Social tagging: Exploring the image, the tags, and the game

Elena Konkova[1], Ayşe Göker[2], Richard Butterworth[1] and Andrew MacFarlane[1]

[1]Centre for Interactive Systems Research, City University London, Northampton Square, London EC1V 0HB {Konkova.elena87@yandex.ru, richard@richardbutterworth.co.uk, andym@city.ac.uk}

[2] School of Computing Science and Digital Media, Robert Gordon University, St Andrew St, Aberdeen, AB25 1HG {a.s.goker@rgu.ac.uk}

ABSTRACT

An increasing amount of images are being uploaded, shared, and retrieved on the Web. These large image collections need to be properly stored, organized and easily retrieved. Tags have a key role in image retrieval but it is difficult for those who upload the images to also undertake the quality tag assignment for potential future retrieval by others. Relying on professional keyword assignment is not a practical option for large image collections due to resource constraints. Although a number of content-based image retrieval systems have been launched, they 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. An alternative to professional image indexing can be social tagging -- with two major types being photo-sharing networks and image labeling games. Here we analyze these applications to evaluate their usefulness from the semantic point of view. We also investigate whether social tagging behaviour can be managed. 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.

1. INTRODUCTION

A large quantity of social media data (text, audio, video, images, etc) is uploaded to the web constantly. With the popularity of digital photo cameras and mobiles (with cameras), the reduction in cost for image storage and editing, and the popularity of social networks, the Web now abounds with images differing in quality, context and target audience. People upload, browse, share and comment on thousands of images every day. Images are searched on the Web, purchased from stock libraries, and shared on photo-sharing websites. Moreover, photo sharing is known to be the leading activity on social networks (Universal McCann, 2008). These large image collections need to be properly stored, organized and easily retrieved.

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). An alternative for professional image indexing is claimed to be social tagging, which emerged around five years ago together with the Web 2.0 era.

The main aim of this work is to investigate whether 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: (1) What are the facets of image tags in a popular photo-sharing social network? (2) How do these tag facets change in a gaming environment? and (3) 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.

First of all, the analysis of existing tagging behaviour provided us with information about facets of popular image attributes used for image description. The work subsequently also covers a new trend in crowdsourcing using Games With A Purpose (GWAP), which is widely used to support image indexing. Two types of games were created to evaluate the influence of collaboration on image tagging, viz the unrestricted and guided gaming environments. 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 and plans for future work.

2. RELATED WORK

2.1 Image Retrieval Systems

According to Ferecatu et al (2008), the value of interpretative and semantically rich keywords for image retrieval is undeniable. However, these keywords cannot 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. The metadata could be title, natural language description, author, date and time of creation, and assigned keywords (either with the help of controlled vocabulary, professional natural language description, or through social tagging). For web image search engines words in the anchor text of a link, filename, etc. could also be additional contextual information (Westman, 2009). Rui, et al (1999) have outlined several main difficulties with context-based image retrieval. These systems are time- and labor-consuming and subjective (as the same image may be perceived differently depending on place, time and purpose of its use).

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. Hence, they tend to be used more in retrieving subsets of specific visual attributes in domain specific systems, in experimental projects or as an extra feature of existing context-based retrieval systems. The discrepancy between low-level visual features and high-level semantic concepts is often referred to as the problem of the semantic gap (Sawant et al, 2010; Eakins and Graham, 1999).

Chu (2010) assumes that the integration of context-based and content-based approaches "seems to be an ideal road to take in representing multimedia", as keywords and tags can capture the semantic content of images, whereas image attributes like colour and texture, which are hard to name, could be recognized by CBIR. In her extensive survey of image

users' needs and search behaviour, Westman (2009) cites Eakins and Graham (1999) who claim that text-based search mostly operates with semantic terms, whereas syntactic attributes (colour, shape, texture, etc.) are selected from a menu (e.g. drop down).

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". Even if it is possible to label all the objects on the image it does not guarantee that the semantics will be captured, as semantics is more about relationships between objects, relationship with the world at large and some broader context. As Enser et al (2007) concluded, bridging the semantic gap has drawn the attention of a lot of researchers in the image retrieval community. We can characterize the semantic gap in two ways. 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

The main human-based alternative for traditional indexing is social input. Sawant et al (2010) have identified a number of challenges that they stated could be potentially addressed by tagging using a social approach. These include motivation and therefore tagging outcome, cultural differences, tag spamming and specialized knowledge of different user groups that

could cause problems in interpreting tags like “d50” and “hauptstadt”, which are meaningless to a global audience. Although tagging is thought to be subjective, in fact collaborative work helps to alleviate the problem, revealing the ‘wisdom of the crowd’ and with the potential to improve the metadata quality of photos over time.

