Exploiting User Comments for Audio-visual Content Indexing and Retrieval

Carsten Eickhoff[1], Wen Li[1], and Arjen P. de Vries[2]

[1] Delft University of Technology, Delft, The Netherlands,
{c.eickhoff,wen.li}@tudelft.nl
[2] Centrum Wiskunde & Informatica, Amsterdam, The Netherlands,
arjen@acm.org

Abstract. State-of-the-art content sharing platforms often require users to assign tags to pieces of media in order to make them easily retrievable. Since this task is sometimes perceived as tedious or boring, annotations can be sparse. Commenting on the other hand is a frequently used means of expressing user opinion towards shared media items. This work makes use of time series analyses in order to infer potential tags and indexing terms for audio-visual content from user comments. In this way, we mitigate the vocabulary gap between queries and document descriptors. Additionally, we show how large-scale encyclopaedias such as Wikipedia can aid the task of tag prediction by serving as surrogates for high-coverage natural language vocabulary lists. Our evaluation is conducted on a corpus of several million real-world user comments from the popular video sharing platform YouTube, and demonstrates significant improvements in retrieval performance.

1 Introduction

In recent years, content sharing platforms have become very popular. In particular, video sharing platforms have experienced massive growths in both, the amount of shared content as well as the number of viewers. A recent survey attributed YouTube as being solely responsible for approximately 10% of the global Internet traffic [5]. Content sharing services typically enhance the publishing and distribution of pieces of media by social networking features such as friend relationships, messaging, collaborative tagging and commenting functionalities. In order to make content available to the users, most state-of-the-art content sharing platforms rely on tagging. The step of assigning tags, however, is often left to the user community.

While there are users who relish this task, and some platforms even integrate it into games to make it more entertaining, there are many who regard it as a rather tedious burden. Ames and Naaman [2] studied user tagging behaviour and found that a frequently expressed motivation for tagging lies in the necessity to do so in order to make the content available to the user base. Additionally, they noted a significant share of tags to be strongly dependent on the tagger's socio-context, rendering them less useful for users that do not share the same

context (i.e., friends, place of residence, cultural background).
To overcome this challenge in related domains, automatic tagging mechanisms have been proposed that extract keywords from textual meta data and content. In the case of shared multimedia content, however, this is often not feasible with satisfying precision, as meta data can be sparse or ambiguous and concept detection from audio-visual signals is still considered more difficult than text-based alternatives [14]. For example, many videos on YouTube feature only a title and a brief textual description. Statistical tag prediction approaches face significant problems when operating in such resource-impoverished domains.
Commenting, on the other hand, appears to be a more natural activity for most users. We can observe extensive threads of comments related to shared media items. In this work, we propose the use of time series analyses for audio-visual content with sparse meta data. The investigation is not targeted towards the actual content and meta data but will focus exclusively on people's comments towards the content. To this end, we employ a language modelling approach to utilise the naturally created community information on content sharing platforms, to infer potential tags and indexing terms. In this way, we aim to mitigate the vocabulary gap between content and query. In the past, the usefulness of user comments for retrieval tasks was frequently doubted due to the high proportion of noise in the chat domain [17]. However, given the large scale at which user comments are currently available, we will show that informed means of interpreting noisy natural language communication streams as well as aggregation with orthogonal types of (social) media can help to identify valuable pieces of information in the abundant underlying noise. The novel contributions of our work are threefold: (1) We apply a language modelling method for tag prediction of sparsely annotated multimedia content from potentially very short and noisy user comments on the Web. (2) We demonstrate the use of time series analyses to further exploit the inherent structure of natural language conversations. (3) We inspect independent sources of evidence from the Web, such as Wikipedia, in order to further improve tag prediction results.
The remainder of this work is structured as follows: After an overview of related work in Section 2, we describe a time series analysis scheme for resource filtering prior to the tag predicition step (Section 3). In particular, we make use of the online encyclopaedia Wikipedia as a surrogate for natural language vocabulary lists to enhance tag prediction performance. In Section 4, we demonstrate the merit of our method on a real-world data sample of several million user comments collected from YouTube. Section 5 revisits the outcomes of the study and discusses salient examples as well as key challenges and opportunities for both practical application and future research. Section 6 closes with a concluding overview of future directions.

