Enrichment and Ranking of the YouTube Tag Space and Integration with the Linked Data Cloud

Smitashree Choudhury[1], John G. Breslin[1,2], and Alexandre Passant[1]

[1] DERI, National University of Ireland, Galway, Ireland
[2] School of Engineering and Informatics, National University of Ireland, Galway, Ireland
{smitashree.choudhury,john.breslin,alexandre.passant}@deri.org

Abstract. The increase of personal digital cameras with video functionality and video-enabled camera phones has increased the amount of user-generated videos on the Web. People are spending more and more time viewing online videos as a major source of entertainment and “infotainment”. Social websites allow users to assign shared free-form tags to user-generated multimedia resources, thus generating annotations for objects with a minimum amount of effort. Tagging allows communities to organise their multimedia items into browseable sets, but these tags may be poorly chosen and related tags may be omitted. Current techniques to retrieve, integrate and present this media to users are deficient and could do with improvement. In this paper, we describe a framework for semantic enrichment, ranking and integration of web video tags using Semantic Web technologies. Semantic enrichment of folksonomies can bridge the gap between the uncontrolled and flat structures typically found in user-generated content and structures provided by the Semantic Web. The enhancement of tag spaces with semantics has been accomplished through two major tasks: (1) a tag space expansion and ranking step; and (2) through concept matching and integration with the Linked Data cloud. We have explored social, temporal and spatial contexts to enrich and extend the existing tag space. The resulting semantic tag space is modelled via a local graph based on co-occurrence distances for ranking. A ranked tag list is mapped and integrated with the Linked Data cloud through the DBpedia resource repository. Multi-dimensional context filtering for tag expansion means that tag ranking is much easier and it provides less ambiguous tag to concept matching.

1 Introduction

A key feature of the Social Web is the change in the role of a user from simply being a consumer of media: they are now content creators. It is not just textual content that can be shared, annotated or discussed, but any multimedia content such as pictures, videos, or even presentation slides. With tools like iMovie for video creation and digital cameras with built-in WiFi for instant uploads, web users can easily add their multimedia content to social media websites. With this ease of creation, there is an ever increasing amount of multimedia in various formats becoming available on the

A. Bernstein et al. (Eds.): ISWC 2009, LNCS 5823, pp. 747–762, 2009.

Social Web. Recently YouTube[1] reported that 20 hours of video were being uploaded per minute, which amounts to 28,800 hours of video uploaded in one day to that site.

All of these videos are being annotated by users with free unstructured keywords. Some video sharing sites also permit sharing and collaboration in the tagging process by allowing other users to tag a video, thereby giving a sense of collective intelligence. Current techniques to retrieve, integrate and present this tagged media to users are deficient and could certainly benefit from improvement. Semantic technologies make it possible to give richer descriptions to media, facilitating the process of locating, combining diverse media from various sources and personalising content recommendation.

A major problem is that the textual annotations vary in terms of quality and their ability to describe the video content. The tags include not only the content but also information about the user, their subjective opinion of the content, misspelling, and emerging co-joined tags. Given the ambiguity, subjectivity and noise in tags, one of the fundamental problems is to learn the relevance of the tag corresponding to the content. Unstructured and informal descriptions rule out any kind of interoperability of the resources across similar and related content. An attempt to give a well-defined structure and to formalise the tag space for user-generated videos will be the first step towards a desired solution. Moreover, we believe that this could be an efficient way to add relevant semantics to videos on the Web, in combination with existing initiatives, such as the MPEG7 [10] standard and its associated RDF(S)/OWL mappings (that can be used to represent image regions and add particular annotations about them) and the current tasks of the W3C Media Annotation Working Group, as defined in their document on “web video”[2].

In this study, we have designed a framework to explore the contribution of various types of contextual data to the tag space and their relevance in ranking. Information embedded in video contexts such as social, spatial and temporal contexts are a good source for video tag suggestions. Enriching tags, though helpful for more reliable descriptions of content, can at the same time add noise to the resulting video metadata. In order to attenuate the noise from the tag space, we need to rank the tags. Studies have been carried out recently on the relevance of ranking tags for documents and images, but to our knowledge there is no study yet to rank tags for user videos on the Web. After tag ranking, we consider linking this enriched data to the open Web following the principles envisioned in the Linking Open Data (LOD) initiative [15]. This rich data cloud gives each object and concept a unique identifier (URI) which is referenceable and linkable on the Web, such that they can make reference to each other irrespective of the vocabulary used. Three video resources may be described with three different tags “new york city”, “nyc” and “big apple” by three different users, but the intended meaning is the same, i.e. the city of New York. When we look for “new york city”, we may not find the other two even though both of them are describing the same content. If we can disambiguate these three and link to one identifier, this makes retrieval much easier. To address this problem a solution is to disambiguate each tag to an ontological concept identified by its own URI [27]. Since tags are simple uncontrolled keywords, they inherit the same IR-related

[1] http://www.youtube.com/

[2] http://www.w3.org/2008/WebVideo/Annotations/

problems of synonyms and polysemy, as described in [26] and [27]. A robust disambiguation method is needed for direct tag-to -concept matching. In the present study, we have not described the tag-to-concept matching module in much detail, but rather we have described the applicability of tag-to-concept matching and the benefits of interlinking to the structured world. The final output of the framework is a set of RDF triples describing the video and its contextual metadata with the support of a video model and various existing lightweight ontologies such as Dublin Core, SIOC, MOAT, FOAF, etc.

