Automatic Tag Extraction from Social Media for Visual Labeling

Shuhua Liu and Thomas Forss
Arcada University of Applied Sciences, Jan-Magnus Janssonin aukio 1, 00560 Helsinki, Finland

Keywords: Visual Labeling, Image Annotation, Social Media Analysis, Twitter.

Abstract: Visual labeling or automated visual annotation is of great importance to the efficient access and management of multimedia content. Many methods and techniques have been proposed for image annotation in the last decade and they have shown reasonable performance on standard datasets. Great progress has been made especially in recent couple of years with the development of deep learning models for image content analysis and extraction of content-based concept labels. However, concept objects labels are much more friendly to machine than to users. We consider that more relevant and user-friendly visual labels need to include "context" descriptors. In this study we explore the possibilities to leverage social media content as a resource for visual labeling. We developed a tag extraction system that applies heuristic rules and term weighting method to extract image tags from associated Tweet. The system retrieves tweet-image pairs from public Twitter accounts, analyzes the Tweet, and generates labels for the images. We elaborate on different visual labeling methods, tag analysis and tag refinement methods.

1 INTRODUCTION

Visual labels are the primary component of large searchable multimedia collections. With the fast spread of user-generated content in social media and on the cloud, we have seen explosive growth of image and video data - a huge part of which have no labels at all and thus hard to be accessed with user-friendly text queries. Automatic visual labeling is thus an essential tool for getting access to a great amount of multimedia content.

Automatic visual labeling is a very challenging task that has attracted great attention and immense interest of machine learning researchers in the field of computer vision (Makadia et al, 2008). Image and video annotation has been a topic of on-going research for very long time and many techniques have been developed, which can be classified into three general approaches: (1) to generate candidate labels through image content analysis and visual object recognition; (2) to formulate candidate descriptors through analysis of textual context information of visual content; and (3) to explore the relationships between user queries and images as well as similarities between images, to treat image annotation as a retrieval problem (Makadia et al, 2008; Sun et al, 2011).

Image annotation research has demonstrated success on test data for focused domains. Unfortunately, extending these techniques to the broader topics found in real world data often results in poor performance (Liu et al, 2009). In recent few years, great progress has been made fast and very impressive results achieved in "content-based labeling", where candidate-labels are generated through image content analysis using deep learning methods (Chen et al, 2013; Sjöberg et al, 2013a, 2013b). However, formal concept and object labels from visual content analysis are much more machine-friendly than user-friendly. There is the need to explore more relevant and user-friendly visual labels that include "context" descriptors.

Context-based labels can be personal or social or domain and topic dependent. Our research aims to leverage the potential of social media resources for the benefit of visual annotation, to make use of social context information in facilitating image organization, access and utilization. Given an input image, our goal of automatic image annotation is to assign a few relevant text keywords to the image that reflect not only its visual content but also its social context.

In such a setting, we have developed a tag extraction system that applies heuristic rules and term weighting method to extract image tags from

Liu, S. and Forss, T..
Automatic Tag Extraction from Social Media for Visual Labeling.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 504-510
ISBN: 978-989-758-158-8

associated Tweet. The system retrieves tweet-image pairs from public Twitter accounts, analyze the Tweet, and generate labels for the images. The baseline system was then extended to include new functionality of extracting Named Entities, and use of machine translation to handle content in multiple languages.

Our research idea coincides with a number of earlier studies on social content and community tagging, with belief that tags associated with social images the potential of being a valuable information source for superior image search and retrieval experiences (Sun et al, 2011; Sawant et al, 2011). Our study shares partly similar goal as the T3 project (http://t3.umiacs.umd.edu) aiming at enhancing the usability of social tags, and recognizing that social inputs from social media such as Twitter with their added context represent a strong substitute for expert annotations or content based automatic annotations in helping the semantic interpretation of images (Sawant et al, 2011).

In the following we first introduce the image-tagging problem in section 2. We then present our system, report our initial experiments with the system in section 3. We also discuss visual labels methods making use of textual information in image annotation. In section 4 and 5 we elaborate on tag refinement issues and summarize the paper.

2 VISUAL LABELS: CONTENT BASED VS CONTEXT BASED

Visual labels can be broadly viewed as three types: (1) Content Labels - labels generated from directly analysis of visual content from computer vision and machine learning studies. The most recent developments in the field try to identify types of physical objects or activity/actions in images or videos through learning from datasets labeled with object concept terms, much promoted by ImageNet and TRECVID evaluations. (2) Context Labels: Location/space, Time, Social/Cultural/Personal Life events (Social media, FB), Domain/Topics. (3) Subjective Quality Labels: opinions, sentiments.