2 Related Work

While tag prediction from short, noisy user communication has not been extensively studied, there are several prominent methods for keyword extraction

directly based on content. Hu et al. introduced a graph-based method for discussion summarisation through sentence extraction from weblog posts [11]. Budura et al. [4] propagate tags along the edges of a Web page similarity graph that is built based on a range of content features. Matsuo et al. [15] present an approach of extracting keywords from single documents without the need for a background corpus. Using intra-document term distributions, the authors report performances that approximate those of *tf/idf*-based methods. Wartena et al. [22] propose to infer keyword candidates from the semantic relationships between terms in academic abstracts and BBC news stories. Tomokiyo et al. [21] present a language modelling approach to keyword extraction from longer coherent news articles. Their use of the divergence between term frequency distributions is based on an intuition similar to our method. Due to the high amount of noise in user comments, additional steps are required to successfully apply their method in this domain. To this end, we apply time series analyses to identify informative comments. Amodeo et al. investigated temporal relationships between time of publication of blog posts and their probability of relevance [3]. The authors employ a notion of activity bursts similar to the one proposed in this work. However, where their approach applies time series analyses directly to documents in order to prune the list of pseudo relevant results, we aim to improve the general indexing quality by broadening the document vocabulary.

Tag prediction is most prominently used to describe pieces of textual content, as semantic concepts can be conveniently observed in the form of term occurrences. However, there are several pieces of work dedicated to predicting tags directly from multimedia content. Eck et al. [6] present an approach of predicting tags from the audio signal of music pieces. Similar approaches for other types of media include Siersdorfer's automatic video tagging method which propagates tags across videos containing redundant or similar content [20], or Wu et al.'s photo tagging scheme [23].

While the previously discussed publications concentrate solely on extracting tags from actual content, we can identify a body of work that makes additional use of community-created information. As an example, Mishne et al. first employed user comments to enhance weblog retrieval [16]. Heymann et al. [9] predict tags from a range of local Web page features enriched by information from social bookmarking services. In 2009, Yee et al. [24] presented a method of improving search performance by utilising user comments by means of a *tf/idf*-based method. Most recently, Filippova et al. employ user comments to aid content classification performance [8]. The promising results achieved by previous work support the feasibility of our goal: Describing content exclusively based on user comments. We will employ statistical language models aided by time series analyses and external web resources such as Wikipedia, to find potential index terms and evaluate their quality in a series of TREC-style experiments.

3 Comments as Bursty Streams

Common methods for characterising individual documents d within a collection C are often based on the intuition that some terms will occur more frequently locally in d than in the collection-wide average. This notion is for example expressed in the popular tf/idf family of formulae but is also implicit in the language modelling framework [10]. The same method can be applied to the video retrieval setting, in which each shared video corresponds to a distinct d. We assume a unigram collection model LM_C comprised of all comments in C and dedicated document models LM_d based on the comment thread of document d. Subsequently, we assume good descriptors of d can be determined by the term-wise KL-divergence between both models (LM_C and LM_d), identifying locally densely occurring terms w (those that display a high negative value of $KL(w)$).

$$KL(w) = P(w|d)\ log\frac{P(w|d)}{P(w|C)} \tag{1}$$

This method has been applied for a wide number of settings and is known for its robustness and generalizability [21]. The domain at hand, however, imposes a number of specific challenges on automatic keyword extraction. There are several sources of comment noise that require appropriate treatment.

Firstly, there is a significant share of comments that are uninformative for the task of keyword extraction, either because they are off-topic (spam) or because they simply do not convey much meaning (e.g., "Cool."). In order to address this type of messages, we introduce a resource selection step that identifies informative comments based on Kleinberg's burstiness criterion [13]. When analysing the usage statistics of his personal email account, Kleinberg noticed that his incoming email was subject to sudden, typically short, peaks of activity. A first investigation in the domain of shared Web videos showed that most comment threads (98%) display the same peaking behaviour.