The rest of the paper is organised as follows. Section 2 describes various related studies in tag suggestion ranking and semantic integration. Section 3 describes the system architecture and its modules. Section 4 describes the integration of the enriched video tag space and metadata into the Linked Data Cloud. This is followed by experiments and evaluation in section 5, after which we will conclude with some remarks and future directions in the final section.

2 Related Tag Studies

Strongly descriptive and unambiguous tags are the first step toward more effective retrieval and interoperability across Social Web data sources. Much research has been carried out recently in refining user-generated tags to make them more semantically interoperable. Numerous studies [1], [8] have been carried out to suggest relevant tags for media documents based on supervised learning techniques, where the models are built for specific domains and co-relations of low-level features to tags are learned. However, due to the numerous amounts of visual variations, many efforts are far from satisfactory and moreover are restricted to a small domain of applications. Manual and collaborative tagging is one of the alternatives adopted by most popular media sharing sites such as Flickr and YouTube. This of course adds other problems as described in the first section of this paper. These problems have led to many studies in the field of folksonomies, user-tagging behaviours, semantic tagging and tag refinement. We will describe some of the studies relevant to the present study and outline how they differ from the present study. These studies mainly come under three different groupings: tag suggestion, tag ranking and tag semantics.

2.1 Tag Suggestion

In the field of tag suggestion, different but simultaneous approaches have been pursued by researchers to improve both automatic annotations and multimedia annotation quality. Researchers from the machine vision community are now focusing on gathering contextual data together with content processing to bridge the semantic gap [19], while other researchers [17] are purely focusing on social data combined with a knowledge base to augment media with social annotations[3]. Though both approaches have their valid points, harvesting social data is not only inexpensive but can contribute significantly to bootstrapping the content understanding process.

The informal nature of tagging means that semantic information cannot be directly inferred from an annotation, as any user can tag any resource with whatever strings

[3] http://acronym.deri.org/

they wish. However, studying the collective tagging behaviour of a large number of users allows emergent semantics to be derived [14]. Through a combination of such mass collaborative 'structural' semantics (via tags, geo-temporal information, ratings, etc.) and extracted multimedia 'content' semantics (which can be used for clustering purposes, e.g. image similarities or musical patterns), relevant annotations can be suggested to users when they contribute multimedia content to a community site by comparing new items with related semantic items in one's implicit and explicit networks.

2.2 Tag Ranking

Research into tag ranking began with studies [1] and [8] where ranks were assigned with respect to visual content as the result of supervised machine learning approaches, where models map relationships between visual features and semantic concepts. Uncontrolled visual content where there are a vast number of concepts involved makes the above approach less effective, and led to another approach for tag ranking which followed usage statistics by studying tag co-occurrence over a large corpus. Sigurbjörnsson et al. ranked Flickr tags [3] by means of co-occurring tags. Hotho et al. [11] suggested Folkrank for community detection in Delicious tags. Relevance ranking by means of frequency counting for "neighbouring" images (in terms of visual similarity) was conducted by Li et al. [4], where they selected common tags from neighbouring images for higher ranking. A recent study on tag ranking by Liu et al [9] proposed a tag rank for Flickr images by means of a random walk. Our ranking module is in the same general domain with the exception that we enriched our tag space before ranking to tackle the problem of tag sparsity in YouTube videos.

2.3 Tag Semantics

Studies in tag semantics fall into two broad categories: a corpus-based or statistical approach and a knowledge-based approach. Initial studies [2] on folksonomies explored means of leveraging the statistical co-occurrence relations between tags to define their semantics, and knowledge-based approaches refer to external knowledge sources such as thesaurus and ontologies to define the tag meaning [20]. Rattenbury et al. [5] explored tag-usage statistics to determine the events and place semantics from Flickr tags using burst detection analysis. Research in [6] used online ontologies and WordNet [17] to map tags for Flickr tags to concepts. Simon et al. [7] used Wikipedia categories and template structures to classify Flickr tags and these were mapped to WordNet concepts.

Other works on the topic include studies regarding the emergent semantics of tagging systems. Among others, [22] used an approach based on related co-occurrences of tags to extract hierarchical relationships between concepts, modeled in RDFS, while [23] defined a socially-aware approach for building ontologies by combining social network analysis and clustering algorithms based on folksonomies. More recently, FolksOntology [29] and FLOR [28] also provide frameworks for automated semantic enrichment of tagged data.

Finally, various models have been developed to capture the semantics of tagging systems using lightweight ontologies, such as SCOT [25], MOAT [24] or CommonTag[4].

3 System Architecture

In this section, we will give a detailed description of the tag expansion and ranking system that we have built. We begin with a general overview of the different modules, followed by an explanation of the tag filtering and expansion step, then we will describe the tag graph creation process, and finally we will detail the tag ranking methods by means of spreading activation over the tag graph.