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,* and these descriptive keywords are known as tags (Motive, 2005). Several works (Rorissa, 2010; Rafferty and Hidderley, 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. Crowdsourcing has proved itself to be “a reasonable substitute for repetitive expert annotations” (Sawant et al, 2010). Recent crowdsourcing systems like LabelMe and Amazon Mechanical Turk Internet enlist the user for image labeling tasks. People are provided with detailed instructions about a particular task and are given a small cash reward in return for satisfactory completion. The tasks are usually split into smaller units to encourage people to do as many tasks as possible. Nevertheless, the two major sources of image annotations are considered collaborative image labelling games (Games With A Purpose - GWAP) and tagging communities in social networks.

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 commonly take the form of a tag cloud, where the size of each tag depicts the frequency of the word in the system. 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). Cattuto et al (2008) investigate the properties of such tagging systems by focusing on one particular social bookmarking system del.icio.us.

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". Social tagging is also used in other applications such as museum collections (Trant and Wyman, 2006).

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 Dabbish, 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. They are usually paired randomly in order to prevent cheating and increase the quality of the game results. GWAP has been used in various applications including affect for a database of messages (Pearl and Steyvers, 2010), to produce domain specific sentiment lexicons (Weichselbraun et al, 2011), detecting

passivized intransitive verbs in Turkish sentences (Gencer et al, 2012), and building ontologies for the semantic web (Siorpaes and Hepp, 2008).

There is also a variety of games for eliciting annotations (in the form of tags), for example:

- Games for video annotation: PopVideo (GWAP, N.D.), OntoTube (OntoGame, N.D.).
- Games for image annotation: ESP Game and Matchin (GWAP, N.D.).
- Games for audio annotation: TagATune (GWAP, N.D.), Herdit (FaceBook, N.D.).
- Games for text annotation: Verbosity GWAP, N.D.), Phrase Detectives (N.D.)

According to Goh et al (2010b), human computation games are mostly *collaborative* in nature. Players *cooperate* in order to score points. However, there is another recent *competitive* type (Ho et al, 2009), where participants play against each other. It is mostly used to address quality issues in collaborative games. Mobile games are usually used for location-based annotation and are competitive by their nature. Examples are PhotoCity (N.D.), Eyespy (Bell et al, 2009), Gopher Game (Casey et al, 2007), MobiMissions Grant et al, 2007), CityExplorer (Matyas et al, 2008) and Indagator (Lee et al, 2010). Web-based applications are mostly designed for keywords assignment or similarity judgments. According to von Ahn and Dabbish (2008) collaborative games can be put into three categories: *output agreement* (player attempts to generate the same output based on the common input), *input agreement* (players have to decide whether they have the same input or not based on the independently generated descriptions of each other), and *inversion problem* ("Describer" – "Guesser" principal) games.

3. METHODOLOGY

3.1 Overall Approach

Human intervention is still required for effective image retrieval, despite the advances in CBIR. Automatic tagging which relies on extraction techniques alone is not sufficiently reliable in multimedia generally (Wang et al, 2012). User-provided tags are usually noisy and incomplete (Wang et al, 2012), and some kind of quality control is therefore desirable. 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 a gaming environment.

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. It also allows 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 a number of previous researches (e.g. Van Zwol and Sigurbjornsson, 2010). It 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 analyzed 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. The quality of GWAP results is usually evaluated based on the descriptiveness and usefulness of the tags for future retrieval. There are a number of existing papers about GWAP design, implementation and evaluation mostly for one particular application (Ho et al, 2009; Lee et al 2010; von Ahn et all, 2006; Bell et al, 2009; Šimko, J and Bieliková, 2012). There are also a number of comparative studies (von Ahn and Dabbish, 2004; 2008). However, they are mostly oriented towards game design and purpose, rather than being an evaluation of actual game outcomes. The only article which is similar to this research is a comparative study of Goh et al (2010b) across three different types of tagging application evaluating quality of computations and user perceptions. As there are neither customisable games that could be used for research purposes nor available data about the outcome of existing image tagging games, it was decided to design and develop two types of games for further game-based tag analysis. The purpose of the games would be to assign tags to a selected image set from CoPhiR through playing regular and restricted image-labelling games. Both games are collaborative and the output is a set of agreed tags.