These so-called *bursts* can be related to external triggers such as a famous musician winning an award, causing a sudden increase of attention and commenting activity on his music videos. Often, however, the trigger is of internal nature, e.g., caused by controversial comments that spark an avid discussion. This latter class of triggers lets us assume that comments submitted within an activity burst may be more informative than regular ones. We formulate a variation of Kleinberg's original burst detection scheme to better fit the notion of threaded chat communication: We consider each coherent sequence of messages $m_i...m_j$ with inter-comment intervals $\delta_t(i, i+1)$ shorter than a threshold value δ_t as candidate bursts. In this work, we set δ_t to be the median time between comments for each document, however, further tuning of this parameter could prove beneficial. In order to select informative bursts, we apply a burstiness function $b(i, j)$, according to which we rank all candidates. The underlying intuition is that a "good" burst should cover many comments in as little time as possible. This is represented by $length_{rel}(i, j)$, the relative share of comments contained in the burst, divided by $\delta_{rel}(i, j)$, the relative amount of time for which the burst lasted. Consequently, we pool all comments from the n highest-ranked bursts to

train LM_d. This filtering step eliminates a significant proportion of unrelated "background noise" comments from the modelling step.

$$b(i,j) = \frac{length_{rel}(i,j)}{\delta_{rel}(i,j)} \tag{2}$$

3.1 Modelling Burst Causality

Considering the merit of using bursty comments, and assuming them to be triggered within the comment stream, we further suspect that the event triggering the increased commenting activity may be of import as well. In order to verify this hypothesis, we use a history of h comments immediately preceeding each burst as an alternative resource. Manual qualitative investigations showed an optimum in extracted tag quality at $h = 7$ history comments preceeding each burst.
In order to harmonize the evidence from pre-burst histories and actual bursts, we turn to the simplest setting of Ogilvie's method for language model combination [18]. Instead of directly estimating the probabilities of observing given terms from the whole comment thread, we now use a weighted combination of two such models. $P_B(w|D)$ is based on the maximum likelihood estimate of term occurrence according to the comments within bursts. $P_H(w|D)$ is based on the 7-comment pre-burst history. The mixture parameter λ determines the relative importance of burst comments over history comments. Higher values of λ give more weight to comments within the bursts.

$$P_{HB}(w|D) = \lambda P_B(w|D) + (1-\lambda)P_H(w|D) \tag{3}$$

In order to assess tag extraction quality, we randomly sampled 50 videos from YouTube, applied our four tag prediction methods (based on the entire comment thread, on bursts, on pre-burst histories, and, on the burst/history mixture) and measured the overlap of the respective with the gold standard tags as assigned by YouTube users. Figure 1 shows tag prediction performance as we vary the composition of the model mixture. Best results could be achieved for settings of $\lambda = 0.65$. Language models trained on the entire comment thread resulted in an F_1 score of 0.061, significantly below any of the compared settings in Figure 1 (tested using Wilcoxon Signed rank test with $\alpha < 0.05$).

3.2 Wikipedia as a Surrogate for Natural Language Vocabulary

Previously, we addressed noise in the form of unrelated and uninformative comments within the thread. The second source of noise are misspellings, abbreviations, chatspeak and foreign language utterances, all of which are frequently encountered in on-line chat communication. To address this, we use the online encyclopedia Wikipedia for regularization. We formally introduce the $\eta(w)$ criterion. Terms w that do not have a dedicated article in the English version of

Wikipedia are assumed to be noise and, subsequently, rejected from the list of candidate terms. Due to Wikipedia's high coverage, the number of false positives, valid terms rejected by this filter, has been found to be negligible.

$$\eta(w) = \begin{cases} 1 & \text{if } w \text{ has an English Wikipedia article,} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

With our noise filtering components in place, our final extraction scheme is:

1. Within a comment thread d, find all message sequences (so-called bursts) with inter-comment intervals no longer than δ_t.
2. Rank the bursts according to their burstiness $b(i, j)$ (Eq. 2) and keep top n.
3. Train LM_d on the previously selected most bursty comments (Eq. 3).
4. Rank all terms w according to (Eq. 1).
5. Return top k terms $w_1 ... w_k$, rejecting all w with $\eta(w) = 0$.