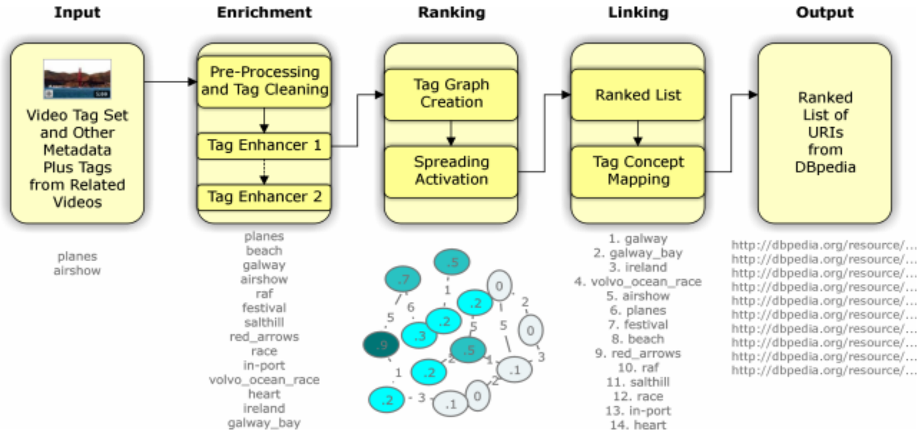

Input
Enrichment
Ranking
Linking
Output
Video Tag Set and Other Metadata Plus Tags from Related Videos
Pre-Processing and Tag Cleaning
Tag Enhancer 1
Tag Enhancer 2
Tag Graph Creation
Spreading Activation
Ranked List
Tag Concept Mapping
Ranked List of URIs from DBpedia
planes
airshow
planes
beach
galway
airshow
raf
festival
salthill
red_arrows
race
in-port
volvo_ocean_race
heart
ireland
galway_bay
.5
.7
0
1
2
6
.2
5
.2
5
5
0
.3
.9
.5
3
2
1
.1
1
.2
2
.2
3
.1
0
1. galway
2. galway_bay
3. ireland
4. volvo_ocean_race
5. airshow
6. planes
7. festival
8. beach
9. red_arrows
10. raf
11. salthill
12. race
13. in-port
14. heart
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...
http://dbpedia.org/resource/...

Fig. 1. Work flow of the tag enrichment, ranking and linking processes

The goal of this work is to enrich the user-generated tag space, and to rank and interlink the tags to DBpedia concepts for greater integration with other datasets. DBpedia is considered as a central node in the LOD cloud (the DBpedia nucleus), and linking to DBpedia also allows one to reach other datasets, thanks to the network effect of this project. There are three main modules in the system, each of which consists of many sub-modules. Figure 1 shows the normal work flow of the system: (1) context analysis and tag expansion; (2) tag ranking, and (3) concept mapping and linking to the Semantic Web.

3.1 Context Analysis and Tag Expansion

In this section, we describe our first module that implements the tag expansion strategy. Because of the sparseness of video tags, we need to expand the tag base with various other contextual sources such as social, temporal and geographical contexts.

[4] http://commontag.org/

We will begin with a description of our pre-processing step. User-generated tags consist of three broad categories of tags: functional tags (meaningful and mostly single keywords), noisy tags, and compound or emerging tags. Compound or emerging tags are those tags consisting of two or more keywords without any white space such as "friendsoftheearth", "iswc09" (used for friends_of_the_earth, ISWC_2009 respectively). There are other categories of tags which are subjective or judgmental tags, as studied by [21] and these reflect a user's view point rather than the video content, for example, "funny", "wonderful", "watch this", etc. In our system, we excluded tags with less than three characters, subjective tags, non-English tags, and tags describing usernames for the purpose of this study. However, there can be some difficulty with compound tags as these tags are not common words, and further work must be performed to identify meaningful tags from these composite sets. The textual content from the video title and descriptions are subject to the same kind of pre-processing described above, including stopword removal.

3.2 Semantic Tag Space Enrichment

In this module, we work on the tag space enrichment process where the sparse video tag space is enriched with multiple contextual sources. The sources are of various natures and exist in the context of the video in question:

1. Other textual contexts such as title and description of the video
2. Geospatial contexts, such as the place where the video has been recorded (latitude and longitude coordinates available through the YouTube API)
3. Temporal contexts, e.g. recording time
4. Social contexts, e.g. groups or playlists that include the tagged video as an item
5. Related videos, i.e. videos sharing some specific characteristics such as tags or time and space
6. User contexts, such as the type of user that includes the video in their bookmarks or favorites list
7. Context from the Web itself, i.e. other websites delivering information about these tags

We have considered the first five contextual sources to increase the tag space, and omitted the last two, which may be the subject of another study. Textual contexts such as video titles, descriptions and categories are used to rank the tag weights and sometimes add extra tags that are missing in the tag space itself. To avoid noise propagation, weights are added to different sources.

A *playlist* "extreme sailing" can include videos whose tag space is more compact and clustered from the general "sailing" tag space. Playlist and group structures where videos and users are members can propagate tags to the individual video items [18].