Using data from Delicious, Golder and Huberman (2006) identified seven functions that tags perform, including identifying what or who it is about, what it is, who owns it, refine categories, and self-reference. Primary motivations for social tagging can be sociality (whether the tag's intended usage is for self or others) or function (whether the tag's intended use is for organization or communication). For example, tagging for organizing purpose (search and browse), for attracting attention, for making contribution and sharing, for express opinion or self-presentation, and for social communication (adding context for friends, family, and the public) (Stvilia and Jorgensen, 2010).

Content-based visual labels are mostly "functional", as they indicate what it is and what it is about, using terms from general or domain specific formal concept taxonomy, following standard schemes for classification/tagging. They are professionally defined, accurate, consistent, controlled vocabulary, restrictive and static in nature, so easily run into coverage and scaling issues.

For images with little textual information around, annotation using visual content is a natural solution. Therefore, many content-based annotation algorithms have been proposed since 1999.

Social tags and text associated with images and videos on popular social media sites Flickr, Instagram, Facebook, Twitter, Youtube, are sources of rich semantic clues and context clues for broader indexing. They are author-given, can be general or personal, subjective or objective. They are user generated metadata with user choice of terminology, flexible, unstructured, and no vocabulary control, informal with non-hierarchical flat organization, thus there can be lots of noise, errors, irrelevance, redundancy and ambiguity, with varying level of granularity. They are emergent in nature and constantly evolving. The consolidation of social tags leads to a collective vocabulary that forms folksonomy - an informal, organic assemblage of related terminology (Vander Wal, 2005).

What are good labels and what type of tags is needed depends on the purpose of labeling, their intended usage or user preference: user oriented vs resource oriented, better representation of the visual resource for search and retrieval (labeling the visual resource), or better representation of user (labeling people's interest and hobby activities). Tag quality can be measured by tag relevance, accuracy and specificity, tag discrimination power, tag relatedness, tag representativeness and tag completeness. Level of granularity is important – a right mix of high-level (abstract) concepts or mid-level concepts and low-level concept instances would be able to meet the needs of labels for different purpose and different usage. Contextual clues may well offer a middle ground between generalized and specialized annotators.

3 EXTRACTING VISUAL LABELS FROM TWITTER

In this section we describe the TwitterAnalyzer/Tag Engine developed during our project on visual labeling. Our idea is to extract image tags from the associated Tweets of images. The social tags will then be merged with formal tags from content-based analysis at a later stage.

3.1 Core of the Tag-engine

The system retrieves tweet-image pairs from public Twitter accounts, analyze the Tweet to extract a number of items as candidate tags: Named Entities (location, people, organization), Hashtags, key words and phrases (tf-idf weighted words, frequency weighted bigram and trigrams). The ranking algorithm examines the extracted items and removes any redundancy between named entities, hashtags and ngrams (represent topics). Then post-processing is done to remove noisy tags, currently based on heuristic rules, which will later on be extended with more advanced methods. The system output would be a balance of different types of tags, depending on the targeted usage. The system also contains components to detect the language of the Tweet and automatically translate a non-English Tweet into English to be analyzed. The extracted tags are then translated back into original language.

Pre-Processing of Twitter text is very straightforward, only needs to pay attention to some special characters and adds to stop word lists. Emoticons are removed for the time being, as we only target content and context labels, not sentiment labels.

For named entity recognition we applied Stanford Named Entity Recognizer, which identifies names of people, places, organizations quite satisfactorily. Other types of proper nouns, e.g. names of products, books, magazines, movies, sports, other events and activities can often be identified in the hashtags or key phrase list.

N-gram extraction allows us to extract tags with size up to the specified amount N. This means that it is possible to extract multi-word or phrase labels to better describe any entities, topics and activities. When we set n-grams as 4, the system will extract unigrams, 2-grams, 3-grams, and 4-grams. Each of the ngrams are then weighted by adding together the unigram weights for each word the ngram contains. N-grams that start with or end with stop-words and punctuations are omitted.

Keyword and key phrase extraction help select a small set of words or phrases that are content bearing. As Tweets have very short text body, we take the simplest method for word weighting: TF-IDF for individual word weighting.