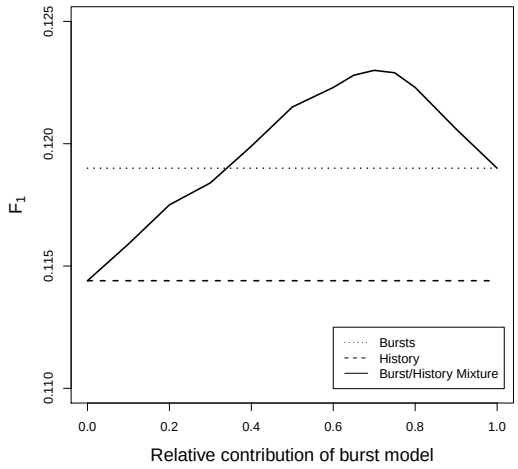

Fig. 1. Performance of burst-history mixture models for varying weights.

4 Evaluation

Previously, we investigated our method's performance at replicating the gold standard labels assigned by YouTube users. Ultimately, however, we aim to improve retrieval performance of shared video content by extracting representative terms *a priori* at indexing time. In this way, we can enrich sparsely annotated content (e.g., in the audio-visual domain) by harnessing community knowledge in the form of user comments.

Our evaluation dataset is comprised of 4.7 million user comments issued towards

more than 10.000 videos. It was collected between December 2009 and January 2010. The crawling process was limited to textual information, omitting the actual audio-visual content, and was started from a diverse selection of manually formulated seed queries, following the "related videos" paths. On average, every video in this collection has 360 ($\sigma = 984$) dedicated user comments and 14 tags ($\sigma = 11.8$) assigned to it. The only source of textual meta information are titles and video descriptions provided by the uploader.

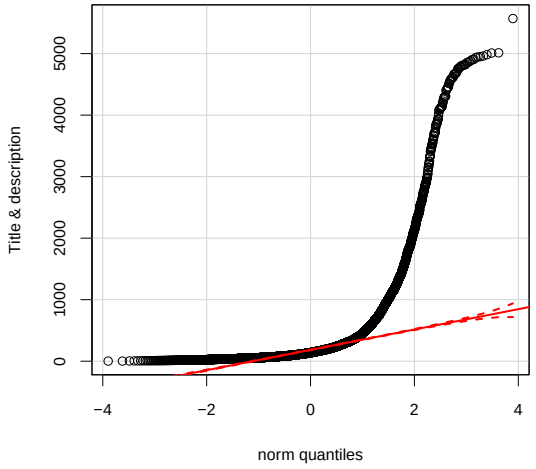

Fig. 2. Quantile distribution of YouTube video meta data length in characters.

To evaluate our method's merit at indexing time, we conduct a TREC-style retrieval experiment. We use the Lucene search engine library (`http://lucene.apache.org/`) and a BM25F retrieval model [19]. We manually designed a set of 40 topics that are well represented in our collection (e.g., *"Lady Gaga Music Video"* or *"Swine Flu 2009"*). Finally, we obtained binary relevance judgements for the top 10 retrieved results per query via crowdsourcing. On average, 36 results per query were evaluated. [1] describes a similar setting for collecting pairwise query/document judgements, concluding that a group of untrained workers can produce relevance judgements of a quality comparable to that of a single domain expert. As a consequence, we collected 10 redundant binary judgements per unique topic/video pair and aggregate the results in a majority vote. The task was offered on the crowdsourcing platform Amazon Mechanical Turk (AMT) at a pay level of 2 cents per single judgement, as suggested by previous studies [7]. In order to ensure result quality, we employ gold standard judgements as well as honey pot questions as recommended by [12]. Our method's parameter settings were determined on the first 5 topics of our data set by varying the number of most intense bursts, n, and the number of terms extracted per video, k. These training queries were not used further for evaluation. The best observed performance could be achieved at $n = 20, k = 15$. Table 1 compares

