

Broadcast yourself on YouTube – Really?

Gijs Kruitbosch
University of Amsterdam
Kruislaan 419

1098 GB Amsterdam
+31 20 525 6377

gijskruitbosch_at_gmail.com

Frank Nack
University of Amsterdam
Kruislaan 419

1098 GB Amsterdam
+31 20 525 6377

nack_at_uva.nl

ABSTRACT

One essential reason for people to publish on the web is to express themselves freely. YouTube facilitates this self-expression by allowing users to upload video content they generated. This paper investigates to what extent the videos on YouTube are self-generated content, instead of amalgamated content that was mainly professionally authored in the first place. Results show that most of the popular content on YouTube was professionally generated, even though a random sample shows that there is plenty of user-generated content available – it just does not make the cut. As a result we propose that YouTube is more of a social filter, allowing anyone to share content they find interesting rather than a way for aspiring creative people to show their creative abilities to the world. The outcome is a set of requirements which describe better means for YouTube to support better authoring and presentation of video, where the core research direction is focused on the self-representation of humans in the realm of their creative possibilities on one side as well as the stimulation of new insights on existing material to stimulate new creative impulses.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]:

General Terms

Measurement, Design, Human Factors

Keywords

YouTube, user-generated content, search for video content

1. INTRODUCTION

YouTube's slogan "Broadcast Yourself" implies that the focus of this web site is to facilitate users to upload their video. The aim is to provide a platform where people can express their creativity. To some extent YouTube has been immensely successful, averaging 65,000 daily uploads [1] and being visited by 20% of Internet users every day [2].

Thus, it is often considered a good example of a website focused around user generated content (UGC) [1], [3], [4]. The prevalence and importance of UGC were also stressed by Times magazine, who in 2006 found "You" to be the Person of the Year [5], and a recent survey found that the executives of large media companies

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HCC'08, October 31, 2008, Vancouver, British Columbia, Canada.
Copyright 2008 ACM 978-1-60558-320-4/08/10...\$5.00.

thought user generated content was the biggest threat to their businesses [6].

Due to the large amounts of traffic it receives, YouTube also attracts large amounts of professionally created content. Some of this content is uploaded by professionals or the companies that manage them, such as Sony or MTV [7]. Some of it is uploaded by users who happened to be able to capture it somewhere and now wish to share it with the rest of the world. In other words YouTube is an environment full of mash-ups, parodies, malapropisms or simple copies.

This article investigates the role and impact of user generated content on YouTube. In order to verify this we looked at the uploaded material of a week and analyzed the content, namely the visual as well as the audio material, of 100 randomly selected video as well as the content of the 100 most popular videos. The analysis focused on a classification of the available material in a range between 'professional' and 'user-generated'. The aim of the analysis is to get a better grip on methods necessary to support access to the material once it is made available for sharing. Here we go a bit further than earlier research that has shown that 10% of the videos on YouTube account for 80% of the views on the site [3], which indicates that users wish to access 'interesting' material.

The result of the analysis forms the basis of the description of a research agenda, as outlined in section 5. This agenda aims at supporting two important aspects of social video-based interaction. First, we will develop new authoring mechanisms that enhance the creativity of a typical YouTube author by helping to avoid the general flaws, compared to film theory, while the material is assembled or by providing a basic quality evaluation while the video is uploaded to YouTube. The second research direction exploits the already available metadata in YouTube to establish new presentation forms of the material, such as biased presentation of news or automatically generated gossip shows. The core development focuses on the self-representation of humans in the realm of their creative possibilities on one side as well as the stimulation of new insights to stimulate new creative impulses.

2. METHOD

In order to analyze the data available on YouTube with respect to its production background we took a random sample of 100 videos uploaded during one week in March 2008. We also took a sample of the 100 most popular videos of the same week. These videos were downloaded and then analyzed based on a categorization we developed.

2.1 Sampling details

YouTube provides an astonishingly limited access to its videos. It is not possible, for example, to query for a 'random' video, and consequently it is impossible to obtain a completely

random sample of the videos that are currently available on YouTube. Instead we used a feed that YouTube publishes with the most recent uploads to its website. A small sample showed that the feed contains 15 items and is changed approximately every 20 minutes.

The 1080 videos contained in this feed do not cover the total number of uploads to YouTube per day (65.000 [1]) and they are certainly selected by YouTube using some sort of filtering but the sample obtained from this feed can still be considered representative.