"Related Videos" in YouTube are those videos that are considered similar to the original video in some aspects. YouTube provides a related video feed for each video. It is not known on what basis YouTube ranks the relatedness of a video, and sometimes the results are unexpected. Moreover, YouTube feeds cannot be filtered with complex queries such as "*give me the videos related to the query where relatedness is based on a shared tag space, should be from the same place, and must be within a*

time range, but not from the same user” without a lot of work, so we decided to generate a list of related videos for each video from our own data set. The related videos are judged based on mutual content information in tag space. We adopted a space and time normalisation criteria in selecting the related videos. To explain this, if videos share a time and place value with the original video, they are ranked higher in relatedness. The intuitive explanation for this is that videos from the same place and same time are more likely to capture the same events and content [5]. Videos from Galway (a geographical area) from 30-05-2009 to 01-06-2009 are more likely to contain the events “Salthill air show” and “Volvo Ocean Race”, so accordingly there is a definitive pattern of high-frequency tags such as “Salthill”, “air show”, “red arrows”, “beach”, “Volvo Ocean Race”, etc.

Spatial context is information regarding the geolocation where the video has been recorded or the place that the content describes, which can be extracted from the geo coordinates. *Temporal context* is the time of video recording (not publishing). Table 1 shows the comparative tag spaces of related videos with and without time and space filters. This contextual information not only expands the initial tag space, but it also adds weights to the tags. The intermediate list of tags is the input for the final phase of tag expansion and recommendation based on tag co-occurrence. Table 2 shows the first phase of tag expansion.

Table 1. Comparative tag spaces of related videos with and without filters

Original Video Tags	*Related Video Tags Without Filter*
Planes, Air show, Galway	Planes, Air show, red_arrows, Volvo Ocean Race, Galway, Ireland, Panasonic, NV-GS330, NV, GS, 330, NV-GS
Original Video Tags	*Related Video Tags With Time and Space Filters*
“MOV04687”, “Galway”, “Ireland”, “air show”	“heart”, “festival”, “raf”, “in-port”, “race”, “beach”, “galway_bay”, “salthill”, “Volvo Ocean Race”, “red arrows”, “beach”

Table 2. Multi-contextual tag expansion

Title	Planes Salthill Galway	“Planes”, “beach”, “Galway”, “Air show”, “raf”, “festival”, “salthill”, “red arrows”, “race”, “in-port”, “volvo_ocean_race”, “heart”, “Ireland”, “galway_bay”
Description	same planes!!	
Tags	Planes, air show	
Related Videos	“heart”, “festival”, “raf”, “in-port”, “race”, “beach”, “ galway_bay,” “salthill”, “Volvo Ocean Race”, “red arrows”, “beach”	
Geolocation	Galway, Ireland	

Tag co-occurrence is one of the key enablers towards creating a more comprehensive semantically-related tag space for the video. Co-occurrence between two tags occurs when both the tags are used to label the same resource. We opted for a second phase of tag expansion based on tag co-occurrence if the tag set (N < 5) is less than five after the first stage of expansion. Raw co-occurrence gives a weak relationship as there may be many occurrences of less descriptive tags such as “news” for all news category videos. Therefore, it is natural to normalise the count in order to reduce the bias.

Intuitively, one resource will not be tagged with equivalent tags but rather with related tags in which case the distance between them will not be symmetrical: $d(t_1, t_2) \neq d(t_2, t_1)$. We have adopted an asymmetric approach of measuring co-occurrence using the equation:

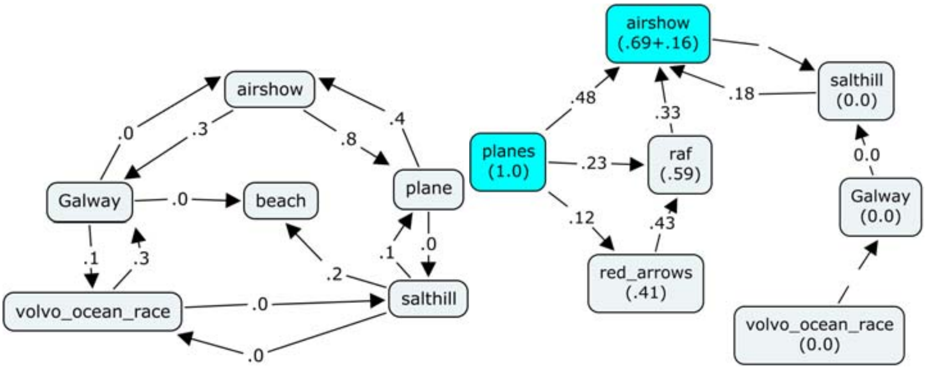
$$cd(t_1, t_2) = |t_1 \cap t_2| / |t_1| \tag{1}$$

It captures how often tag 2 (t_2) occurs with tag 1 (t_1) given the total number of occurrences. It gives a more diverse tag space when compared to a symmetric co-occurrence coefficient.

When we get a list of co-occurring tags for each of the tags from the list above we need a mechanism to aggregate them so that we can prepare the final list. Aggregation can be a simple voting mechanism where frequent candidate tags are ranked higher.

3.3 Tag Ranking

In this section we describe the detection of ranked nodes in the graph, beginning with an overview of the tag graph creation, and then describing spreading activation over the graph to rank the nodes.