the retrieval performance of various BM25F indexes, using either only original meta information, extracted terms, or combinations of both. We measure result quality in terms of *Mean Reciprocal Rank* of first relevant results (MRR), *Mean Average Precision* (MAP) as well as precision at rank 10 (P@10). In a baseline performance run, we rely exclusively on video titles and textual descriptions, each of which becomes an individual field in the retrieval model's index. This is comparable to the information based on which the standard YouTube search API operates (`https://developers.google.com/youtube/2.0/developers_guide_protocol_api_query_parameters#qsp`). Unless stated differently, all experiments were conducted on the full video corpus regardless of the number of comments per video. Statistically significant performance improvements over the baseline are denoted by the asterisk character (tested using a Wilcoxon signed rank test at $\alpha = 0.05$-level). In a second experiment, we exclusively use the top $k = 15$ terms extracted by our method to form the index. We can note a significant and consistent improvement over the original index's retrieval performance. When combining extracted terms and original meta data by interleaving a pool of k terms from both source selections, we experience another significant performance gain. Indexing the full comment thread alongside the original meta data introduces a high number of false positives, ultimately hurting retrieval performance. As a point of comparison, we include runs for extracted terms based solely on bursts (not using the pre-burst history), as well as those not using Wikipedia regularization. In both cases, we note performance drops as compared to the regularized mixture setting.

Table 1. Retrieval performance on shared video content.

Index type	MRR	MAP	P@10
Title & description	0.81	0.48	0.46
k extracted terms	0.85*	0.52*	0.51*
k extracted terms (bursts only)	0.80	0.49	0.46
k extracted terms (no regularization)	0.63	0.32	0.25
k random comment terms	0.08	0.03	0.05
Title, description & extracted terms	**0.89***	**0.67***	**0.64***
Title, description & full comment thread	0.48	0.33	0.34

The domain at hand is particularly challenging, since a high percentage of videos is annotated only sparsely. Our investigation shows that both titles and descriptions contain only small amounts of text (titles have an average length of 32.8 ($\sigma = 12.8$) characters, and, descriptions average at 211 ($\sigma = 220$) characters each). Figure 2 shows the quantile distribution of video description lengths in our data sample. A significant percentage (58%) of videos in our corpus is described with no more than 140 characters each. This represents the same amount of information that could be conveyed in a single tweet. For video titles, we observed a similar behaviour with more than 50% of all titles being shorter than 35 characters. In combination, this lack of explicit content annotation may hinder

successful retrieval. In order to confirm this assumption, we repeat the retrieval experiment and restrict the corpus to those videos that are sparsely annotated. More concretely, we index only those videos that feature either less than 35 title characters *OR* less than 140 description characters. The resulting set contains 7840 videos, an equivalent of 77% of the original collection.
Table 2 details the performance of the previously-introduced indexes when textual information is sparse. We can see that performance scores are consistently lower, while the performance-based ranking of approaches remains the same. However, the difference in performance between comment-based and exclusively meta data-based indexes becomes more expressed. Again, we can note a clear merit of using burst / and pre-burst information, as well as Wikipedia regularization. In conclusion, we observe significant performance improvements across all experimental settings when applying keyword extraction to user comment threads for the task of video retrieval on online content sharing platforms such as YouTube.

Table 2. Retrieval performance for sparsely annotated content.

Index type	MRR	MAP	P@10
Title & description	0.74	0.41	0.35
k extracted terms	0.79*	0.44*	0.39*
k extracted terms (bursts only)	0.75	0.38	0.33
k extracted terms (no regularization)	0.56	0.25	0.27
k random comment terms	0.08	0.04	0.05
Title, description & extracted terms	**0.82***	**0.63***	**0.59***
Title, description & full comment thread	0.41	0.31	0.25

