A Classification Scheme for Content Analyses of YouTube Video Comments

1 Introduction

YouTube is one of the most visited sites on the Internet, and attracts 800 million unique visitors per month (Google, 2011[1]). It allows users to upload and share video clips on a diverse array of topics, and also incorporates a growing number of additional features that allow users to interact with the content and other users. Users can review or rate what they have watched, and associate comments with videos to express their opinions or respond to the video content.

The user comments facility has become a medium for other forms of communication unrelated to the video content. Comments are used for self-expression, providing emotional support, reminiscence, grieving and advice, as well as direct comments on the video itself. In this paper, we explore the types of comments provided on YouTube, proposing a classification schema to categorise comments on YouTube. This scheme was based on a content analysis of 66,637 user comments. Our classification schema consists of 10 broad categories referring to the major classes of comments and then a detailed breakdown into 58 subcategories which allow the classification of comments into coherent groups of comments.

Existing literature is reviewed in section 2 in order to contextualise a study of YouTube as a communication area and consider the challenges inherent in classifying social media comments. In section 3 we present the methodology employed to sample comments, the process of creating and refining a classification schema for YouTube comments, and a test of the inter-coder agreement of our scheme, and, in section 4, the details of the classification scheme created. Implications and future work are considered in section 5.

2 Background

In recent years the way in which people use the Internet has been evolving, with a perceivable shift towards increased user participation in uploading content (photos, videos, audio and textual information), sharing and recommending content, and leaving comments and ratings on the resources they access (Gill et al, 2007, p.15; Heckner and Wolff, 2009).

The sites that have facilitated this shift, by simplifying the processes of posting and sharing material to be accessible and practical for a widespread user base, are referred to under the heading of "Web 2.0" or "the social web". Keenan and Shiri (2009, p.439) identify two main types of social websites, people-focused and activity-focused. People-focused sites such as Facebook and MySpace are centred on users' profile pages, which contain personal information about the user, whereas activity-focused sites focus on the presentation of content such as videos (YouTube), photos (Flickr), audio material (last.fm) or links to other websites (Digg, del.icio.us).

The boundaries between these website types are becoming increasingly blurred, as Facebook and MySpace allow users to upload, link to and embed a variety of content types on their profile pages. On the other hand, some activity-focused sites like YouTube incorporate people-focused features like profiles, subscription and friending options, and user commenting facilities. These people-focused features can collectively be referred to as the "community aspect" of the sites. In this paper we are interested in how people use this community aspect of YouTube and for what communicative purposes.

[1] http://www.google.com/adplanner/static/top1000[Accessed 22/06/2012]

2.1 YouTube

YouTube was founded in 2005 and developed rapidly to become the largest video-sharing website on the Internet (Gill et al, 2007, pp. 15-16). Following its acquisition by Google, the site has continued to grow in popularity and now attracts almost half a billion unique users per month. The site allows users to upload an unlimited number of video clips, which can be viewed and linked to by anyone.

YouTube provides a variety of facilities to allow registered users to interact with each other and share their responses to content. One option is for users to post direct video responses to what they have viewed. Other options are afforded by the user profiles, or "channels": there is a private messaging service and a more visible commenting option on the profile, and users can choose to store and display lists of their favourite videos. They can also befriend other users on the site, or become "fans" by subscribing to receive alerts when they post new material. The site also provides sharing, rating and commenting options on individual videos.

YouTube also provides the option to set up a user comments facility underneath each video, which resembles a message board and allows registered users to post messages[2]. The uploader of the video can change the comment settings so that only certain users can comment, can require that comments are subject to moderation before they become publicly visible, or can choose to disable the comments facility completely. Comments are displayed in reverse chronological order, and if there are large numbers of comments they break over multiple pages.

There are certain restrictions placed on the commenting facility: there is a 500-character limit, and as an anti-spam measure it is forbidden to insert URLs or html tags. The uploader can remove comments that they do not wish to see displayed on the video, and users can flag up inappropriate comments or suspected spam for moderation.

Users can also "rate" comments using thumbs-up and thumbs-down buttons, as they can for videos (These buttons become visible when the cursor rolls over them.) The comments that gather the greatest number of positive votes are displayed above the main bulk of comments.

The comments facility was intended to be "a section of text for users to provide information related to a video"[3] and where users could express their opinion on the video. However, in practice this is not always what happens. As Hutchby (2001) discusses, technology provides affordances but is not prescriptive, and so can be put to uses other than that for which it was originally intended. This is the case with the comments facility: it is quite possible for a user to write things in the comments box that are not relevant to the video.

2.2 YouTube Research

Researchers are increasingly studying YouTube, in order to investigate user behaviour, measure video popularity and harness content for marketing purposes. Kousha et al. (Kousha et al., 2012) provide a particularly good review of the use of YouTube videos in a multitude of domains including marketing, medicine and management. Most of these studies, however concentrate on the video content rather than, as here, the content of user-generated comments.