(a) (b)

Fig. 2. (a) Tag graph for a video and (b) spreading activation from the node “planes”

Given a video $v \in V$ and an extended tag set $ET = \{t_1, t_2, \dots t_n\}$, we create a local graph of tags. The tag graph is a directed weighted graph with tags as nodes and the links between nodes are weighted edges. The edge weight is an asymmetric correlation based on their co-occurrence. If the correlation value is less than a threshold (τ) the tags are not connected. The co-occurrence relation is calculated as per equation 1. Figure 2 (a) shows a tag graph for the video.

We have used the tag list as a loose semantic network based on their correlations, and processed the network using spreading activation. Spreading activation as an information processing algorithm is based on the theory of cognitive science [16] and human memory. It works on a semantic network of nodes and links where links are connections between nodes based on certain relationship. Information processing starts when the activated node starts spreading its energy towards the neighboring nodes. At the end of processing, all nodes will have some activation value which was contributed through its relation with neighboring nodes.

In our work, we have used the video tags as nodes of the network and their co-occurring relations as the weighted link between nodes. The relationship between tags can be semantic as in WordNet and other ontologies, or it can be based on co-occurrence patterns as observed in web data. We have opted for co-occurrence relations to connect the tags as a network. The activation process includes the following steps:

1. The graph nodes are assigned an initial value of 0 except for the firing node which has an activation value of 1.0.
2. It spreads its activation to all nodes in its immediate neighborhood that are connected to the source node.
3. Output activation is a function of (*initial activation* + (*initial activation* * *edge strength*) * *d*), where d is a decay factor set up experimentally. In our case it is .85.
4. If the node value exceeds a threshold, we fire the node again.
5. The node activation value is the weighted sum of its contributing nodes.
6. Each node activates once in the process.

Following the above steps, we rank our extended tag list turned local graph. Experimentally we set up the decay factor to be .85 and the iteration was 1 for all the nodes as this graph is not a complex nested graph. The activation starts with the top node in the tag list. All the nodes except the firing nodes are assigned a value of 0. Once the energy propagation starts, the node spreads its energy to the connected nodes and the receiving amount of energy is a function of relationship strength and decaying energy factor. Figure 2 (b) shows the activation process starting with the node “plane” and spreading to three nodes “air show”, “raf”, “red arrows” (this node again spreads and contributes to the “air show” node).

3.4 Linked Data Creation

Once the tags are cleaned and ranked, they can then be connected to other similar and related resources. For example, there are videos of the “Volvo Ocean Race” on YouTube as well as on other media sharing sites such as Vimeo or Joost. To discover the entire spectrum we need to create a mapping mechanism from user tags to ontological resources so that they can be more connected and discoverable. This is where the Linked Open Data initiative fits in. As part of its principles, one needs to identify every entity (object, concept, event, people) with a unique web identifier or URI on the Web. Following one URI will lead to the discovery of some more related information. This simple yet powerful idea has rapidly gained momentum recently. The Linked Open Data cloud now consists of more than a hundred datasets and billions of interlinked facts as entities.

There is the question of how best to integrate user-generated videos with the existing structured LOD cloud. Automatic mapping from tags to concepts is desirable, but challenging due to multiple contexts of the concept. Studies in MOAT[5] showed that users are willing to do this manually when they realise the benefits of such an effort, for instance, if they get advanced browsing or querying features.

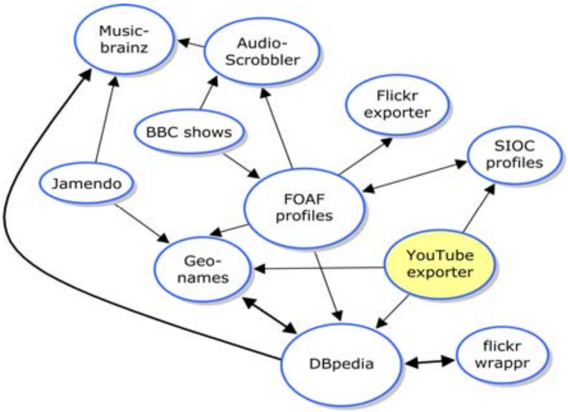


Fig. 3. Connecting to multimedia datasets already interlinked within the Linked Data cloud

In the present study, since we have a limited domain, the number of mappings is quite small so we used the semantic indexing engine Sindice[6] to query DBpedia and select the most appropriate URIs. Automatic mapping of tags to concepts is ongoing work to be reported on later. Links can also be added to user accounts (SIOC) and locations (GeoNames) obtained from the YouTube API (Figure 3).

3.5 Tags-to-Concept Mapping

Tag-to-concept matching is not yet fully implemented and will be part of our future work. As part of the experiment we have used DBpedia resources. We presume that cross-resource mapping to other sources from the LOD initiative (such as Freebase) can easily be adapted. Depending on the context, some particular datasets may also be considered, e.g. a genes database when dealing with medical videos. Here, we will briefly describe our approach for tag-to-concept matching. Once the tags are finalised, we use a two-step process for assigning concept identifiers. The tags are fed into a local WordNet module and some simple heuristics are followed:

1. If the tag matches with a WordNet noun, and if there is only one matching synset, we select the corresponding WordNet URI in DBpedia (Figure 4).

