

Towards A Collaborative Framework for Image Annotation and Search

Yi Hong* and Stephan Reiff-Marganiec

Department of Computer Science
University of Leicester, UK, LE1 7RH

Abstract. Users tag images with plain text information, which is then used as the basis for search. For the large amount of digital images available on the web this becomes challenging because the tags are abstract concepts whose relationship is undefined. For effective search which requires reasoning on concepts and their relations one requires richer data structures for tagging and one needs to take into account the confidence and credibility of the tagging user. In this paper, we introduce a novel collaborative framework for image annotation, which allows users to create tags that are based on a concept repository which provides a hierarchical context for them as well as allowing to define relationships among said concepts. It also provides a new and systematic way to establish user credibility as well as to compute the truthfulness or reliability of a particular statement, which are used for ranking search results. A prototype has been implemented using this approach and we will show some examples to explain our methodology in detail.

1 Introduction

In recent years the number of images available online has increased significantly leading to a problem of “information overload” in the sense that finding what one is looking for becomes hard. Tagging or annotating has become a popular way of adding searchable information to images, especially in shared environments such as social networking websites. An image tag is a small piece of plain text or some keywords attached to a specific area of an image. It helps users in organising and searching image content. However, it raises a challenging question about how to structure metadata to enable users to describe, extract and search information based on images in a more accurate and efficient way.

Currently there are a number of tagging approaches available. Generally these are based on keywords, but there have also been some efforts centered around ontologies. We will now highlight some problems of the keyword-based tagging approaches and some known issues existing in semantic-oriented tagging frameworks [8]. Keyword-based tagging approaches have the following drawbacks:

Ambiguous semantics. In traditional image tagging system, a tag is normally a freely-chosen, non-hierarchical keyword or term. The tag in several pictures

* This work was partially supported by Leverhulme Trust Funded “Tracing Networks” Programme.

can be identical but the meaning maybe ambiguous. For example, plain text “date” tagged on an image might have different interpretation. It may refer to a day on the calendar, a fruit or the image is showing someone out on a date. Because the word has several meanings and the context of its use is not mentioned it is unclear what is meant. Similarly, a system is unable to tell whether the picture with tag “mouse” is referring to a computing device or an animal, something totally different.

Inadequate support for describing relationships. Current tagging systems are focused on labelling elements in the picture rather than the relationships among them, but in the real world specifying relationships between entities is as important as identifying the entities themselves. For example, a user can select two tagging areas on the image and annotates them with “cat” and “mouse”, but most annotation systems do not provide enough support to describe the relationship between them in a formal way e.g. “the cat is chasing the mouse” (but not “the mouse is chasing the cat”, or even “the cat is sleeping near a [computer] mouse”).

Ability to perform automatic reasoning. Unstructured plain text tags do not allow to perform any reasoning tasks along with the search. For example, given a query “display all images with two animals on it”, the system will not be able to answer this question because it does not have sufficient knowledge to reason about the facts, such as e.g. cats and mice being in fact animals. Due to the previous point there can also not be any reasoning for searching for pictures where for example “mice are chasing cats”.

To overcome the limitation of traditional tagging approaches, semantic tagging applications have been developed e.g. [11]. These are also not perfect:

Maintaining domain-specific ontology without domain experts. Many of these systems are developed specifically for a particular domain (e.g. Medicine, Bioinformatics etc.) and are normally implemented on the basis of a domain-specific ontology. Although these ontologies provide a formal knowledge representation as a set of concepts and relationships within this particular domain which fits and works quite well, the ontologies need to be maintained by domain experts in collaboration with ontology developers. Long-term maintenance may become an issue.

Describing truthfulness and reliability of a statement. Current semantic tagging applications do not have the functionality to say how truthful or reliable user’s statement is. The system does not support statements such as “I have strong evidence to believe that there is a cat in the tagging area” or “It is probably a cat in the picture, but I am not sure about it”, nor a way to express these in any other form.