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 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 to think like his/her partner and enter the same tag, which helps 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 (Jaimes 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). Secondly, 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.

<table>
<tr><th colspan="2">Employed classification</th><th colspan="2">Jorgensen</th></tr>
<tr><td colspan="2">Non-visual features</td><td>Art historical information</td><td>Interpretive</td></tr>
<tr><td colspan="2">Primitive syntactic features, which include colours, shapes, textures, orientation and arrangement</td><td>Color
Visual Elements
Descriptions
Location</td><td rowspan="2">Perceptual</td></tr>
<tr><td colspan="2">Visible objects/people in the image, as well as generic spatial features, which could be recognized by global colour analysis</td><td>Literal objects
People</td></tr>
<tr><td rowspan="4">Semantic (conceptual) features involving interpretation of the meaning and purpose of the visual features</td><td>Who?</td><td>People qualities</td><td rowspan="4">Interpretive</td></tr>
<tr><td>What?</td><td>Content/Story
Abstract concepts</td></tr>
<tr><td>Where?</td><td>Content/Story</td></tr>
<tr><td>When?</td><td>Content/Story</td></tr>
</table>

Table 1 - Comparison of our employed classification with Jorgensen's framework

The chosen classification scheme is derived from the literature and corresponds to existing frameworks (Jaimes 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 interpretations 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 RESULTS

Here we describe the results of analysis of the photo-sharing network (Flickr) tags and the two games that followed. The Flickr tags showed that people used a considerable number of semantic features (Level 4) without any prompting. These tended to focus around the context of the individual. Differences between social tagging and GWAP results are clear. Under a collaborative game scenario, users displayed a balanced use of perceptual and interpretive tags. When some restrictions were added to the game along with guidelines, the potential to increase the number of interpretive tags was shown.

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 (N.D.). After data

analysis the plural forms of nouns and spelling variations were eliminated leaving 134 tags for further classification.

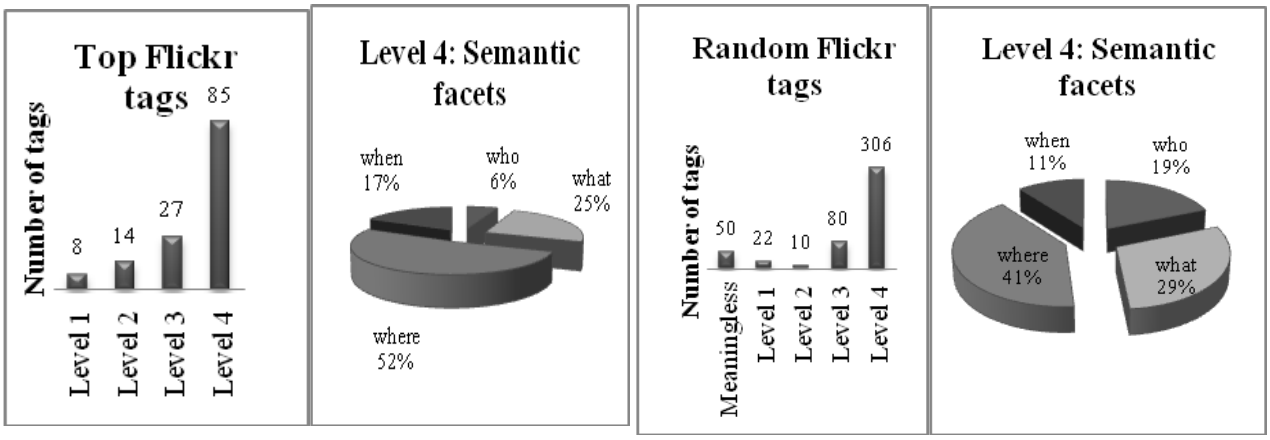

Figure 1. Top Flickr tags' distribution.

Figure 2. Random Flickr tags' distribution.

This research 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 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 (*summervacation*). For the purpose of tag content analysis solid words were disjoint. 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 possible the non-English words (Spanish, French, etc) were translated with the help of Google Translator (N.D.) and the meaning of a number of words was checked with Wikipedia (N.D).

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 (½ï¿½, 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 postgraduates 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 on-line 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. The facet analysis and distribution of these tags are shown in Figures 3-6.

