# An Evaluation of Search Strategies for User-Generated Video Content

Christopher G. Harris
Informatics Program
The University of Iowa
Iowa City, IA 52242

christopher-harris@uiowa.edu

## ABSTRACT

As the amount of user-generated content (UGC) on websites such as YouTube have experienced explosive growth, the demand for searching for relevant content has expanded at a similar pace. Unfortunately the minimally-required production effort and decentralization of content make these searches problematic. In addition, most UGC search efforts rely on notoriously noisy user-supplied tags and comments. In this paper, we examine UGC search strategies on YouTube using video requests from several knowledge markets such as Yahoo! Answers. We compare crowdsourcing and student search efforts to YouTube's own search interface and apply these strategies to different types of information needs, ranging from easy to difficult. We evaluate our findings using two different assessment methods and discuss how the relative time and financial costs of these three search strategies affect our results.

## Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *Relevance feedback, Search process*

## General Terms

Experimentation, Design, Verification.

## Keywords

Crowdsearch, crowdsourcing, search strategies, user-generated content, YouTube

## 1. INTRODUCTION

Until recently the content available to consumers using video-on-demand systems has been tightly controlled by a limited number of media producers. Within the last decade, *user-generated content* (UGC), propelled by Web 2.0 content-sharing initiatives, have re-shaped the online video market enormously. As a result, hundreds of millions of Internet users worldwide have now become self-publishing consumers.

Currently the most popular UGC video-on-demand website is YouTube; each month 90 billion videos are viewed by 800 million unique users [1]. In December 2011, "youtube" was the third-

most searched term on Google [2] and was the third-most visited website as measured by Alexa [3]. Many visitors do more than simply view content – more video content is uploaded to YouTube in 60 days than the three major US networks have created in the past 60 years [1].

The lower required production effort, exponential growth, and decentralization of the UGC videos often make searches for specific content challenging: to compensate, searchable content on UGC websites is often restricted to producer-supplied categories and tags or obtained from viewer-supplied comments. YouTube comment text is frequently noisy and insufficient to produce a set of content and/or context terms from which to search effectively [4]. In addition, despite the Web 2.0 features YouTube has integrated to encourage user participation, an examination by Cha *et. al.* [5] found the level of active user participation is remarkably low - comments on YouTube videos are provided by a mere 0.16% of total viewers. This limited contribution to searchable text also has a negative impact on search quality.

Categories used on UGC websites are often too broad and lack the discriminative power for use in most searches; YouTube, for example, contains 15 broad categories with labels such as *Autos & Vehicles*, *Comedy*, and *Education.* In contrast, producer-supplied tags on UGC websites are usually quite sparse and do not always represent the true video content. In a study of more than one million YouTube videos conducted by Geisler and Burns in [4], the median number of tags applied per video was 6.0. One of the study's findings was many tags did not adequately describe the actual video content. Rarely do the terms used by video content producers match those used in searches, as Bischoff *et. al.* illustrated in [6]. For example, people tagging music videos would likely use terms associated with its genre, such as "rock," whereas people generally do not search for music videos via genre, instead opting for searches containing song title and/or artist.

Despite these shortcomings, the search function on YouTube's website remains the most frequently-used method to find videos, according to Zhou *et. al.* [7], yet many user queries for UGC go unsatisfied. Knowledge market websites, such as Yahoo! Answers[1] and Answers.com[2], contain unfulfilled and partially-fulfilled user requests for videos; as of January 2012, Yahoo! Answers alone had more than 250,000 requests for assistance to locate videos for a specific information need. Some studies, such as one conducted by Dearman and Troung [8] and another by

Bian *et. al.* [9] found that inadequate phrasing of a question and/or corresponding answer on knowledge market websites negatively affects utility. Consequently, the ability to effectively search for UGC, particularly on rare or noisy topics, remains a challenge.

Crowdsourcing may provide a viable solution for searching UGC. The use of the crowd as a search strategy is compelling; it introduces diversity of search terms since different members of the crowd will apply different search strategies based on their familiarity with the search topic. Moreover, the crowd has been shown to provide good quality in studies involving relevance judgments. Even with diversity, we can still expect search quality: some studies on prediction in crowdsourcing systems demonstrate that reliability of the average of predicted scores by the crowd improves as the size of the crowd increases [10, 11]. Likewise, search quality is expected to improve as the number of searchers in the crowd expands. Crowdsourcing contrasts with knowledge markets in level of engagement; Nielsen mentions in [12] that over 90% of knowledge market group participants fail to contribute; therefore the crowdsourcing aspect introduces some financial incentive to motivate task participation.