Measuring user credibility in collaborative environment. Most collaborative image tagging applications do not take into account the trustworthiness of a statement and the reputation of a user based on their expertise. E.g. Alex is a zoologist and she believes the animal in the picture is a house cat without a doubt while a 5 year old child tagged it as a lion. The trustworthiness of the opinion or statement should be determined by both the

credibility of a user in a certain field along with how certain they are in making the judgement.

In this paper, we introduce a novel collaborative framework for image annotation, which allows users to create tags that are based on a concept repository which provides a hierarchical context for them as well as allowing to define relationships among said concepts. It also provides a new and systematic way to establish user credibility as well as to compute the truthfulness or reliability of a particular statement, which are used for ranking search results.

The next section will introduce the framework, section 3 will look at the implementation and evaluation. Section 4 shows related work and section 5 completes the paper with conclusions and an outlook to future work.

2 Collaborative Image Annotation and Search Framework

We propose a framework for image annotation and search in collaborative working environments to address the problems listed above. The framework is built on the basis of an *Image Annotation Ontology*, which functions as an abstract data model for organising and storing tagging data. It provides the infrastructure for annotators to link a tag to the predefined concept within a conceptual-semantic lexical database. This forms our knowledge base, therefore a reasoner can be used to exploit implicit knowledge as the tagging data are structured in a machine readable and understandable way. The tagging framework also allows users to define the degree of uncertainty of a statement. This is called certainty factor and will subsequently be used in conjunction with user credibility to compute the truthfulness of statements. This additional user context will greatly improve the efficiency and accuracy of a query. We will illustrate the main components and steps with a few examples in the following sections.

2.1 Data Model

To allow for reasoning and structuring the tagging information an underlying structure is needed. The basic structure of the Image Annotation Ontology is shown in Fig. 1. This structure allows to store information about a tagging area from the image and the relevant annotations as in conventional image tagging system. In addition to that, assumptions concerning a subject-predicate-object triple or a link between tagged area and resources can be stored. Every assumption is associated with an original certainty factor given by the annotator. Each individual user is assigned a set of credibility values for the different domains in which they annotate according to their expertise and reputation in that domain.

Resources are a group of synonyms defined in the WordNet lexical database, organised into hierarchies by hypernym (or hyponyms). A hypernym is referred as an “is a” relationship, for example the phrase *airplane is a craft* can be used to describe the hyponymic relationship between airplane and craft. The WordNet hierarchy for the word “aircraft” [16] is shown in Table 1.

While linking a tagging area to a term defined within WordNet, the definition of the term as well as the whole branch of the hypernym tree describing the

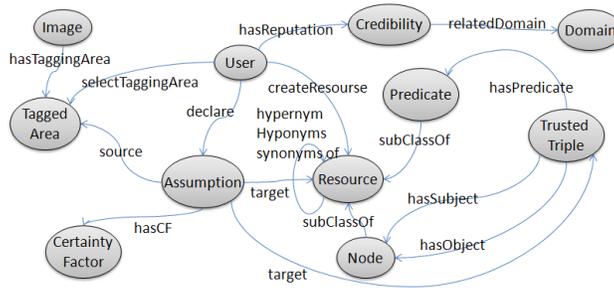


Fig. 1. Image Annotation Ontology Overview

Table 1. hypernym, hypernyms and synonyms of “aircraft” in WordNet

airplane, aeroplane, plane
 heavier-than-air_craft
aircraft
 craft
 vehicle
 conveyance, transport
 instrumentality, instrumentation
 artifact, artefact

semantic relations between these nodes, will be added to the triple store to populate the knowledge base. A user is free to create new terms by adding a new branch to the tree. A new term is fully defined as long as all leaf nodes in its branch are linked to already predefined terms.

Using WordNet provides us with an ontology that is not domain specific, but rather generic and that is maintained in a collaborative fashion. In addition it is not specific to our application and hence maintenance becomes a property somewhat distant from our immediate concerns.

