhinav, Kurapata, Kaushal, Muralidharan, Mridul, Muthu, Sudar, Murdock, Vanessa, Ng, Polly, Ramani, Anand., Sahai, Anuj, Sahai, Anuj, Sathish, Sriram T., Vasudev, Hari and Vuyyuru, Upen-dra. 2010. Faceted exploration of image search results. In Freire, Juliana and Chakrabarti, Soumen, eds., *Proceedings of International Conference of the World Wide Web (WWW'10)*. New York: ACM Press, pp. 961-70.
- von Ahn, Luis and Dabbish, Laura, 2008. Designing games with a purpose. *Communications of the ACM* 51: 58-67.
- Wang, Meng, Ni, Bingbing, Hua, Xian-Sheng and Chua, Tat-Seng. 2012. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM computing surveys* 44: 25:1-24.
- Westman, Stina. 2009. Image users' needs and searching behaviour. In Göker, Ayşe and Davies, John, eds., *Information retrieval: Searching in the 21st century*. Chichester: Wiley.