The objective in this paper is to examine if the crowd can provide a more precise set of UGC search results, given a query, compared with other multimedia search tools. The contributions of this paper are as follows. First we compare the retrieval performance of different retrieval models in terms of precision on several categories using UGC video requests taken from leading knowledge market websites. We then compare YouTube's own search interface with a search conducted by students as well as a search approach using crowdsourcing. We evaluate our results using two methods: mean average precision determined after applying pooling, and a simple list preference, where the entire list of videos judged as relevant by each method are compared.

The remainder of the paper is organized as follows. In Section 2 we put our work in the context of previous work. In Section 3 we discuss our experimental setup. Section 4 offers a discussion of the results. We conclude and provide insight into future work in Section 5.

## 2. RELATED WORK

Even prior to Web 2.0, there has been significant research in multimedia search methods, including several organized competitions that involve traditional search strategies. The popular TRECVid [13] benchmarking competition focuses on the detection of specific features within non-UGC multimedia collections. Wikipedia Retrieval, a task in ImageCLEF [14] involves locating relevant images from the Wikipedia image collection based on a provided text query and several sample images. While Wikipedia Retrieval examines noisy and unstructured textual annotations in Wikipedia multimedia, the semi-structured content evaluated in ImageCLEF is far less noisy and more structured than content searches on YouTube.

Several studies have examined search quality on user-supplied tags in other Web 2.0 applications. Diversity of image tag search results in Flickr using an implicit relevance feedback model is explored by von Zwol *et. al.* [15], concluding that diversity is an important component when retrieval is based on small data sets, such as those found in image tags. Hotho *et. al.* explore folksonomy tagging, which is bound by the same noisy unstructured restrictions as YouTube tags [16], but their study was primarily focused on recommender systems usage of these tags. Others have examined multimedia search effectiveness on knowledge market websites, such as Chua *et. al.* in [17] and Li *et. al.* in [18]; however, their focus is to locate all content addressing a specific question (e.g. "how to" and "why" question types) whereas the focus of our study is on finding and ranking videos that fulfill a specific search request (e.g., "help find a video").

A few studies have examined the effectiveness of crowds on noisy data searches. Steiner *et. al.* demonstrated searches of event detection methods in YouTube videos at the fragment level [19]. Hsueh *et. al.* examined searches in political blogs in [20] which, although noisy, do not experience the restrictions inherent in multimedia tags. In [21], Yan *et. al.* provided an innovative approach called CrowdSearch, which provided near-real-time assessment of images. Although the authors' focus was on labeling images, their approach could feasibly be extended to locating similar media on YouTube.
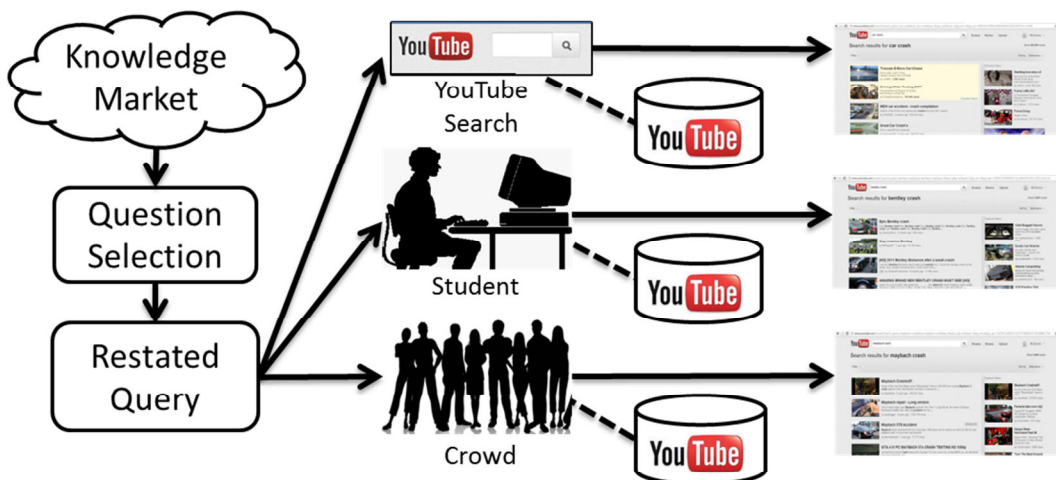