2.2 Uncertainty in Trusted-Triple Graph

The subject-predicate-object relation mentioned in the previous section is common to ontologies and the usual structure used for storage is a Triple Graph. We are proposing an extension to Triple graphs to include a certainty factor (cF)[10] which is a number from -1.0 to 1.0 indicating how accurate or certain a user is about an assumption. Positive certainty means the user basically believes the assumption is a true statement, but he or she is possibly not 100% certain. Negative certainty means a disagreement with a given assumption, but might not rule out the possibility. For example, a statement such as “The blonde girl wearing the blue T-shirt is Alice” might be given a number such as 0.95 from a friend of hers; who would be reasonably certain of the fact. But if you had never met Alice, but heard that she is an Asian girl then you will probably give a

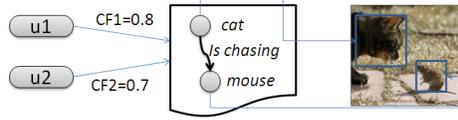


Fig. 2. An example of user’s opinion on a statement

negative certainty number, say -0.7 as generally Asians are not blonde. Assume $cF(u, s)$ is the original certainty factor given by a user u for assumption s . To combine the certainty factor and user reputation we define the concept of a composite user certainty-credibility factor $CF(u, s, r_d)$ applying to an assumption s made by a user u with reputation r in a domain d . CF is defined as follows:

$$CF(u, s, r_d) = r_d \cdot CF(u, s) \quad (1)$$

where s is a subject-predicate-object triple or an assumption about the relationship between tagging area and a predefined concept, $CF(u, s)$ is the certainty factor given by user u on s , and r_d is the user’s reputation r in domain d .

We will come back in section 2.3 to explain how r_d is calculated. To compute the overall certainty-credibility factor CF_s based on all available judgements of all users (e.g. $u1, u2$) over the same statement s , we combine $CF_1 = CF(u_1, s, r_{d_1})$ and $CF_2 = CF(u_2, s, r_{d_2})$ using the parallel function [10] in (2):

$$CF_s = \begin{cases} CF_1 + CF_2(1 - CF_1), & \text{if } CF_1, CF_2 \geq 0, \\ CF_1 + CF_2(1 + CF_1), & \text{if } CF_1, CF_2 < 0 \\ \frac{CF_1 + CF_2}{1 - \min(|CF_1|, |CF_2|)}, & \text{other} \end{cases} \quad (2)$$

For example, consider the statement “a cat is chasing a mouse” shown in Figure 1. For $CF_1 = 0.8$ and $CF_2 = 0.7$, we have $CF = 0.8 + 0.7 \cdot (1 - 0.8) = 0.94$.

In this case, the overall composite user certainty-credibility factor is 0.94, significantly above the average of the individual users’ claims, which might come as a surprise. There is a certain sense in this: if many user’s tend to claim the same thing with reasonable confidence then one can overall be more confident in it actually being true and the effect of people being ‘shy’ in claiming certainty becomes reduced. However, it also stops us from arguing that the a statement is ‘twice as true’ as another just because it has double the score.

2.3 User Credibility Extraction

We need to obtain the reputation [9] r of a user to calculate the user certainty-credibility factor. The expertise factor [7] defines the degree of a user’s competency to provide an accurate prediction in a particular field. In our case, the reputation r of a user u in a domain d , it is defined by (3)

$$r(u, d) = \beta(u, d) \left(1 - \frac{\sum_{s \in D(s)} \sum_{a \in U(s)} |CF(u, s) - CF(a, s)|}{|D(s)|} \right) \quad (3)$$

where $CF(u, s)$ is the original certainty factor given by user u on s , $U(s)$ is a set of users who commented on statement s (excluding user u), $D(s)$ is the statement set in the domain of target statements s , and $\beta(u, c)$ is the activity weighting and is defined as $1 - (1/n)$ (n is the number of comments within this category), meaning that a user will be assigned a higher value of expertise if making more comments for more statements within a particular category¹.

Note that we refer to $r(u, d)$ as r_d if it is clear which user is concerned.

As can be seen in formula (3), the reputation of an annotator is determined by several aspects. In general, (a) the more statements in the same category a user commented the more likely are they to have specific expertise in the field and this will increase their reputation in this category, but only if (b) the opinions the user provided reflect the truthfulness of the actual statement as measured by the similarity between their judgement and judgements made by other users.