---

[2] This description refers to YouTube's commenting facility at the time the research was conducted.

[3] http://www.google.com/support/youtube/bin/answer.py?hl=en&answer=171666 [Accessed 01/05/2011]

2.2.1 User behaviour

A number of researchers have been interested in looking at user interaction patterns on YouTube. Studies of user interaction behaviour are of particular use for marketing purposes, as companies can identify core or hub users and use this information to target their messages and products. This can allow them to disseminate their messages and product advertisements more quickly and effectively.

Paolillo et al.'s 2008 study for example analysed user profiles, including friending, subscription, favouriting and commenting, and identified that certain types of content were cultivated by users from particular social groups with shared characteristics (e.g. being of a similar age, having links to a particular location, or belonging to specific cultural groups) (Paolillo et al, 2008, p. 164). They demonstrated that producers of YouTube content tended to have strong links to others producing similar content, expressed through commenting and subscription (Paolillo, 2008, p.156), and identified an interesting core-periphery pattern within their analysis. It was found that a relatively small number of central participants on the YouTube site exerted a disproportionate influence on social interaction and the development and posting of content on the site (Paolillo, 2008, p.157). Canali et al. (2010) uncovered similar findings. In their study, assessing the strength of links between users in terms of in-degrees and out-degrees, they found that certain users had a noticeably higher proportion of fans in relation to invited friends, and termed these people "hub-users".

Chatzopoulou et al (2010) examined over 37 million videos, analysing properties like view counts, number of comments left, number of ratings given and number of times a video is favourited, in order to see which were most useful in indicating video popularity. They argued that responding by favouriting, commenting or rating was a stronger indicator of popularity than simply viewing a video, because it requires more effort to log in to express a reaction. They also claimed that reaction strength tended to be stronger amongst less-viewed videos. Such approaches are useful for understanding behaviour but also for developing YouTube itself. Cheng, Dale and Liu (2008), for example, show how information on small-world patterns within YouTube's user population can be used to design caching policies to improve system performance.

In contrast to these quantitative approaches, qualitative studies, such as Lange's (2007) study on video sharing, which reveal complex interaction patterns between YouTube users are becoming available. A good example of this is Kousha et al. who examine the type of videos which are cited in academic publications (Kousha et al., 2012).

2.2.2 Commenting behaviour

YouTube comments have, thus far, been comparatively understudied in relation to other aspects of the site, because the sheer number of comments, lack of structured organisation and the variable quality in terms of spelling, grammar and expression have presented considerable difficulties for conducting analyses. Small scale qualitative studies, such as Jones and Schiefflin (2009) have used comments to investigate language use associated with particular genres of video.

However, most research has focussed on interaction patterns relating to comments (Lee et al, 2010), and mining the content of comments in order to assess video popularity (Potthast, 2009; Potthast and Becker, 2010; Potthast et al, 2010; Stein, 2010).

Lee et al. (2010) examined the discussions and disputes in YouTube comments. Rather than looking at the content or language of comments, they focussed on identifying conversation patterns like turn taking between participants, by looking at authorship and timing of comments. They conducted an automated analysis on a large number of videos in order to reveal instances where there was prolonged interaction between small numbers of participants.

Other researchers are interested in summarising the responses of previous users in order to assist users in assessing the quality of a resource. Studies such as (Potthast, 2009; Potthast and Becker, 2010; Potthast et al, 2010; Stein, 2010) have focussed on mining comments and presenting the opinions expressed in a way that was accessible to non-specialist users. They developed a prototype application called "OpinionCloud" which extracts words with positive and negative connotations from comments and summarises them in "tag clouds". This was intended to help users gauge the relative levels of positive and negative responses to a video without having to sift through large numbers of comments.

Previous research has attempted to categorise comments left on other social websites like Twitter (Jansen et al, 2009), Digg and del.icio.us (Park et al, 2008), and these provide some useful starting points for analysis.

Park et al (2008) examined comments from del.icio.us and Digg and drew out a very general set of categories: Summary, Additional Information, Impression, Opinion and Etc. "Summary" referred to comments that repeated or lifted content from the resource being described. "Additional Information" referred to comments that provided information from other sources to contextualise the resource being described. "Impression" referred to comments containing very general responses to and opinions on the resource, using expressions like "Awesome!" or "Great!". "Opinion" referred to comments that provided an explicit judgement or justified expression of opinion regarding the content of the resource. The additional category, "Etc.", contained the comments that did not fit into the other four categories, or that could not be analysed because of issues with the language used in them which the computer software was unable to resolve.