[5] http://www.moat-project.org/

[6] http://sindice.com/

2. If there are more than one WordNet synset, we send the tag and its context tags to a similarity module to compute the cosine similarity between the current tag context and already-existing tag URIs. (The similarity module is based on the Lucene[7] text retrieval Java library and on other work in progress).
3. For those tags that are not part of WordNet, we send them to the semantic indexing engine Sindice to look for resources. Once we get the top *k* URIs for the query, the user can select the URI manually or else it is fed into another disambiguation module where URIs can be contextually disambiguated (not implemented yet).

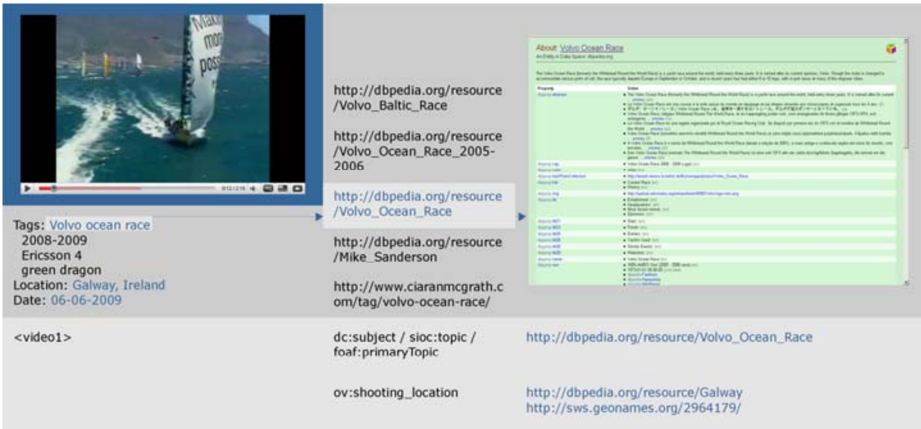

Tags: Volvo ocean race
2008-2009
Ericsson 4
green dragon
Location: Galway, Ireland
Date: 06-06-2009
http://dbpedia.org/resource/Volvo_Baltic_Race
http://dbpedia.org/resource/Volvo_Ocean_Race_2005-2006
http://dbpedia.org/resource/Volvo_Ocean_Race
http://dbpedia.org/resource/Mike_Sanderson
http://www.ciaranmcgrath.com/tag/volvo-ocean-race/
About: Volvo Ocean Race
<video1>
dc:subject / sioc:topic / foaf:primaryTopic
http://dbpedia.org/resource/Volvo_Ocean_Race
ov:shooting_location
http://dbpedia.org/resource/Galway
http://sws.geonames.org/2964179/

Fig. 4. Matching YouTube tags to DBpedia concepts

4 Experiments and Evaluation

4.1 Results

We have collected 3,990 YouTube videos. All video metadata including the metadata of related videos was collected through the YouTube API. We collected videos of specific categories such as “skiing”, “sailing” and “cricket”. The data includes video tags, dates, places (if available), titles, descriptions and group tags (if available). The total number of unique tags is more than 11,900 which includes many misspellings, number tags, co-joined tags, and subjective as well as meaningless tags. There are 2,261 distinct users in the data set. On average, one user has less than two videos. Since users tag differently depending on their background and expertise, we can assume a relatively heterogeneous tag source. We did a preliminary filtering of tags by removing stop words, tags with two characters and number tags. Though the tag list is far from clean, this reduces a lot of noise. This tag set is used for extracting the co-occurrence statistics.

[7] http://lucene.apache.org/java/docs/

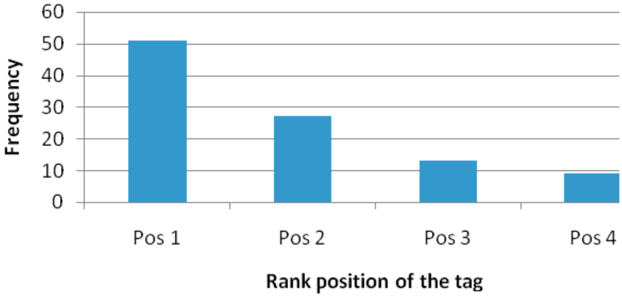


Fig. 5. The number of times the most relevant tag was suggested at different positions

We conducted a preliminary evaluation to explore the quality of our ranking method and tag enhancement. We randomly selected 100 videos from the larger set to explore potential benefits and problems. Three users familiar with the topics were asked to rank the tag lists of these videos on a scale of 1 to 4: “most relevant”, “relevant”, “partially relevant” and “irrelevant”, according to their depicted content. We computed the amount of times the most relevant tag was ranked as “most relevant” by users, and we will now discuss the quality of our tag enhancement.