**Figure 1. Overview of the video retrieval process involving YouTube's search interface, students, and the crowd.**

# 3. EXPERIMENTAL SETUP

## 3.1 Retrieval Process

Our objective is to compare the search results obtained from crowdsourcing, human searchers, and YouTube's own search interface. YouTube's search interface is a version of Google's search that has been refined for YouTube, and represents a significant share of Google-based searches. Since late 2008, metrics confirm that video searches on YouTube account for more than a quarter of all Google search queries in the U.S, and a similar share in a most other countries [22]. We began by extracting a set of 100 questions randomly taken from four knowledge market sources (Yahoo! Answers, Answers.com, Blurtit[3] and Allexperts[4]) containing the terms "find" and "video" and remained either unanswered or partially-answered (i.e. the requestor did not indicate their query had been satisfied). We pared our list of questions down to 45 by removing those where the requestor's need could not be clearly determined or we could not find any candidate videos on YouTube's website that appeared to meet the stated criteria through a preliminary search.

Our method is similar to that used by Kofler *et. al.* in [23]. For each request, we removed noisy terms from the original request (e.g., only retain those that support the identified information need); we call this a *Restated Query*. An example of this query refinement procedure is shown in Table 1. We classify each request into one of three categories based on our own assessment of query difficulty using the Restated Query using the following guidelines. Requests classified as "easy" are relatively straightforward to find one (or more) videos that match the stated request - likely listed as a result of requestor laziness or inexperience with search tools. Requests classified as "medium" require some additional refinement, such as an expansion of terms or enhancement using synonyms. Requests classified as "difficult" require significant term refinement to obtain links to YouTube videos. Our final set of queries contained 15 of each difficulty level. This retrieval process is outlined in Figure 1. Examples of Restated Queries categorized as "easy", "medium", and "difficult" appear in Table 2.

**Table 1. The process of developing a restated query from an unanswered knowledge market query.**

| Original Knowledge Market Query | Restated Query |
|---|---|
| Hey everyone, my friend told me about this one guy who drove his Bugati (totally expensive car) into the river as part of an insurance fraud scheme. Help me find the video of it. | Bugatti Veyron driven into water as part of a possible insurance fraud scheme |
| Where can I find video of President Bush conducting an orchestra? I saw this last night on the news, but I can't find it on Yahoo News and there are some articles on ABC and other newspapers but no video. Anyone, help? | President Bush conducting an orchestra |

---

**Table 2. Examples of restated queries categorized as easy, medium and difficult**

> **Easy**
>
> **E1** "Seinfeld" where George Costanza says "serenity now!"
>
> **E2** Coca Cola warehouse that a forklift dropped a LOT of bottles
>
> **E3** Bud Light "real men of genius" bad toupee wearer
>
> **Medium**
>
> **M1** British television show "Top Gear" goes to the deep south of US and makes the locals angry
>
> **M2** song that's called "Millionaire" in the Madden NFL video game
>
> **M3** Michael Jackson's sister as an abused kid on "Good Times" television show from the 1970s
>
> **Difficult**
>
> **D1** fans moving a large rock for Tiger Woods
>
> **D2** clay-mation figures help each other to climb a wedding cake
>
> **D3** girl eating dinner reading comic book gets pulled into an animated music video from the 80s

We then run each of the Restated Queries using YouTube's own search interface, obtain the YouTube video links related to the query. We limit our retrieval results only to the top 40 ranked videos that appear on YouTube's website, since early research in web search indicated a large majority of users examine fewer than four pages of web results [24]; we assume video searchers behave no different than web searchers.

For the student search method, we asked five university students to perform each search. We paid an hourly rate of $10 per hour to search each of the Restated Queries, a typical wage for this type of task in our area. Each student was instructed to provide a list (of up to 40) YouTube video links for each Restated Query. Although given unlimited time, the student group took an average of just under 90 minutes to complete all 45 queries. Participants were told they could use any available search methods or tools.