5 Discussion

The previous sections detailed concrete, task-driven performance evaluations of our method. In this section, we will dedicate some room to lessons learned and will discuss several observations that could not be confirmed to be statistically significant but yet deserve attention as they may become more salient in related applications or domains.
In order to give qualitative insights into comment-based keyword extraction, let us visit an example that we encountered during the manual inspection of extraction results on the YouTube dataset and that is representative for a large number of cases. The video in question shows scenes from a Mafia-related computer game followed by several action film shooting scenes. While the original title ("Mafia Shootout") and description ("Mafia members in a huge shooting.") are very brief and uninformative, the results of our term extraction method show convincing tendencies. The highest-ranked term was "Mafia", which, considering that we do not peek into the actual meta information of the video, is a very good

match. Subsequent ranks contained further unsurprising terms such as "shoot" or "gun". The interesting matches, however, were "Corozzo" and "Guarraci", referring to Joseph "Jo Jo" Corozzo, Sr. and Francesco "Frank" Guarraci, two infamous criminals. Additionally, the term "Mississippi" ended up on a high rank. At first we considered it a false positive, before looking more deeply into the matter and discovering the Dixie Mafia, an organization that heavily operated in the southern U.S. states in the 1970s. Considering this example, we can see how comment-based keyword extraction manages to discover novel aspects of a topic rather than exclusively sticking to the literal content of a video item. The general observation was that our method often picks up very specific topical aspects of a given piece of content. As a consequence of relying on locally densely occurring terms, we discover "Guarraci" rather than "criminal".

One particular application that became obvious throughout the course of our research is using term extraction from comments as a means of summarizing the discussed content. When manually inspecting the output of our methods, we arrived at the impression that the set of top-ranked keywords was sufficient to convey a reliable description of the content itself. We aim to further confirm this notion and determine the method's merit for content summarisation in a dedicated series of future experiments.

In this work, we investigated the usefulness of user comments for two tasks, (1) reproducing the user-assigned YouTube tags without using any form of video-related meta information, and, (2) Improving retrieval performance of shared videos by expanding the index by terms extracted from user comments. In the future, it would be interesting to evaluate the degree to which our findings generalize to different domains and media types.

The step towards alternative media is not assumed to introduce significant changes to the method since we did not make any assumptions on the content other than the existence of time-stamped user comments. Therefore, our method should be conveniently portable to platforms such as Flickr (images) or Last.fm (music). A more challenging but also potentially more interesting generalization step could be taken to explore novel domains besides shared and commented media content. Examples of this include the Blip.tv corpus used for the Mediaeval benchmarking initiative (`http://www.multimediaeval.org/`). This corpus consists of tweets that contain links to shared videos. The data structure looks initially similar to our YouTube setting. Both have short textual messages dedicated to a given piece of content, we expect significant new challenges as a consequence of less pronounced causal relationships among tweets.

Finally, we would like to address several logical extensions to this work. There are four major directions that we aim to address in the future: (1) Rather than using Wikipedia as a single source of external evidence, an aggregate of different sources should be explored. For example, collaborative bookmarking services such as del.icio.us may contain valuable information that could be used for annotation. Candidate term regularization might benefit from a broader multi-external-source architecture. (2) Currently, our keyword extraction methods are exclusively comment-based and do not take into account the actual content or

meta information that is being annotated. While this makes for an elegant and challenging setting for this initial study, a more industrial approach should aim for a better exploitation of potential synergies between content, meta data and comments. A first step towards this end would be to not only regularize in terms of general existence/ non-existence of certain terms but also in terms of their likelihood of being related. Such a setting could determine a measure of relatedness along which to score extraction terms and to create a probabilistic framework of term inclusion. Candidate measures of conceptual relatedness include the frequency of co-occurrence between terms in the content/gold standard keywords and potential keywords or their distance in an ontology such as WordNet. (3) The proposed method currently could face cold-start issues for very new videos that have not been commented on. In order to address this problem, we propose employing a smoothed mixture model of original meta information that gradually is enriched by more community-based tags as the volume of comments increases. (4) Content sharing platforms with a high coverage typically contain a multitude of languages. In order to succumb this challenge we would like to further study the potential cross-language applicability of our method by using resources such as Wikipedia that are assumed to easily bridge the language gap.