During the period of one week, we retrieved the recent uploads every 5 minutes, so as to make sure not to miss any sudden changes. We took 2 randomly chosen videos from every 15 videos referenced for analysis and made sure that no overlap of videos in the sample occurred. This potentially provided 1008 videos for investigation. However, we missed out on 19 videos that were taken down by their owners within the time between their insertion into the feed, and the time at which we attempted to download them (a time span of 5 minutes or less). In addition there were two time periods in which YouTube was not available. The first down period was caused by problems with our own Internet connection (downtime: 6 hours) causing 3.5 % of the potential sample to be unavailable. The second downtime was caused by a Pakistani ISP who accidentally blocked YouTube from the entire world (downtime: 3 hours [8]). As the second downtime was global nothing was uploaded in this time frame, which clearly means it cannot have affected the validity of the sample. In the end, 811 videos were available for investigation.

From the 811 videos, we randomly selected 100 for analysis, a number corresponding to the 100 videos listed in the "Most Viewed Videos" feed. To compare results from this sample with that of the most popular videos, we used YouTube's list of "Most Viewed Videos- this week" at the end of this one week period. This list contains the 100 videos that were accessed most often during the past week. 3 of the 100 videos in this list were removed because of copyright claims before we could download them.

2.2 Categorization

The content on YouTube consists of both video and audio material encoded in Flash video (.flv) using the H.263 codec. While there is always an audio track, uploaders are free to submit silent videos. Likewise, not all videos contain actual video sequences: some contain only the representation of a slideshow of text or photos. Furthermore, while YouTube is used both for user-generated and for professional content, there is also a large gray area of videos created by users from cuts of existing professionally created content, as well as videos that are user-created recordings of professional performances, such as theater shows, concerts or sports events.

In order to accurately represent these distinctions, we categorized the actual audio and video material separately, and also considered the editing performed to create the final product as a separate property of the video. This allows for distinguishing videos that were taken and edited by the user, videos that were professionally recorded and edited, and videos that are made by users from existing professionally recorded material. The possible values and rationale for assigning them is explained below:

A: Video recording :

User-generated - content planned/observed/experienced and recorded by the user using a webcam, handcam, mobile phone or similar, as well as user-created still photography that has been edited into a video.

User-generated of a professional performance - content created with the same technology as described above but making use of professional performances, like concerts, theater shows, dance performances, and so on.

Professional recording - this is content recorded for professional use (e.g. in news, feature films, etc.) For cases that a slideshow is produced, this content covers material produced by professional photographers or visual artists.

B: Audio recording

User recording - applies to content planned recorded using a personal audio recorder, mobile phone or camera.

User recording of professional performance - using the technical means as above, but applied to professional performances, like concerts, musicals, etc.

Professional recording - consists primarily of professionally created music, interviews, talk shows and so on.

C: Editing

User edited - content is edited unprofessionally. It might have been recorded professionally, but has been edited afterwards by the user who uploaded it.

Professionally edited - content is edited professionally, for publication on various channels, such as tv or films, and various purposes (e.g. advertising, trailer, infotainment, etc).

All 200 videos were classified using the outlined taxonomy.

3. RESULTS

Based on our categorization of user-generated and professional content Figure 1 to Figure 3, on the following page, describe the percentage of this type of content in the 200 analyzed videos. Content classified as 'unknown' represents material that was withdrawn between the time it was registered and our attempt to download the video itself.

For getting a better understanding of the presentation style of the videos we also calculated the average length of the videos, both for all the videos and for the subsets of user edited and professionally edited videos. With that information we wanted to see if the presentation style has any influence on the access behavior of YouTube visitors. In addition we collected data on the number of times each video was accessed. These results are presented in Table I and II on the following page.

4. EVALUATION

The statistics provided in the previous section can be summarized into the following three statements:

- user-generated content is, no matter how it is measured, not as present in the list of most viewed videos, even though it is clearly there when one observes the randomly selected videos;
- user-edited videos consistently get fewer views than professionally edited ones;
- user-edited content is considerably shorter than professionally edited content.

Individual cases illustrate, however, that it is in fact possible for user-generated content to appear in the most popular videos. One video, for example, was created by the father of a 3 year old girl. In the video the girl tried to explain Star Wars, to the hilarity of many people, as the comments attached to the video demonstrated. This particular video was viewed by more than 5