For the crowdsourced search method, we use the Amazon Mechanical Turk[5] platform (MTurk) to list tasks, and provide each worker with Restated Query for each question with instructions to return at least 10, but not more than 40, of the most relevant YouTube video links. Using MTurk, we created 45 queries, called Human Intelligence Tasks (HITs), amounting to one HIT for each Restated Query. As with the student searchers, crowdsourcing participants were told they could use any search tools they desired and thus were not constrained to using YouTube's search interface. We paid participants $0.10 per completed HIT, which is a typical wage for this type of crowdsourcing task; to maximize the use of the crowd model and differentiate it from the student search model, crowdsourcing participants were not able to participate in more than one HIT.
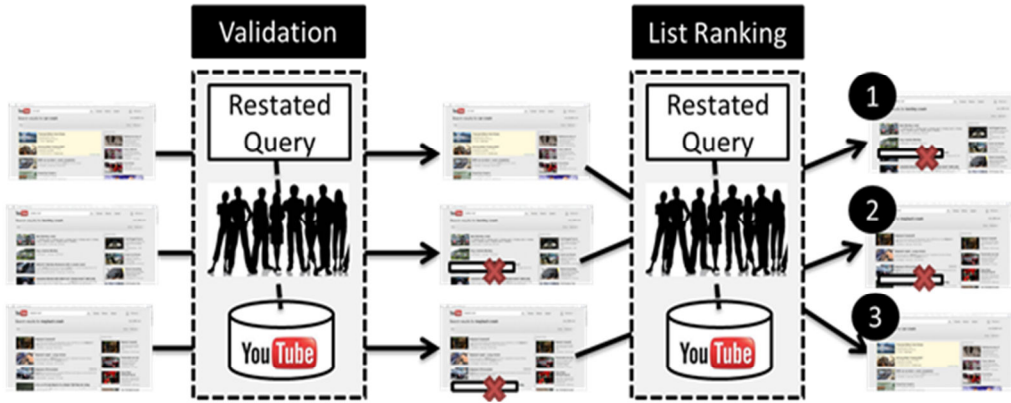
---

**Figure 2. The video validation and ranking steps for the simple list preference method. Each step is performed by different subgroup from the crowd.**

## 3.2 Evaluation

The result sets were scored and ranked two different ways: pooling, which has been used in TRECVID, and simple list preference, where the each result set is first validated and compared as a whole.

### 3.2.1 Pooling

Therefore, the following pooling technique is used instead. We employ the pooling method used in TRECVID [25]. First, a pool of potentially-relevant YouTube video links is obtained by gathering the sets of links returned by the YouTube query, the human searchers, and the crowdsourcing group. These sets are then merged, duplicate links are removed, and the relevance of only this subset of YouTube video links is assessed.

The performance measure used to evaluate and rank the results is *average precision* (AP):

$$\text{AP} = \frac{1}{R} \sum_{k=1}^{A} \frac{R \cap L^k}{k} \lambda(l_k)$$

where $L^k = \{l_1, l_2, ..., , l_k\}$ is the rank version of the answer set, A. At any given rank $k$, let $R \cap L^k$ be the number of relevant videos in the top $k$ of $L$, where $R$ is the total number of relevant videos. Indicator function $\lambda(l_k) = 1$ if $l_k \in R$ and 0 otherwise. Since the denominator $k$ and the value of the indicator function $\lambda(l_k)$ are dominant in determining average precision, it can be understood that this favors relevant videos appearing towards the top of the list. Mean average precision (MAP), which is the mean of the average precision values over a set of queries, has been a key standard evaluation measure in TRECVID.[6] We used the list of all relevant videos for each question as our determination of ground truth.

### 3.2.2 Simple List Preference

Perhaps a more holistic metric is the *simple list preference*, which utilizes the lists returned by each of the three search strategies as entities. The videos on each list are validated for relevance

against the video request and those that are judged irrelevant are removed. The remaining lists are evaluated in pairwise fashion. Figure 2 gives a high-level overview of this evaluation method.

For each of the 45 requests, we have 3 result sets, one for YouTube search, one for student search and one for Crowdsourcing. The mean size of the 135 result sets was 27.7 video links, with a standard deviation of 6.8.