To be able to order the Named Entities according to relevance we can then increase the significance of the Named Entities so that they appear higher in the TF-IDF weighting. Another approach to sorting Named Entities by relevance is to find the highest weighted TF-IDF word that is linked to each Named Entity and sort them according to these TF-IDF values.

Removing noisy tags is very important and takes the most of our efforts. As expected, the extracted tags are of varying level of granularity in users' flexible terminology. There is a lack of consistency and lack of relationships between the tags when comparing with content-based labels. Through many debugging and testing we tried to add automatic filters to remove noisy tags by refining post-processing component.

3.2 Machine Translation and Dictionaries

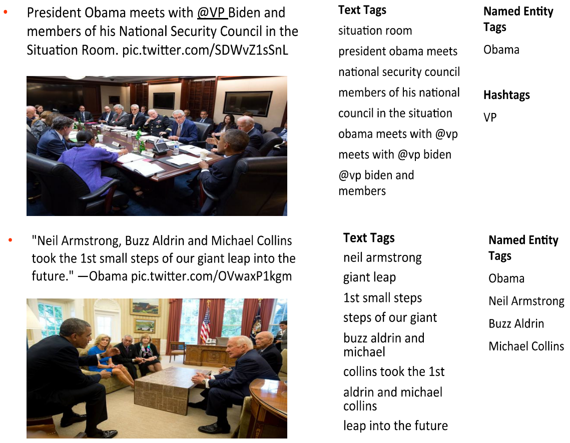
The system translates non-English text into English for processing and analysis. The extracted tags are then translated back to the original language.

Here we used an approach that combines dictionaries for slang, abbreviations, and intentionally misspelled words and interests with machine translation. We use the freely available Yandex translation resource to translate content of any language into English. A separate stop word list is not needed for the extra languages since stop words will be translated to English and then removed by the English stop word list.

We create dictionaries for abbreviations, slang and intentionally misspelled words to supplement the system. When extracting information we first remove the abbreviation, slang and misspelled words by going through the dictionaries and match them to the profile that is being parsed. After that we translate the text into English. When we have the profile text in English we can use our interest and hobby dictionaries to supplement the TFIDF extraction. For IDF database we found the one provided in MEAD very effective.

3.3 Examples and Debugging

Two examples of the tags extracted by our system are shown in the following figures. Fig. 1 shows two original tweet-image pairs retrieved from the public Twitter account The White House (@WhiteHouse), with the tweets in English. Fig. 2 shows two original tweet-image pairs from Aku Ankaa (https://twitter.com/akuankka_313), with its original tweets in Finnish language.



- President Obama meets with @VP Biden and members of his National Security Council in the Situation Room. pic.twitter.com/SDWvZ1sSnL

Text Tags
situation room
president obama meets
national security council
members of his national
council in the situation
obama meets with @vp
meets with @vp biden
@vp biden and members

Named Entity Tags
Obama

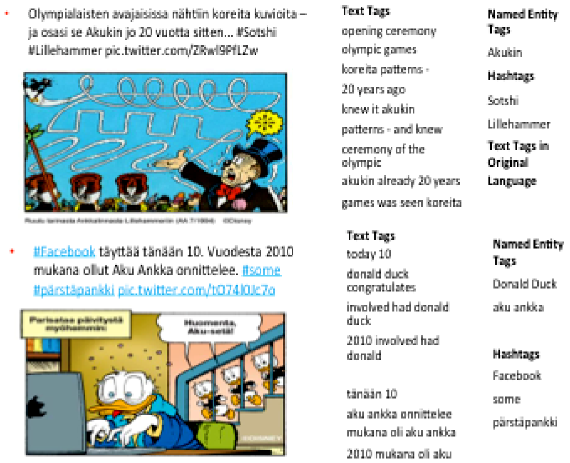
Hashtags
VP

- "Neil Armstrong, Buzz Aldrin and Michael Collins took the 1st small steps of our giant leap into the future." —Obama pic.twitter.com/OVwaxP1kgm

Text Tags
neil armstrong
giant leap
1st small steps
steps of our giant
buzz aldrin and michael
collins took the 1st
aldrin and michael collins
leap into the future

Named Entity Tags
Obama
Neil Armstrong
Buzz Aldrin
Michael Collins

Figure 1: Tag extraction example: English tweet, political domain.