For example, our zoologist Alex commented on a lot of pictures regarding wildlife and her judgement is quite accurate in most cases, so normally her opinion in this domain does not differ much from the general public including other zoologists. Also, the annotations she created about wildlife were generally accepted by the community. In this case, Alex should have a high reputation in the wildlife domain. While (say) a computer scientist called Tom also made lots of comments on wildlife, but his opinions received negative feedback; therefore he is less credible than Alex in this domain.

2.4 Query and Reasoning

So far we have presented how image information is enhanced with semantic data and other aspects. This is the preparation for successful searches. The Annotation Ontology model we used for structuring the tagging data and user context makes it possible for the system to run intelligent queries in conjunction with a reasoning component. Queries can be defined by graph patterns rather than simple keywords. It allows for more flexible ways of obtaining answers for queries. A modified version of our example in the introduction could be:

“Display all images with two animals in them, along with what is happening between them”

A specific variant of that query might be ‘find all pictures where mice are chasing cats’. This query is concerned with entities and their relationships in the picture. Our trusted-triple repository, which is built on top of the Image Annotation Ontology, contains the information about the tagged areas and their relationships. It also contains the hypernym hierarchy (e.g. the fact that a cat is an animal) derived from the WordNet database. In other words, the background knowledge we need to answer this query is already captured by the data model. Apart from this, we will need a set of inferencing rules to perform the reasoning. To formulate the questions in a formal way, we rewrite it as listed below.

¹ Bear with us! We are not claiming that just because someone makes lots of comments they are more qualified to do so.

```

Ask for: ?a1 ?relation ?a2
Image(?i),appearsOn(?a1,?i),appearsOn(?a2,?i)
Animal(?a1),Animal(?a2),?relation(?a1,?a2)

```

This query will have to be run together with several deductive reasoning rules (rules of the form “antecedents \rightarrow consequent”), to get the search results. Examples of these are:

```

Mouse(?a),hyponymOf(Mouse, Animal)-> Animal(?a)
Cat(?a),hyponymOf(Cat, Animal)-> Animal(?a)

```

Clearly if one had to specify all of these this would not be sensible. However, the reasoning rules can be rewritten in a more generic and reusable form. The relation “hyponymOf” is a transitive property and the second rule will compute the transitive closure over the relation “hyponymOf”.

```

Concept(?c1),Concept(?c2),?c1(?x),hyponymOf(?c1, ?c2)-> ?c2(?x)
hyponymOf(?c1,?c2),hyponymOf(?c2,?c3)->hyponymOf(?c1,?c3)

```

The query and inference rules are in fact graph patterns and they can be translated to SPARQL [2] queries and SWRL [3] rules. Though unbounded predicate support (such as *?relation*) is currently not available in many rule-based query languages such as SQWRL [13], we can still translate it to SPARQL-DL.

In addition to this, as we are using a trusted triple graph, we can also carry the information on uncertainty through the reasoning by attaching a certainty factor to each custom reasoning rule as suggested in [10].

2.5 Ranking

Once the search results are obtained they are ranked by the degree of truthfulness. This is calculated for every sub graph in the result set that matched the pattern. The weight of the subgraph is defined by the function:

$$W_G = \prod_{s \in T} CF_s \quad (4)$$

where G is a trusted subgraph matching the search pattern, T contains all triples in G , and CF_s is the overall composite certainty-credibility factor of triple s .

W_G will then be used in the ORDER BY clause of the concrete query implementation providing a result set with the most trusted results at the top.