The first step is validation. We separated the 3743 video links in groups of 15, comprising 250 separate HITs. To each HIT, we introduced 2 "trap" links - clearly irrelevant links added to ensure assessor attention to detail. We posted these 250 HITs, each containing 17 binary relevance judgments and paid $0.25 per completed HIT. Each of the 250 crowd assessors was only permitted to evaluate a single HIT. Thirteen of the HITs were rejected and had to be relisted due to the assessor failing the trap link judgment. This validation step reduced the 3743 YouTube links by just over eighty percent to 728, averaging 5.4 relevant video links for each of the 135 result sets.

Using the validated links, within each result set, we presented the lists in pairs along with an individual thumbnail from each video as a HIT. For each of the original 45 requests, lists were presented in random order to avoid selection bias, along with the Restated Query. We asked each worker to choose the list that best answered the restated query. We posted each pairwise judgment at least twice in order to ensure that the highly-subjective determination of ground truth was made by two different people. Workers were paid $0.10 per judgment and were restricted from rating more than one query. If the two raters had a difference in list preference or the resulting list preference was cyclical (i.e., 1>2, 2>3, 3>1), we hired an additional rater from the crowd to establish a clear preference order. Two assessors each made 3 pairwise judgments across the 45 Restated Queries, with a Cohen's kappa of 0.624, representing a reasonably strong inter-annotator agreement. Of the 270 pairwise decisions 21 required the use of a tiebreaker, and no cyclical references were encountered. For each set of results obtained by our Restated Query, we then apply a Condorcet method to each pairwise preference among strategies and evaluate based on the lists of relevant UGC videos they contain.

---

[6] In recent years, average precision has been replaced by inferred average precision (IAP), which closely approximates the AP measure but requires only a subset of the pooled results to be manually evaluated.

# 4. RESULTS

## 4.1 Pooling

Using the pooling evaluation method, we calculate the MAP scores for each of the search efforts. These are given in Table 3. While these scores seem reasonable, it is likely due to two issues: our calculation of ground truth and, for most searches, there were only a small percentage of YouTube videos were considered relevant. The crowdsourcing search strategy and the student search strategies performed better than the YouTube search interface as measured by MAP, a result that is statistically significant (two tailed, $p<0.05$).

**Table 3. Overall MAP scores for each search strategy.**

| Search Strategy | MAP |
|---|---|
| Student Search | 0.524 |
| Crowdsourcing | 0.512 |
| YouTube Search | 0.359 |

Since Restated Queries were grouped into three separate categories (easy, medium, and difficult), we evaluated them separately for each search strategy. The results are reported in Table 4.

**Table 4: MAP scores for each search strategy, broken down by search category.**

| Search Strategy | Easy | Medium | Difficult |
|---|---|---|---|
| Student Search | 0.636 | 0.516 | 0.421 |
| Crowdsourcing | 0.619 | 0.512 | 0.404 |
| YouTube Search | 0.508 | 0.344 | 0.224 |

Table 4 raises some interesting points for discussion. First, MAP scores for easy queries are much more consistent across strategies compared with those for medium or difficult searches. This is likely a result of a heavier reliance for students and the crowd on the standard YouTube search interface for the easier queries, limiting the advantages of human computation. As more difficult queries are encountered, the value of human computation becomes a more important consideration.

Second, although the MAP score gap is small between student search and crowdsourcing, we do notice that the five students consistently performed slightly better than the crowd. Each student performed all 45 queries, refining their sources and techniques as they encountered each new query – all five participants performed faster and provided better search results towards the end of their query session than in the beginning (we cannot observe this improvement with the crowd as each crowd participant provided results for only a single query). The crowd had the smallest deviation in MAP scores across the 3 search categories, primarily because the larger number of people searching reduces the variation, as discussed in [10] and [11].

Third, we can see the value of using human input in these MAP scores, but Table 4 does not take the costs in both time and money into consideration. We make the assumption that YouTube's search has no cost in terms of time and money and use it as our baseline. We kept track of the elapsed time taken by the crowd and for the students as well, so we can evaluate this in aggregate. This is reported in Tables 5 and 6.