- Olympialaisten avajaisissa nähtiin koreita kuvioita – ja osasi se Akukin jo 20 vuotta sitten... #Sotshi #Lillehammer pic.twitter.com/ZRwl9PfLZw

Text Tags
opening ceremony
olympic games
koreita patterns -
20 years ago
knew it akukin
patterns - and knew
ceremony of the olympic
akukin already 20 years
games was seen koreita

Named Entity Tags
Akukin

Hashtags
Sotshi
Lillehammer

Text Tags in Original Language

- #Facebook täyttää tänään 10. Vuodesta 2010 mukana ollut Aku Ankka onnittelee. #some #pärstäpankki pic.twitter.com/tO74l0Jc7o

Text Tags
today 10
donald duck congratulates
involved had donald duck
2010 involved had donald

tänään 10
aku ankka onnittelee
mukana oli aku ankka
2010 mukana oli aku

Named Entity Tags
Donald Duck
aku ankka

Hashtags
Facebook
some
pärstäpankki

Figure 2: Tag extraction example: Finnish tweet, comics.

As we can see, the labels are a mixture of concepts at different levels, sometimes can be overlapping with high level concept and formal tags, but not in most cases. Hashtags can have much overlapping with Named Entities and text tags (key words and ngrams). We use heuristic rules for first layer post-analysis and processing of the tags: (1) keep all Hashtags; (2) Named entities in hashtags considered more relevant, so they get higher priority; (3) Ngram weighting adjusted by individual word tf-idf weighting; (4) we consider variety a necessity and priority of a good tag set, to include location, time, organization, people and topics.

A debugging site was set up to enable us test and manually assess the extracted tags, to find ways fine-tuning our ranking method and algorithm (at a later stage, we will add the feedback mechanism to incorporate user feedback into the system directly).

Debugging hopefully helps us to exploit the potential of heuristic rules to a great extent. Still we found problems with the Topic tags that – most favorable tags are not being top weighted.

4 TAG ANALYSIS AND REFINEMENT

Social tags are assigned by different users with different motivations for tagging, different understandings of relatedness between tags and images, or even different interpretations of the meaning of tags arising from knowledge or cultural diversity (Sun et al, 2011). In the nus-wide dataset (Flickr photos), more than 420K distinct tags have been used to annotate 269K images, many of which do not describe the visual content of these images.

Klavans et al (2011) reported a linguistic analysis of a tag set containing nearly 50,000 tags collected as part of the steve.museum project (http://www.umiacs.umd.edu/research/t3/link.shtml). The tags to 1,785 works describe images of objects in museum collections. They stressed the importance of leveraging the tags and relationships between them, utilized NLP tools and formal resources such as WordNet and domain ontology to help normalize the tags (Klavans et al, 2011).

In our context, we consider such tag refinement is important smoothing only when we already collected good candidate tags and removed common noises – the second level refinement. The immediate challenge for us is to attain a good balance between rich tags and relevant tags. So our current focus is to re-rank the tags of a tagged image such that the most relevant tags appear in top positions, while to make sure important tags are covered – the first level refinement.

4.1 First Level Refinement

Wang et al (2006) proposed a tag re-ranking method using Random Walk with Restarts (RWR) for image annotation refinement. Measuring tag relatedness would be another approach, which helps to remove e.g. tags of self-reference and provide suggestions for tags that describe image content.

Tag relatedness can be based on tag associations as well (Tag-by-association). Associations between tags can be computed based on tag co-occurrence, Google distance, Jaccard coefficient etc. (Liu et al, 2009, 2010).

Tag relatedness and refinement can also be based on visual similarity or visual-representativeness, which indicates the effectiveness of a tag in describing the common visual content of its annotated images, and is mainly applicable to "content related" tags. A tag is visually representative if its annotated images are visually similar to each other, containing a common visual concept such as an object or a scene (Wang et al, 2006). If multiple people use same tags to label visually similar images, then these tags are likely to reflect the visual contents of the annotated images (Liu et al, 2009, 2010).

With our tag extraction system, we will be able to gather large amount of images with shared tags and visual similarities, to find the common visual theme from the shared tags, and assess their visual representativeness.

Finally, tag relevance can eventually be evaluated by matching images to a given tag query or user query.