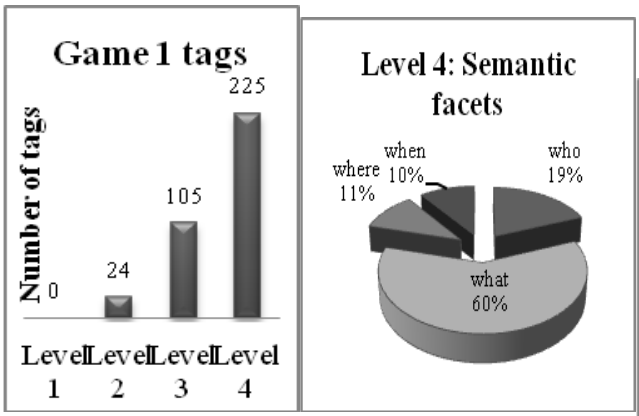

Figure 3. Game 1 tags' distribution.

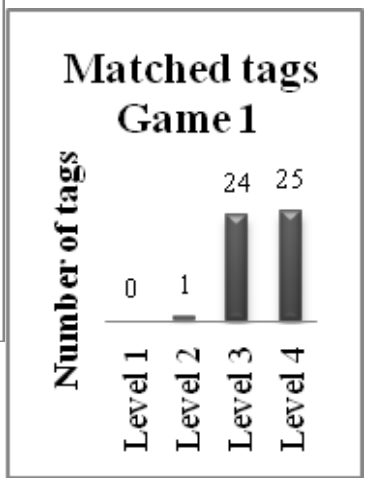

Figure 4. Game 1 matched tags.

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.

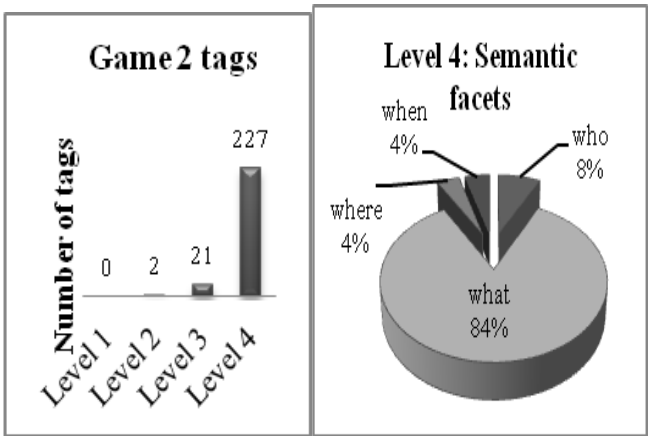

Figure 5**. Game 2 tag distribution.**

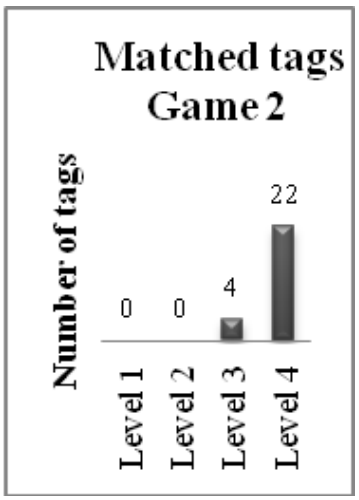

Figure 6**. Game 2 matched tags**

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. 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 description 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.

Figure 7: An example of an image within a collaborative game with restrictions: Stop words were *two, woman, red, blue, white, women*.

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 much 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.* Figure 7 shows an image along with the restrictions/stop words that were applied. 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 *Halloween/Holloween/Hallowen* etc.

5. 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. 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. 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.

There is also a need for future research into contextual image-labeling games, to provide players with some context for image tagging (e.g. further use of images in advertising) and thus improve the quality of tags. This could be achieved by adjusting the rules of the game or through the change of game genre to role-playing or action, which has not been explored yet in social tagging (Goh et al, 2011). Extending others' work, we devised a framework for classification of image attributes and in particular expanded on the semantic level of attributes for both analysis and targeting tag generation in these facets.

The experimental part of the research could be repeated with a larger number of input images and participants. Replication of the study with more diverse groups of participants (e.g. age ranges, educational and professional backgrounds) would be useful for better understanding of tagging trends in GWAPs. Moreover, other game types could be tested in terms of output analysis. The games could be released on the web, thus increasing the number of potential participants and providing researchers with an opportunity to use crowdsourcing systems for selection of participants (e.g. people with initial tagging experience). Moreover, adding an auto-correct word function could enhance the number of matched tags by reducing misspellings and typing errors.