**Table 5: Increase in MAP scores over the YouTube search interface divided by additional time taken (in minutes).**

| Search Strategy | Easy | Medium | Difficult |
|---|---|---|---|
| Student Search | 0.125 | 0.101 | 0.061 |
| Crowdsourcing | 0.059 | 0.066 | 0.040 |

**Table 6: Increase in MAP over the YouTube search interface divided by additional cost (in US cents).**

| Search Strategy | Easy | Medium | Difficult |
|---|---|---|---|
| Student Search | 1.332 | 1.656 | 2.723 |
| Crowdsourcing | 2.155 | 1.953 | 1.111 |

To illustrate, in Tables 5 and 6, for Restated Queries classified as "difficult", to obtain an increase in MAP of 0.001 using students, we would expect to spend 0.06 minutes and incur a cost of 2.723 cents. To obtain an equivalent increase in MAP using crowdsourcing, we would expect to spend, on average, 0.04 minutes and incur a cost of 1.111 cents. Note that these numbers represent long term averages. Thus, we observe that using the crowd, as compared with students, requires 40% of the cost and takes two thirds the time, on average, to raise MAP by an equivalent amount. Thus, when obtaining more precise results is our paramount objective, using students or the crowd is expected to provide the best results. If time or financial costs are also a consideration, our results show that using the crowd will provide the best tradeoff between time, financial cost, and precision.

## 4.2 Simple List Preference

We apply Copeland's pairwise aggregation method, described in [26, 27], is a Condorcet method used to evaluate pairwise preferences. Copeland's pairwise aggregation method examines two lists for a given query in a pairwise fashion and records the assessor's preference as a "victory". Search strategies are ordered by number of victories over each opponent to determine an overall winner. We examine each pairwise preference for the three result lists for all 45 queries. These comparison results are given in Table 7.

**Table 7: Pairwise comparisons of list preference using Copeland's pairwise aggregation method, as assessed by the crowd.**

| Comparison | Result | Winner |
|---|---|---|
| Student Search vs. Crowdsourcing | 24 vs. 21 | Student Search |
| Student Search vs. YouTube Search | 39 vs. 6 | Student Search |
| Crowdsourcing vs. YouTube Search | 37 vs. 8 | Crowdsourcing |

From Table 7, we observe that student search is our Condorcet winner, beating all other search strategies in pairwise comparisons. As with the pooling assessment method, there was a slight preference of student search results over the crowdsourcing supplied video lists. However, when financial costs are disclosed to the assessors along with the scores, crowdsourcing is our Condorcet winner, as observed in Table 8.

**Table 8: Pairwise comparisons of list preference using Copeland's pairwise aggregation method when financial cost is considered, as assessed by the crowd.**

| Comparison | Result | Winner |
|---|---|---|
| Student Search vs. Crowdsourcing | 10 vs. 35 | Crowdsourcing |
| Student Search vs. YouTube Search | 17 vs. 28 | YouTube Search |
| Crowdsourcing vs. YouTube Search | 31 vs. 14 | Crowdsourcing |

## 5. CONCLUSION

This study has examined the effects of using students, crowdsourcing, and YouTube's search interface on UGC searches. We observe that human computation efforts provide better MAP scores than YouTube's own search interface across all categories. In addition, our study examines the costs, in terms of time and money, of this MAP score increase for each search strategy. Although this study didn't explicitly vary the financial incentives offered to students or the crowd, we do observe there is a tradeoff between better precision and search costs (in terms of time and money); it is up to each search requester to decide if these costs outweigh the need for improved precision.

We also examine the retrieval lists as complete entities. We see that a simple list preference favors the student search strategy when costs are not considered; if time and cost are to be considered, crowdsourcing achieves additional consideration due to the cost savings it offers over student search. This reinforces the findings observed through pooling evaluation.

In future studies, we plan to examine the financial incentives offered to examine the marginal benefit of achieving better precision. Similarly, we plan to investigate whether we can incentivize the crowd to increase their performance without increase time and financial cost. We also plan to examine different types of searches, such as those specific to a particular domain to observe if searches can be performed effectively when specific domain knowledge is required.

## 6. REFERENCES

[1] YouTube Press Statisics. http://www.youtube.com/t/press_statistics. Retrieved January 8, 2012.

[2] Google Insights for Search. http://google.com/insights/search. Retrieved January 8, 2012.