4.2 Second Level Refinement

Liu et al (2009) noticed that topic/subject tags account for more than half of the searches in tag-based image searching and browsing, while photos of specific named entities (scientist, politician, building and scenery) relate to a very small portion of topic/subject tags. Our second level refinement will mostly concentrate on topic/subject tags.

Here the major issue is to map social tags to or associate social tags with formal concepts, i.e., to relate social tags to its related upper level concepts. This would be the foundation for connecting and integrating context social tags and content-based concept labels.

Different methods have been proposed to fuse content-based visual features with noisy social labels. Jin et al (2005) approached image annotation refinement with using WordNet to prune the irrelevant annotations. However, their experimental results show that although the method can remove some noisy words, many relevant words are also removed, and many tag words simply do not exist in the lexicon of WordNet. Noel and Peterson (2013) also proposed to leverage WordNet synsets for selection of appropriate annotations. It converts words surrounding an image into WordNet synsets related with ImageNet concept labels.

In addition to the above methods, with more extensive social ontology resources become available, for example Freebase, DBPedia, BabelNet, which are all databases about millions of things from various domains, they could form potential new bridges between taxonomy and folksonomy, and could be useful for us to integrate the social tags with formal visual content tags.

For our system, it will be possible to test, compare and integrate two approaches: one trying to make use of existing lexical and ontology resources, the other being based on statistical methods for similarity and clustering analysis to find closely related concepts.

5 SUMMARY

In this study we explore the possibilities to leverage social media content as a resource for visual labeling. We developed a tag extraction system that applies heuristic rules and term weighting method to extract image tags from associated Tweet. The system retrieves tweet-image pairs from public Twitter accounts, analyzes the Tweet, and generates labels for the images. To put our work in context and preparing for future work, we discuss different types of visual labels methods that make use of textual information in image annotation. We also elaborate on tag analysis and refinement methods and techniques, and the integration of context-based labels with content concept labels.

Overall, the system is simple, generic, handling multiple languages. Feedback mechanism can be incorporated at a later stage to help refine and control the quality of the tags, and collect user approved tag sets. user feedback mechanism to help improve the tags sets by taking into consideration of user feedback in testing process.

From information retrieval point of view, Folksonomy is often criticized because tags are not drawn from a controlled vocabulary. The aggregated terminology drawn from tagging is expected to be inherently inconsistent, and therefore flawed, according to theories of indexing (Trant, 2009).

On the other hand, solely content-based concept labels are not enough to meet needs of users in image search and organization. Social tag based search is an important way of searching or browsing images. Complete image tag sets should contain tags that offer visual clues, semantic clues and context clues (Ref). Context information is becoming more and more important for enhancing retrieval performance and recommendations. Social media offers tremendously rich content for the extraction of contextual visual labels.

Comparing with content-based labels, social tags assigned to an image may not necessarily describe its visual content, but instead describe time, location, people or social event, and a mixture of concepts at different levels. They are able to offer the advantage of both generalized and specialized annotators, and can be expected to have more semantic correlation to user queries and be more user friendly.

Our study is only at the beginning. Our immediate next step is deeper tag analysis and implementing tag re-ranking and refinement techniques. The extracted social tags will be studied in the context of formal labels and user queries. By integrating results from our tag extraction system with content-based methods, we will be able to study many things and move towards the construction of a social image database with rich set of tags covering both formal concept tags and social context tags.

ACKNOWLEDGEMENTS

This research is supported by the Tekes funded DIGILE D2I research program, Arcada Research Foundation, and our industry partners.

REFERENCES

Liu, X., Zhang, S., Wei, F., & Zhou, M, June. Recognizing Named Entities in Tweets. 2011. ACL (pp. 359-367).

Chen M., A. Zheng, K. Weinberger, Fast Image Tagging, 30th International Conference on Machine Learning (ICML), 2013

Sjöberg M., M. Koskela, S. Ishikawa and J. Laaksonen. Large-Scale Visual Concept Detection with Explicit Kernel Maps and Power Mean SVM. In Proceesdings of ACM ICMR 2013, Dallas, Texas, USA

Sjöberg Mats., J. Schlüter, B. Ionescu and M. Schedl. FAR at MediaEval 2013 Violent Scenes Detection: Concept-based Violent Scenes Detection in Movies. In Proceedings of MediaEval 2013 Multi-media Benchmark Workshop, Barcelona, Spain, October 18-19, 2013.