6 Conclusion

In this work, we investigated the potential use of user comments for indexing purposes on content sharing platforms such as YouTube. We found that it was possible to deduce meaningful tag candidates from comment streams without using any form of direct annotations such as titles or video descriptions. Results improved significantly when incorporating time series analyses to identify informative regions in the discussion. We were able to benefit from external resources such as Wikipedia by using them to reduce the background noise of the chat domain. After a series of experimental runs against a set of gold standard tags, we confirmed the usefulness of the extracted terms for retrieval purposes in a sizeable TREC-style experiment based on several million user comments. We showed, that including only a high-precision set of tags extracted from user comments achieves better retrieval performance than either ignoring comments altogether or indexing the full comment stream.
Future directions based on this work should include an inspection of additional domains that may benefit from the proposed method (e.g., shared pieces of music, images, tweets or SMS), further exploitation of external resources such as collaborative tagging services and a stronger utilisation of the available content meta information. Additionally, it would be interesting to investigate means of further incentivising commenting on shared content. This could for example be done by means of community-powered games with a purpose or a reputation concept that more directly reflects the quantity and quality of comments contributed by a given individual. A growing volume of comments, especially for new and niche videos would greatly facilitate content indexing, and, subsequently, retrieval performance.

References

1. O. Alonso and S. Mizzaro. Can we get rid of trec assessors? using mechanical turk for relevance assessment. In *SIGIR 2009 Workshop on the Future of IR Evaluation*.
2. M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *SIGCHI 2007*.
3. G. Amodeo, G. Amati, and G. Gambosi. On relevance, time and query expansion. In *CIKM 2011*.
4. A. Budura, S. Michel, P. Cudré-Mauroux, and K. Aberer. Neighborhood-based tag prediction. *The Semantic Web: Research and Applications*, 2009.
5. X. Cheng, C. Dale, and J. Liu. Understanding the Characteristics of Internet Short Video Sharing: YouTube as a Case Study. *ArXiv e-prints 2007*.
6. D. Eck, P. Lamere, T. Bertin-Mahieux, and S. Green. Automatic generation of social tags for music recommendation. *NIPS*, 20, 2007.
7. C. Eickhoff, C.G. Harris, A.P. de Vries, and P. Srinivasan. Quality through flow and immersion: gamifying crowdsourced relevance assessments. In *SIGIR 2012*.
8. K. Filippova and K.B. Hall. Improved video categorization from text metadata and user comments. In *SIGIR 2011*.
9. P. Heymann, D. Ramage, and H. Garcia-Molina. Social tag prediction. In *SIGIR 2008*.
10. D. Hiemstra. A probabilistic justification for using tf× idf term weighting in information retrieval. *JDL 2000*.
11. M. Hu, A. Sun, and E.P. Lim. Comments-oriented blog summarization by sentence extraction. In *CIKM 2007*.
12. G. Kazai. In search of quality in crowdsourcing for search engine evaluation. *ECIR 2011*.
13. J. Kleinberg. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, (4), 2003.
14. M. et al. Larson. Automatic tagging and geotagging in video collections and communities. In *ICMR 2011*.
15. Y. Matsuo and M. Ishizuka. Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, (1), 2004.
16. G. Mishne and N. Glance. Leave a reply: An analysis of weblog comments. In *WWE 2006*.
17. A. Oghina, M. Breuss, M. Tsagkias, and M. de Rijke. Predicting imdb movie ratings using social media. *ECIR 2012*.
18. P. Ogilvie and J. Callan. Combining document representations for known-item search. In *SIGIR 2003*.
19. S. Robertson, H. Zaragoza, and M. Taylor. Simple bm25 extension to multiple weighted fields. In *CIKM 2004*.
20. S. Siersdorfer, J. San Pedro, and M. Sanderson. Automatic video tagging using content redundancy. In *SIGIR 2009*.
21. T. Tomokiyo and M. Hurst. A language model approach to keyphrase extraction. In *ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment*.
22. C. Wartena, R. Brussee, and W. Slakhorst. Keyword extraction using word co-occurrence. In *Database and Expert Systems Applications (DEXA), 2010*.
23. L. et al. Wu. Distance metric learning from uncertain side information with application to automated photo tagging. In *ACM Multimedia 2009*.
24. W.G. Yee, A. Yates, S. Liu, and O. Frieder. Are web user comments useful for search? *Proc. LSDS-IR*, 2009.