[3] Alexa Top Site Info for YouTube. http://alexa.com/siteinfo/youtube.com. Retrieved January 8, 2012.

[4] Geisler, G. and Burns, S. Tagging video: conventions and strategies of the YouTube community. ACM. 2007.

[5] Cha, M., Kwak, H., Rodriguez, P., Ahn, Y.-Y. and Moon, S. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement* (San Diego, California). ACM. 2007

[6] Bischoff, K., Firan, C. S., Nejdl, W. and Paiu, R. *Can all tags be used for search?* ACM. 2008.

[7] Zhou, R., Khemmarat, S. and Gao, L. *The impact of YouTube recommendation system on video views*. ACM. 2010.

[8] Dearman, D. and Truong, K. N. *Why users of yahoo!: answers do not answer questions*. ACM. 2010.

[9] Bian, J., Liu, Y., Agichtein, E. and Zha, H. Finding the right facts in the crowd: factoid question answering over social media. ACM 2008.

[10] Surowiecki, J. *The Wisdom of Crowds*. Anchor Press. New York. 2005.

[11] Pennock, D. The wisdom of the ProbabilitySports crowd. http://blog.oddhead.com/2007/01/04/the-wisdom-of-the-probabilitysports-crowd/. Retrieved January 12, 2012.

[12] Nielsen, J. Participation inequality: Encouraging more users to contribute. http://www.useit.com/alertbox/participation_inequality.html. Retrieved January 12, 2012.

[13] Smeaton, A. F., Over, P. and Kraaij, W. Evaluation campaigns and TRECVid. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval* (New York). ACM. 2006.

[14] Tsikrika, T., Popescu, A. and Kludas, J. Overview of the wikipedia image retrieval task at ImageCLEF 2011. Amsterdam. 2011.

[15] van Zwol, R., Murdock, V., Pueyo, L. G. and Ramirez, G. Diversifying image search with user generated content. In *of the 1st ACM international conference on Multimedia information retrieval* (Vancouver 2008). ACM, 2008.

[16] Hotho, A., Jäschke, R., Schmitz, C. and Stumme, G. Information retrieval in folksonomies: Search and ranking. *The Semantic Web: Research and Apps,* 2006. pp, 411-426.

[17] Chua, T. S., Hong, R., Li, G. and Tang, J. From text question-answering to multimedia qa on web-scale media resources. ACM. 2009.

[18] Li, G., Ming, Z., Li, H. and Chua, T. S. *Video reference: question answering on YouTube*. ACM. 2009.

[19] Steiner, T., Verborgh, R., Van de Walle, R., Hausenblas, M. and Vallés, J. G. Crowdsourcing Event Detection in YouTube Videos. In *Proceedings of the 10th International Semantic Web Conference* (Koblenz, Germany, 2011).

[20] Hsueh, P.-Y., Melville, P. and Sindhwani, V. Data quality from crowdsourcing: a study of annotation selection criteria. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing* (Boulder, Colorado, 2009). ACL, Strousburg, PA. 2009.

[21] Yan, T., Kumar, V. and Ganesan, D. CrowdSearch: exploiting crowds for accurate real-time image search on mobile phones. ACM. 2010.

[22] *ComScore: YouTube Now 25 Percent Of All Google Searches*. http://techcrunch.com/2008/12/18/_comscore-youtube-now-25-percent-of-all-google-searches/ Retrieved January 22, 2012.

[23] Kofler, C., Larson, M. and Hanjalic, A. To seek, perchance to fail: expressions of user needs in internet video search. Advances in Information Retrieval. 2011. pp. 611-616.

[24] Spink, A., Wolfram, D., Jansen, M. B. J. and Saracevic, T. Searching the web: The public and their queries. Journal of the American society for information science and technology, 52:3 2001. pp. 226-234.

[25] Over, P., Awad, G., Smeaton, A. F., Foley, C. and Lanagan, J. Creating a web-scale video collection for research. In Procedings of the 1st workshop on Web-scale multimedia corpus (Beijing, China, 2009). ACM, 2009.

[26] Copeland, A. H. *A reasonable social welfare function*. mimeo. University of Michigan, 1951.

[27] Moulin, H. Choosing from a tournament. *Social Choice and Welfare*, 3:4 1986. pp 271-291.