Jin, Y., Khan, L., Wang, L., and Awad, M. Image Annotations By Combining Multiple Evidence & Wordnet. Proc. of ACM Multimedia, Singapore, 2005

Jain R. and P. Sinha. Content without context is meaningless. In Proceedings of the international conference on Multimedia (MM’10), pages 1259–1268, Firenze, Italy, 2010. ACM.

Sun A. and S. S. Bhowmick. Image tag clarity: in search of visual-representative tags for social images. In Proceedings of the first SIGMM workshop on Social media (WSM’09), pages 19–26, Beijing, China, 2009. ACM.

Sun A. and S. S. Bhowmick. Quantifying tag representativeness of visual content of social images. In Proceedings of the international conference on Multimedia (MM’10), pages 471–480, Firenze, Italy, 2010. ACM.

Sun Aixin, Sourav S. Bhowmick, Khanh Tran Nam Nguyen, Ge Bai, Image-Based Social Image Retrieval: An Empirical Evaluation, American Society for Information Science and Technology, 2011.

Tang J., S. Yan, R. Hong, G.-J. Qi, and T.-S. Chua. Inferring semantic concepts from community-contributed images and noisy tags. In Proceedings of the 17th ACM international conference on Multimedia (MM’09), pages 223–232, Beijing, China, 2009. ACM.

Rorissa A., A comparative study of flickr tags and index terms in a general image collection. Journal of the American Society for Information Science and Technology (JASIST), 61(11):2230–2242, 2010.

Ding, E. K. Jacob, M. Fried, I. Toma, E. Yan, S. Foo, and S. Milojevic. Upper tag ontology for integrating social tagging data. Journal of the American Society for Information Science and Technology (JASIST), 61(3):505– 521, 2010.

Liu D., X.-S. Hua, L. Yang, M. Wang, and H.-J. Zhang. Tag ranking. In Proceedings of the 18th international conference on World wide web (WWW’09), pages 351–360, Madrid, Spain, 2009. ACM.

Liu D., X.-S. Hua, M. Wang, and H.-J. Zhang. Image retagging. In Proceedings of the international conference on Multimedia (MM’10), pages 491–500, Firenze, Italy, 2010. ACM.

Chua T.-S., J.Tang, R.Hong, H.Li, Z.Luo,and Y.Zheng. Nus-wide: a real-world web image database from national university of Singapore. In Proceeding of the ACM International Conference on Image and Video Retrieval (CIVR’09), pp 48:1–48:9, Santorini, Fira, Greece, 2009.

Sawant Neela, Jia Li and James Z. Wang, `` Automatic Image Semantic Interpretation using Social Action and Tagging Data," Multimedia Tools and Applications, Special Issue on Survey Papers in Multimedia by World Experts, vol. 51, no. 1, pp. 213-246, 2011.

Wang Changhu, Feng Jing, Lei Zhang, Hong-Jiang Zhang, “Image Annotation Refinement using Random Walk with Restarts”, MM'06, October 23–27, 2006, Santa Barbara, California, USA

Liu Dong, Xian-Sheng Hua, Meng Wang, HongJiang Zhang, Microsoft Advanced Technology Center, ICME 2009

Makadia Ameesh, Vladimir Pavlovic and Sanjiv Kumar, A New Baseline for Image Annotation, Google Research, New York, NY, & Rutgers University, Piscataway, NJ, 2008

Noel George E. and Gilbert L. Peterson, Context-Driven Image Annotation Using ImageNet, Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, IEEE 2013

Wang, Feng Jing, Lei Zhang, Hong-Jiang Zhang, Image Annotation Refinement using Random Walk with Restarts, MM'06, Oct 23–27, 2006, Santa Barbara, California, USA

Klavans Judith L., Raul Guerra, Rebecca LaPlante, Robert Stein, and Edward Bachta, Beyond Flickr: Not All Image Tagging Is Created Equal, Language-Action

Tools for Cognitive Artificial Agents, the 2011 AAAI Workshop (WS-11-14)

Stvilia, B. and Jörgensen, C., Member activities and quality of tags in a collection of historical photographs in Flickr. Journal of the American Society for Information Science and Technology, 61:2477–2489, 2010

Vander Wal, Folksonomy Definition and Wikipedia, Nov. 2005, http://www.vanderwal.net/random/entrysel.php?blog=1750

Trant, J. Tagging, Folksonomy and Art Museums: Early Experiments and Ongoing Research, Journal of Digital Information, Vol 10, No 1, 2009