REFERENCES

Bell, Marek., Reeves, Stuart., Brown, Barry., Sherwood, Scott., MacMillan, Donny., Ferguson, John., Chalmers, Matthew. 2009. Eyespy: Supporting Navigation through Play. In Hinchley, Ken., Ringel Morris, Meredith., Hudson, Scott. and Greenberg, Saul. eds., CHI '09: *Proceedings of the 27th international conference on Human factors in computing systems,* ACM Press. pp. 123-132.

Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., Rabitti. F. 2009. CoPhIR: a Test Collection for Content-Based Image Retrieval. CoRR abs/0905.4627.

Casey, Sean., Kirman, Ben., Rowland, Duncan., 2007. The gopher game: a social, mobile, locative game with user generated content and peer review. In: Bernhupt, Regina and Natkin, Stephane. eds., *Proceedings of the international conference on Advances in computer entertainment technology* (ACE'07), ACM Press pp. 9-16.

Chu, Heting. 2010. *Information representation and retrieval in the digital age* (2nd ed.). Medford, N.J.: Information Today.

Chung, Eunkyung. and Yoon, Jungwon. 2011. Image needs in context of image use: An exploratory study. *Journal of Information Science*, 37: 163-177.

Cattuto, Ciro., Benz, Dominik., Hotho, Andreas. and Stumme, Gerd. 2008. Semantic grounding of tag relatedness in social bookmarking systems. In Sheth, Amit., Staab, Steffen., Dean, Mike., Paolucci, Massimo., Maynard, Diana., Finin, Timothy and Thirunarayan, Krishnaprasad. eds., *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*, Springer Berlin Heidelberg, pp. 615-631, .

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. Information Today, HighBeam Research.* Available http://www.highbeam.com/doc/1G1-143628543.html.

Eakins, John and Graham, Margaret. 1999. Content-based Image Retrieval. *The JISC Technology Applications Programme*, Report 39.

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: 465-481.

FaceBook Herdit web site. (N.D.). Available http://apps.facebook.com/herd-it/.

Ferecatu, Marin., Boujemaa, Nozha. and Crucianu, Michel., 2008. Semantic interactive image retrieval combining visual and conceptual content description. *Multimedia Systems*, 13: 309-322.

Flickr Tag Cloud (N.D.). Available http://www.flickr.com/photos/tags/.

Gencer, Adem E., Gungor, Tunga., Gurer, Asli. Ozsoy, A.Sumru., 2012. Input-evaluation: A new mechanism for collecting data using games with a purpose, *IEEE Symposium on Computers and Communications (ISCC)*, IEEE, pp. 239-244,.

Goh, Dion. H. and Lee, Chei.Sian. 2011. Perceptions, quality and motivational needs in image tagging human computational games. *Journal of information Science*, 37: 515-531.

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, *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries (NordiCHI '10).* ACM Press, pp. 659-662.

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: 1311-1324.

Gordon-Murnane, Laura. 2006. Social bookmarking, folksonomies, and Web 2.0 tools. *Searcher Mag Database Prof*, 14: 26-38.

Grant, Lyndsay., Daanen, Hans., Benford, Steve., Hampshire, Alastair., Drozd, Adam. and Greenhalgh, Chris. 2007. MobiMissions: the game of missions for mobile phones. In Swanson, Janese. ed., Proceedings of ACM SIGGRAPH 2007, ACM Press, Article 12.

Google translate (N.D.). Available http://translate.google.com/.

GWAP website (N.D.). Available http://www.gwap.com/gwap/.

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., Sebe, Nicu. eds., *Multimedia Content Analysis, Management, and Retrieval 2006. Volume 6073*. San Jose, California, USA, SPIE (2006) 607309–1–607309–1

Ho, Chien-Ju., Chang, Tao-Hsuan., Lee, Jong-Chuan., Hsu, Jane Y., Chen, Kuan-Ta.,2009. KissKissBan: A Competitive Human Computation Game for Image Annotation. In Bennett, Paul., Chandrasekar, Raman., Chickering, Max., Ipeirostis, P., Law, Edith., Mityagin, A., Provost, F. and von Ahn, L. *HCOMP'09L Proceedings of the ACM SIGKDD Workshop on Human Computation*, ACM Press, pp. 11-14

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*, Vol. 3964, pp. 2-15.

Jorgensen, Corinne. 1996. Indexing Images: Testing an Image Description Template. *ASIS 1996 Annual Conference Proceedings, Baltimore, MD, October 19-24, 1996*, pp.209-213.