3 Implementation and Evaluation

A web-based prototype application has been implemented in Java. Figure 4 illustrates the architecture of the system, which consists of several parts. A web-based annotation interface allows users to annotate the image. An annotator can describe concepts and relationships in the image by constructing a set of trusted triple statements, with help of an extendable WordNet lexical database

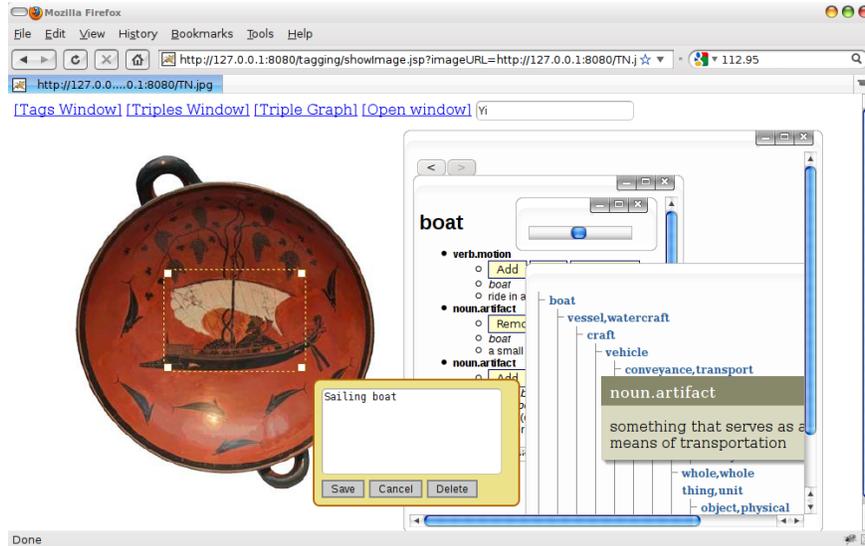


Fig. 3. Screenshot of the prototype implementation

and a remote OpenCalais[1] service. WordNet groups the synonymous words, provides precise definition of terms and defines the semantic relation between these synonym sets. It works in conjunction with the OpenCalais service to help user identify the possible topics (category or domain) of an image. Besides, an annotator can also specify the degree of certainty about an assumption. A screenshot of the interface is shown in Fig. 3. For every triple statement, the overall user certainty and credibility factor will be computed by a certainty calculator and a user credibility calculator respectively; these will then be combined into the composite user certainty-credibility factor used for ranking search results. The query engine retrieves annotation data from the trusted triple repository by means of SPARQL queries and a native OWL API. The system also utilises both description logics (DL) and deductive rule-based reasoning for inferencing. The search screen is currently simply a textbox that allows entry of a query and the results show a ranked list of images that match the given query – both not very exciting to look at in a screenshot.

We are evaluating our approach with users from the archaeology community. “Tracing Networks: Craft Traditions in the Ancient Mediterranean and Beyond” [4] is a joint archaeological research programme, which involves archaeologists and Computer Science researchers. This research programme investigates the network of contacts across and beyond the Mediterranean region, between the late Bronze Age and the late classical period (1500-200 BCE). Researchers in different sub-projects have already gathered massive amounts of image resources and cross-team knowledge sharing and analysis are vital – and being able to retrieve the right pictures is essential. The intention is to apply the presented

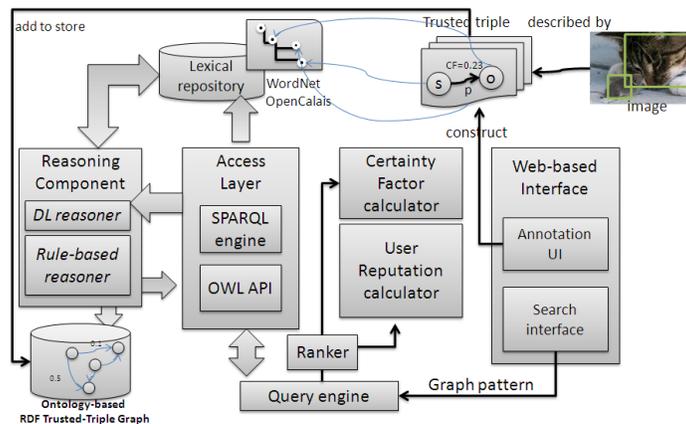


Fig. 4. System architecture overview

approach to enhance the collaboration of teams and enable future research by others.