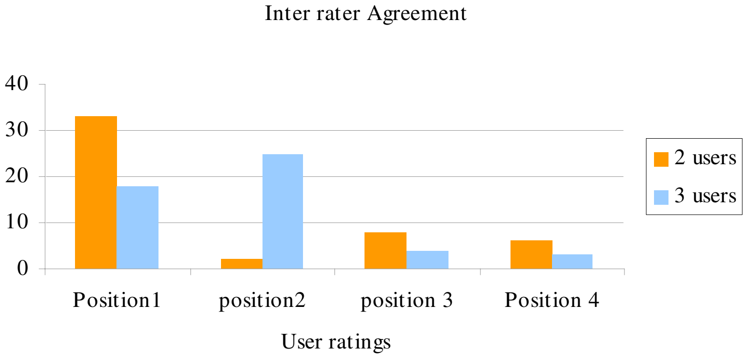


Fig. 6. Inter-rater agreement over the most relevant tag in the first four positions

Though relatively small, the user feedback gave some interesting results. Regarding inter-rater variation, we considered a final rank when a minimum of two users agreed on the same rank. For the 100 videos, Figure 5 shows that the most relevant tag came in at the top position 51 times (where a minimum of two users have agreed), whereas the top tag came second 27 times. Therefore, for almost 80% of the time, the top-ranked tag was in either of the first two positions. Figure 6 shows the inter-rater agreement over the tag relevance in the first four positions.

Similarly, we tried to explore the tag enrichment task and its effectiveness in describing the content of the video. This evaluation was conducted at two stages: before

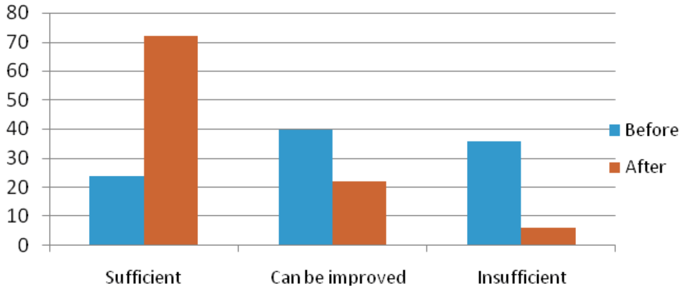


Fig. 7. Comparative evaluation of the quality of the tag space

the tag expansion and after the tag expansion. The users were asked to rank the original tag list on a three-point scale of (1) sufficient for the content, (2) okay, but can be improved, and (3) insufficient. The comparative results are in Figure 7.

The result showed that the tag enrichment process increased the content understanding considerably. Still, 28% of the videos need improvement. There may be many reasons for this such as more specific tag cleaning, insufficiency of the tag list, or perhaps due to noise propagation from the different contextual sources which were used.

In the present study, inter-rater agreement evaluation is only focused on the top-ranked tag. Out of the 51 times where the system-suggested top tag was considered most relevant by users, two users agreed 33 times and three users agreed 18 times. However, in position two, out of 27 total times, all three users agreed that the tag was most relevant 25 times. An exhaustive analysis of user agreement over the relevance of suggested tags is desirable to explore possible room for improvements.

We will conclude this section by discussing some potential benefits in applications such as (1) semantic tag-based search and retrieval, and (2) improved video categorisation.

4.2 Semantic Tag-Based Search and Retrieval

We evaluated our work in a retrieval framework and describe here two use cases of tag-based retrieval. Given a query *q*, the system will retrieve all the videos tagged with *q* but they will be ranked according to the ranked position of *q* in the tag space. Thus, if two videos tagged with *q* are retrieved, the video having *q* in a higher position will be ranked higher. Identifying tags with DBpedia URIs opens up many possibilities of knowledge discovery. A resource tagged with "Volvo Ocean Race" and identified with a URI such as *"http://dbpedia.org/resource/Volvo_ocean_race"* will lead us to discover information about "St._Petersburg, Russia" as it is related to the query by means of destination port. A video tag identified with *"http://dbpedia/resource/Galway"* will lead us to discover more about the culture, events and history of Galway.

4.3 Improved Video Categorisation

Categories in YouTube are selected by users when uploading videos. Sometimes, a user selects a less-relevant category for their content as it is a flat single category system and the choices are also quite limited. In practice, video content may belong to more than one category, and moreover, it may follow a hierarchal structure. Based on our enriched tag space, we can suggest a hierarchical categorisation for a video by exploiting the relations between tags. A video tagged with “news, Japan, earthquake, building, tsunami” can be categorized under News >> Natural Hazard >> Earthquake conforming to the hierarchical structure of existing ontologies such as the Large Scale Concept Ontology for Multimedia (LSCOM) [13]. In LSCOM, “earthquake is a subclass of natural hazard”.

5 Conclusions

In this paper, we propose a roundtrip semantic framework which provides some steps towards solving the above problem. The key modules that have been implemented are for tag enrichment, tag ranking, concept mapping and semantic linking. Tag enrichment involves various contextual analyses of the video and the contribution of these contexts towards content understanding. Context not only includes textual descriptions but also the temporal, spatial and social contexts in which the video is being used and shared. Interplaying a combination of contexts provides for improved enrichment. The advantages of the proposed algorithm for tag enrichment are: (1) multiple sources can make the tags more reliable for content description; (2) the subjective ambiguities are reduced; (3) the method is scalable since it does not require any domain specific model training; and (4) it can evolve with tag usage.

We have also proposed a tag-ranking algorithm performed over a local tag graph by means of spreading activation. The tag graph is based on the enriched tag set and is connected by means of co-occurrence strength. Spreading activation helps to activate the focused nodes and reduces the strength of noisy nodes. We also described the tag-to-resource mapping (with DBpedia as our semantic repository) and outlined how Linked Data principles can aid with linking to user-generated video content. Successful linking to DBpedia web identifiers can harness other related data sources from the LOD cloud and then enrich information discovery from videos.