Lee, Chei S., Goh, Dion H., Chua, Alton, and Ang, Rebecca P. ,2010. Indagator: Investigating perceived gratifications of an application that blends mobile content sharing with Gameplay. *Journal of the American Society for Information Science and Technology*, 61: 1244-1257.

Matyas, Sebastian., Matyas, Christian., Schlieder, Christoph., Kiefer, Peter, 2008. CityExplorer - A Geogame Extending the Magic Circle. In: *Informatik 2008, Workshop on Mobile Gaming, GI-LNI 133*, pp. 503-504.

Motive Web Design Glossary, 2005. Personalised classification. Available http://www.motive.co.nz/glossary/folksonomy.php

OntoGame web site (N.D.). Available http://ontogame.sti2.at:8080/ontogame/

Rafferty, Pauline. and Hidderley, Rob. 2007. Flickr and Democratic Indexing: dialogic approaches to indexing. *Aslib Proceedings: New Information Perspectives*, 59: 397-410.

Pearl, Lisa. and Steyvers., Mark. 2010. Identifying emotions, intentions, and attitudes in text using a game with a purpose. In Inkpen, Diana and Strapparava, Carlo. eds., *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text (CAAGET '10).* Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 71-79.

Phrase Detective website (N.D.). Available http://anawiki.essex.ac.uk/phrasedetectives/register.php.

PhotoCity website (N.D.). Available http://photocitygame.com/about.php.

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: 2230–2242.

Rui, Yong., Huang, Thomas. and Chang, Shih F., 1999. Image retrieval: Current Techniques, Promising Directions, and Open Issues. *Journal of Visual Communication and Image Representation*, 10: 39-62.

Sawant, Neela., Lee, Jia. and Wang, James., 2010. Automatic Image Semantic Interpretation using Social Action and Tagging Data. *Multimedia Tools and Applications*, 51: 213-246.

Šimko, Jakub and Bieliková, Mária. 2012. Personal image tagging: a game-based approach. In Sack, Harald. and Pellegrini, Tassilo. eds., *Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12),* ACM, Press, pp 88-93.

Siorpaes, Katherina., and Hepp, Martin. 2008. Games with a Purpose for the Semantic Web, *Intelligent Systems, IEEE*, 23: 50,60.

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 intelligenc*e, 22: 1349-1380.

Trant, Jeniffer., and Wyman, Bruce. 2006. Investigating social tagging and folksonomy in art museums with steve. museum. In *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland*. Available http://wwwconference.org/proceedings/www2006/www.rawsugar.com/www2006/4.pdf.

Universal McCann's global research (2008), Wave 3. Available http://www.scribd.com/doc/3836535/Universal-Mccann-on-Social-Media.

Van Zwol, Roelef. Sigurbjornsson, Börkur., Adapla, Ramu, Pueyo, Lluis 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, Upendra. 2010. Faceted exploration of image search results. In Freire, Juliana, Chakrabarti, Soumen. eds., *Proceedings of International Conference of the World Wide Web (WWW'10),* ACM Press, 961-970.

von Ahn, Luis. and Dabbish, Laura. , 2004. Labeling images with a computer game. In Dykstra-Erickson, Elizabeth and Tscheligi, Manfred. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press pp. 319–326.

von Ahn, Luis. and Dabbish, Laura, 2008. Designing Games With A Purpose. *Communications of the ACM*, 51: 58-67.

Von Ahn, Luis., Liu, Rouran. and Blum, Manuel. 2006. Peekaboom: A Game for Locating Objects in Images. In Ginter, Rebecca., Rodden, Thomas., Aoki, Paul., Cutrell, Ed., Jeffries, Robin and Olson, Gary *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM Press, pp. 55-64.

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: article 25.

Westman, Stina. 2009. Image Users' Needs and Searching Behaviour. In: Goker, A., Davies, J., ed. 2009. *Information retrieval: searching in the 21st century*. Chichester : Wiley.

Weichselbraun, Albert., Gindl, Stefan. and Scharl., Arno. 2011. Using games with a purpose and bootstrapping to create domain-specific sentiment lexicons. In Berendt, Bettina., de Vries, Arjen ., Fan, Wenfei., Macdonald, Craig., Ounis, Iadh and Ruthven, Ian. eds., *Proceedings of the 20th ACM international conference on Information and knowledge management (CIKM '11* ACM, New York, NY, USA, pp.1053-1060.

Wikipedia web site. (N.D.). Available http://en.wikipedia.org/wiki/Main_Page