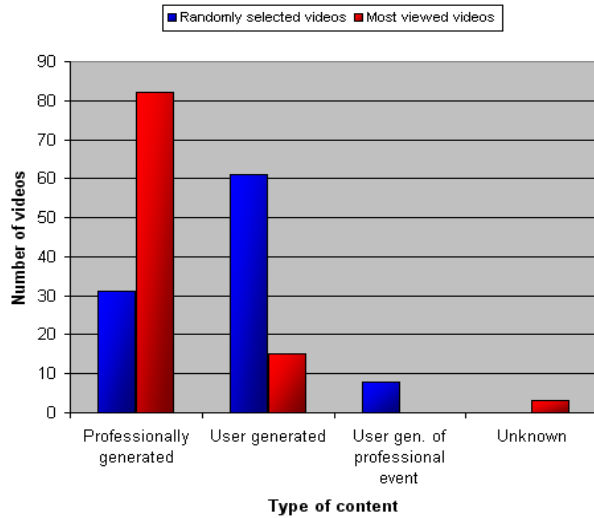


Fig. 1. Categorization of Video Content

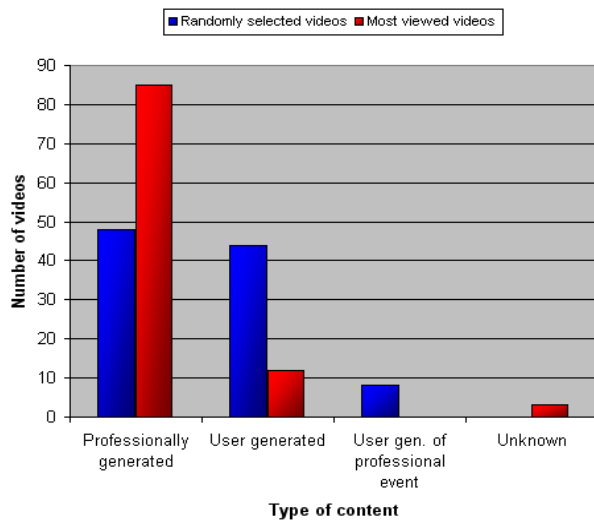


Fig. 2. Categorization of Audio Content

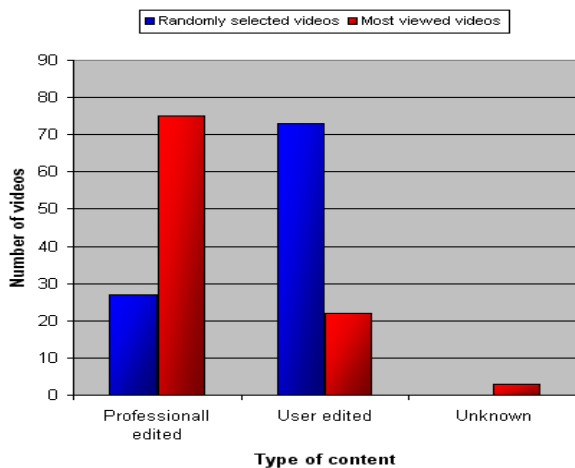


Fig. 3. Categorization of Editing

TABLE I Average video length in seconds

Type of content	Randomly selected	Most viewed
Prof. edited	190	216
User edited	106	211
All	129	212

TABLE II Average number of views per video

Type of content	Randomly selected	Most viewed
Prof. edited	1190	665.285
User edited	812	607.144
All	841	655.597

million times after it was linked extensively across weblogs and social aggregation sites.

As another example, one of the most viewed videos was a creation of a university student, who runs a website that tries to reunite people with their lost cameras. He had mixed a radio interview of himself by CBC (professional content) with some of the lost pictures (user-generated, but not his own) in order to advertise his website [9]. Content like this relies on being able to use professional content without restriction. Research is already being done as to the legality of mixing such content [10].

However, our data sample shows that these cases are rare, and the overwhelming majority of popular content is still professionally made. So while it is clear that user-generated content does feature on YouTube, it does not seem to be the most important part of the website for most viewers. The view of YouTube as being ‘about’ user generated content is therefore clearly incorrect. The remaining question is: what is YouTube then all about?

One comparison that can be made is that with the Hollywood from the twenties to the fifties of the last century: a platform for people to publish. Hollywood studios funded directors, but also owned the theaters, making them a vehicle for the production of movies from start to finish. One of the promoting methods in this scheme were, and still are, the Academy Awards, where actors and other movie professionals vote for the best achievements within their field of that year. In much the same way, YouTube takes care of distributing the uploaded videos, pays some of the uploaders (called “YouTube Partners”) and strikingly started awarding the “YouTube Awards” in 2007. While it is more accessible to the general public than Hollywood, it does not eliminate the need for some level of professionalism, originality and perseverance required of people who want to become popular on the site.

This may well be the reason why so much professionally created content finds its way into ‘user generated’ content - the process of creating something from scratch that is sufficiently polished for popularity on YouTube is too hard and tedious for most users. It is much easier to ‘mix and match’ with content that has already been created. The second best is creating content that requires hardly any technical skills, such as videos which mainly require that the video camera is switched on and some time later off. This content, however, will disappear if it does not offer

something interesting for particular groups or the general public, such as the video of the 'star wars' daughter or the popular 'mishap' videos.