Our future work includes a complete concept-to-URI matching mechanism and an explorative evaluation of the approach with more users.

Acknowledgements

This work was supported by Science Foundation Ireland under grant number SFI/08/CE/I1380 (Líon 2).

References

1. Barnard, K., Duygulu, P., Forsyth, D., Freitas, D.N., Blei, D.M., Jordan, M.I.: Matching words and pictures. JMLR, 1107–1135 (2003)
2. Begelman, G., Keller, P., Smadja, F.: Automated tag clustering: Improving search and exploration in the tag space. In: WWW Collaborative Web Tagging Workshop (2006)
3. Sigurbjörnsson, B., van Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proc. of the International World Wide Web Conference (2008)

4. Li, X.R., Snoek, C.G.M., Worring, M.: Learning tag relevance by neighbour voting for social image retrieval. In: Proc. of the ACM International Conference on Multimedia Information Retrieval (2008)
5. Rattenbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from Flickr tags. In: Proc. of SIGIR, pp. 103–110 (2007)
6. Angeletou, S., Sabou, M., Motta, E.: Semantically enriching folksonomies with FLOR. In: Proc. of the 5th ESWC: CISWeb, Tenerife, Spain (2008)
7. Overell, S., Sigurbjörnsson, B., van Zwol, R.: Classifying tags using open content resources. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM 2009), ACM, Barcelona, Spain (2009)
8. Li, J., Wang, J.Z.: Real-time computerized annotation of pictures. IEEE Transactions on Pattern Analysis and Machine Intelligence (2008)
9. Dong, L., Xian-Seng, H., Yang, L.: Tag ranking. In: Proc. of the International World Wide Web Conference (2009)
10. Hunter, J.: Adding multimedia to the Semantic Web – Building an MPEG-7 ontology. In: 1st International Semantic Web Working Symposium (SWWS 2001), California, USA, pp. 261–281 (2001)
11. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information retrieval in folksonomies: search and ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006)
12. Wu, X., Zhang, L., Yu, Y.: Exploring social annotations for the Semantic Web. In: Proc. of the International World Wide Web Conference (2006)
13. Naphade, M.: Large-scale concept ontology for multimedia. IEEE Multimedia 13(3), 86–91 (2006)
14. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 522–536. Springer, Heidelberg (2005)
15. Bizer, C., Cyganiak, R., Heath, T.: How to publish Linked Data on the Web (2007), `http://www4.wiwiss.fu-berlin.de/ bizer/pub/LinkedDataTutorial/`
16. Quillian, M.R.: Semantic memory. In: Minsky, M. (ed.) Semantic Information Processing, pp. 227–270. MIT Press, Cambridge (1968)
17. WordNet, `http://wordnet.princeton.edu/` (last accessed March 12, 2009)
18. Celma, O., Ramírez, M., Herrera, P.: Foafing the music: A music recommendation system based on RSS feeds and user preferences. In: Proc. of the 6th International Conference on Music Information Retrieval (ISMIR 2005), London, UK, pp. 464–457 (2005)
19. Oge, M., Lux, M.: An exploratory study on joint analysis of visual classification in narrow domains and the discriminative power of tags. In: Proc. of the 2nd ACM workshop on Multimedia semantics, Vancouver, British Columbia, Canada (2008)
20. Schmitz, P.: Inducing ontology from Flickr tags. In: Proc. of the Collaborative Web Tagging Workshop at the International World Wide Web Conference, Edinburgh, UK (2006)
21. Golder, S., Huberman, B.A.: The Structure of Collaborative Tagging Systems. Journal of Information Sciences 32(2), 198–208 (2006)
22. Halpin, H., Robu, V., Shepard, H.: The Dynamics and Semantics of Collaborative Tagging. In: Proceedings of the 1st Semantic Authoring and Annotation Workshop (SAAW 2006), The 5th International Semantic Web Conference (ISWC 2006), Athens, Georgia, USA (2006)

23. Mika, P.: Ontologies are us: A unified model of social networks and semantics. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 522–536. Springer, Heidelberg (2005)
24. Passant, A., Laublet, P.: Meaning Of A Tag: A Collaborative Approach to Bridge the Gap Between Tagging and Linked Data. In: Proceedings of the Linked Data on the Web Workshop (LDOW 2008) at the 17th International Conference on the World Wide Web (WWW 2008), Beijing, China (April 2008)
25. Kim, H.L., Yang, S.K., Breslin, J.G., Kim, H.G.: Simple Algorithms for Representing Tag Frequencies in the SCOT Exporter. In: Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, Fremont, California, USA, pp. 536–539 (2007)
26. Mathes, A.: Folksonomies: Cooperative Classification and Communication Through Shared Metadata. Computer Mediated Communication, LIS590CMC, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign (2004)
27. Passant, A.: Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs: Theoretical background and corporate use-case. In: ICWSM 2007 (2007)
28. Angeletou, S.: Semantic Enrichment of Folksonomy Tagspaces. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 889–894. Springer, Heidelberg (2008)
29. Van Damme, C., Hepp, M., Siorpaes, K.: Folksontology: An integrated approach for turning folksonomies into ontologies. In: Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), pp. 57–70 (2007)