4 Related Work

Various approaches and techniques for tagging have been proposed by different research groups in the last few years, especially to identifying the vocabulary such as folksonomies [15]. A survey of approaches for various semantic annotation techniques is presented in [14]. We have already in the introduction summarised the shortcoming of these works. Many other existing collaborative semantic tagging systems such as SemKey make use of external resources such as WordNet in their approaches [12, 6]. However, most available systems and frameworks are more focused on metadata modelling, concept identification and relationship extraction instead of representing uncertainty, user credibility and taking into account all these factors for reasoning and searching. The closest effort on uncertainty is the W3C Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG), which proposed an Uncertainty Ontology in their report in 2008 [5]; however their focus is more defining the data structure to hold uncertainty information than on its calculation and use.

5 Conclusion & Future Plan

In this paper, we identified problems and limitations of available tagging systems and proposed a framework for image annotation and search in a collaborative environment. We developed an ontology-based data model for identifying concepts, relationships and storing context regarding users. The WordNet lexical database, which describes the semantic relations between different terms is used to disambiguate the keywords and populate the knowledge base. We also

introduced a systematic way to represent uncertainty of a statement as well as user credibility measurement; these two factors are combined into a composite uncertainty-credibility factor, which is used for ranking the search results. We explained with examples how rule-based reasoning can be used in conjunction with graph patterns to help us answer advanced queries. Finally, we presented a prototype implementation that illustrates our methodology.

We are currently looking into several areas of future work, such as (1) evaluating the approach with a variety of users from different communities, (2) considering a more complex user context to provide a more context-aware search and (3) an evaluation of how these factors affect search results and ranking.

References

1. Calais: Connect. Everything. <http://www.opencalais.com/>.
2. SPARQL Query Language for RDF. <http://www.w3.org/TR/rdf-sparql-query>.
3. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. <http://www.w3.org/Submission/SWRL/>.
4. Tracing Networks: Craft Traditions in the Ancient Mediterranean and Beyond. www.tracingnetworks.ac.uk.
5. Uncertainty Reasoning for the World Wide Web. <http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/>, March 2008.
6. Doina A. Cernea, Esther Del Moral, and Jose E. Labra Gayo. SOAF: Semantic Indexing System Based on Collaborative Tagging. *Interdisciplinary Journal of E-Learning and Learning Objects*, 4:137–149, 2008.
7. Jinyung Cho, Kwiseok Kwon, and Yongtae Park. Q-rater: A collaborative reputation system based on source credibility theory. *Expert Systems with Applications*, 36(2, Part 2):3751–3760, 2009.
8. Wai-Tat Fu, Thomas Kannampallil, Ruogu Kang, and Jibo He. Semantic imitation in social tagging. *ACM Trans. Comput.-Hum. Interact.*, 17:12:1–12:37, July 2010.
9. Anna Gutowska and Kevan Buckley. A computational distributed reputation model for b2c e-commerce. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 03*, pages 72–76, Washington, DC, USA, 2008. IEEE Computer Society.
10. David E. Heckerman and Edward H. Shortliffe. From certainty factors to belief networks. *Artificial Intelligence in Medicine*, 4(1):35–52, 1992.
11. Hak-Lae Kim, Simon Scerri, John Breslin, Stefan Decker, and Hong-Gee Kim. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies. In *International Conference on Dublin Core and Metadata Applications*, Berlin, Germany, 2008.
12. Andrea Marchetti, Maurizio Tesconi, Francesco Ronzano, Marco Rosella, and Salvatore Minutoli. Semkey: A semantic collaborative tagging system. In *Proc. WWW 2007 Workshop on Tagging and Metadata for Social Information Organization*, Banff, Canada, May 2007.
13. M. J. O’Connor and A. K. Das. SQWRL: a Query Language for OWL. In *OWL: Experiences and Directions (OWLED), Fifth International Workshop*, 2009.
14. Lawrence Reeve. Survey of semantic annotation platforms. In *Proceedings of the 2005 ACM Symposium on Applied Computing*, pages 1634–1638. ACM Press, 2005.
15. Gene Smith. Folksonomy: social classification (Atomiq), August 2004.
16. Michael M. Stark and Richard F. Riesenfeld. Wordnet: An electronic lexical database. In *Proceedings of 11th Eurographics Workshop on Rendering*. MIT Press, 1998.