The question then is – why do people upload videos if hardly anybody looks at the material? The answer might be that YouTube functions as a social filter or blackboard: people upload content that they think is somehow 'interesting'. It does not really matter whether they created this content themselves, or if they reuse something they found or captured somewhere else: it's a way to facilitate "watercooler" discussions, gossip, and other friendly chat. If someone missed the particular tv series episode, or the funny news item about a politician misbehaving, or the sports event, or your daughter's first steps, that you'd like to talk about - just upload it to YouTube, and everybody can join in. This enables the sharing of experiences within your social circle or for the general public. Additionally, YouTube's features of video responses and user-moderated comments support this view of YouTube as primarily a social tool, rather than a way of enabling promotion of user creativity through user-generated content.

The question then becomes: what determines what people find interesting? If YouTube really is a social filter, then what is it filtering for? Can we determine some set of tags or kinds of content that are more or less likely to be popular, or to be shared more widely? It is interesting to note that even within the limited sample that we took, people were sometimes surprised about the popularity of videos, even their own. The uploader of the aforementioned video of a 3-year-old explaining Star Wars updated the description of the video saying "What was expected to be a short movie that would circulate within a small circle of friends has turned into something much more very quickly." [11]

It would be interesting to see if future research would be able to predict the popularity of content. Based on the data samples we used, there are already some criteria that we can look at:

- From the data collected on the lengths of content, it seems that in general a video needs to be sufficiently long (about 3 minutes, at least).
- Most content that becomes popular is somehow funny, or extremely original.
- The content needs to be attractive to a large group of people: the Star Wars explanation satisfies this criterion very clearly by being attractive to everyone who has ever seen Star Wars, or knows what the story is about, which is a sizable portion of the current Internet user base.

More extensive study on experience representation and exploitation could formulate clearer distinctions of what does and does not become a 'hit' on YouTube. In the end, we have shown that even on YouTube, to be part of the popular top 10%, a video creator will need to put in quite a bit of effort. User generated content currently fails to beat professionally generated content in this popularity contest, and it is clear that popularity does not come easy.

This fact, however, hints to another line of research, namely the support of creativity as well as quality during the authoring process. Our study clearly indicates that user-generated content lacks both aspects and as a result most material will only be seen by the creator and his or her social circle. If tools could be offered that support the creative process cycle of planning, shooting and editing more people could be able to attract larger audiences – which would establish a better general discourse.

5. Proposed research direction

Addressing the outlined challenges for turning sites, such as YouTube, into something where people indeed can represent themselves in the realm of their possibilities, we aim for the following approach. Note: though our approach takes YouTube as the basis (mainly because we would have a proper data set on our hands to evaluate our methods and tools), the interfaces we wish to develop are aimed for being used in any social network oriented video site.

There are two directions we wish to explore. The first is to develop authoring mechanisms that enhance the creativity of a typical YouTube author. We will analyze the most general flaws, compared to film theory, in the captured videos from our sample and then establish aids that either help during the generation process or allow a basic quality evaluation while the video is uploaded to a social environment like YouTube.

The second research direction address the exploitation of metadata provided by YouTube users to establish new presentation forms of the material, such as biased presentation of news or automatically generated gossip shows based on the videos available in YouTube.

6. REFERENCES

- [1] Lee Gomes. Will all of us get our 15 minutes on a YouTube video? Wall Street Journal, August 2006.
- [2] Alexa. youtube.com - Traffic Details from Alexa, February 2008.
- [3] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. In IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pages 1–14, San Diego, California, USA, 2007. ACM.
- [4] Scott Woolley. Video fixation. Forbes, 178:100–106, October 2006.
- [5] Lev Grossman. Time's person of the year: You. Time Magazine, pages 38–41, December 2006. <http://www.time.com/time/magazine/article/0,9171,1569514,00.html>
- [6] Accenture. User-generated content is top threat to media and entertainment industry. Accenture Newsroom, April 2007.
- [7] Candace Lombardi. YouTube cuts three content deals. CNet News.com, October 2006.
- [8] RIPE NCC. YouTube Hijacking: A RIPE NCC RIS case study. RIPE News, February 2008. <http://www.ripe.net/news/study-youtubehijacking.html>.
- [9] Matt Preprost. I Found Your Camera CBC Radio Interview Promo, February 2008.
- [10] Pat Aufderheide and Peter Jaszi. Recut, reframe, recycle: Quoting copyrighted material in user-generated video. <http://www.centerforsocialmedia.org/recut>, January 2008.
- [11] fistofblog. Star Wars according to a 3 year old., February 2008.