Jansen et al (2009) looked at attitudes to popular brands expressed in comments on the Twitter website. They identified 23 different types of comment in their sample, which they collated into the table reproduced as Table 1. Whilst some of these categories (e.g. Order via Twitter, Patronizing) are clearly specific to a marketing study, others (like Supplement, Forwarding, Answer) have the potential to be adapted and applied more generally to comments on a variety of topics.

Action code	**Definition**
Announcement	Declaring the upcoming objects
Answer	Handling question
Chitchat	Casual conversation
Comment	Expressing mixed or neutral feelings regarding objects
Confirmation	Giving assurance or validation regarding objects
Consuming	Drinking or eating objects
Expecting	Looking forward to objects from Starbucks
Forwarding	Pointing to potential useful objects
Maintenance	Managing objects
Missing	Feeling from the lack of objects and expecting to have them back
Negative comment	Critiquing, complaining
Notification	Letting one know on objects
Order via Twitter	Attempting to place order on Twitter
Patronizing	Physically being in objects or going to objects frequently
Positive comment	Complimenting, praising
Question	Expressing confusions or doubts toward objects

Recommendation	Providing positive advice regarding objects
Recommendation request	Seeking advice regarding objects
Request	Asking for objects
Research	Examining objects
Response	Giving unnecessary feedback on objects
Suggestion	Providing ideas to improve objects
Supplement	Adding on to objects

Table 1: Comment functions, reproduced from (Jansen et al., 2009)

YouTube's popularity, and the ease with which comments can be left, has created a large repository of user-generated information which may be mined for studies on communication, information seeking and sociology. One of the few studies to examine the types of YouTube-hosted comments is (Thelwall et al., 2011) who conduct automatic analyses of comments examining typical length, topic (as determined by the official YouTube categories) and sentiment of comments. Such studies are useful in characterising commenting *patterns* and suggesting new ways in which comments can be analysed. Other researchers such as Knautz and Stock (2011) have experimented with YouTube to investigate how emotions can be classified within the videos presented on the site. In their work Siersdorfer et al argue that "the amount of community feedback in YouTube results in large annotated comment sets which can help to average out noise in various forms and, thus, reflects to a certain degree the "democratic" view of a community. (2010, p.895). In their work they focus on classifying comments based on community acceptance within the site.

In this work we focus on a different approach to classifying content: one based on textual analyses which will allow researchers to distinguish between types of comments and, thus, select types of comments for analysis or analyse differences between types of comments left on different videos. Later in the paper we describe a method for classifying user comments and present a classification scheme based on this method.

2.3 Classification

High quality classification "helps to capture relationships and links between different pieces of knowledge" (Dotsika, 2009, p.407). As noted above, the ability to accurately classify YouTube comments opens up multiple areas of research, and the ability to better understand emerging social media behaviour and its potential impact.

In terms of social media, the challenges inherent in attempting to classify the types of comments emphasized above have been discussed in the literature. The complexity of classifying dizzying arrays of diverse content "produced minute by minute throughout the world, in multiple languages, using varied media, to different quality levels, by a huge, diverse population" raises major issues (Cosh, Burns and Daniel, 2008, p.724). Rowe et al agree that the "unprecedented, rapid growth of content is naturally producing new challenges for its analysis and use" (2011, p.315).

The concept of the folk taxonomy, or folksonomy, has become an inherent part of Web 2.0 culture (Keshet, 2011, p.145). While a useful approach for the organisation of the media content, it is arguably less useful in the categorisation of comments about those media, since the discussions can often veer off the topic at hand and into debates and arguments that have no bearing at all on the media posted.

On the other hand, classifying content in social media necessitates an understanding of the medium itself. In other words, when classifying a book, the content is all; however in classifying comments

on a social media site, the context is as important. As Mai argues more generally, "to understand and evaluate the classification one needs to understand and, to a certain degree, be part of the social context in which the classification is used." (Mai, 2010, p.721)

3 Methodology

Our approach to creating a classification scheme is based on qualitative content analysis. Content analysis is a research methodology that involves making the content of messages manifest through identification of characteristics in as objective a way as possible (Bryman, 2008, p.273). The intention is to create a coding schema, or "membership categorisation device", which encompasses a collection of categories and a set of rules on how to attribute data to these categories (Silverman, 2005, p.379). The schema is developed iteratively by looking at the data, identifying possible categories and then testing these on further data to see if they can be applied in their current form.

A key element of the process is known as deviant case analysis: this is where any data that does not fit into the existing schema is reconsidered, and then categories are either created or adjusted in order to accommodate this (Silverman, 2005, pp. 214-215, 377). The aim is to approach a point of theoretical saturation whereby new data does not reveal any new insights but instead confirms what is already known (Bell, 2005, p.20). In the case of the categorisation schema, this would be the stage at which there were enough categories to accommodate all the comment types.

Our overarching aim was to produce a general coding schema that could be used on comments across all genres of videos, and so the corpus of comments used to create the schema therefore had to contain examples from a wide range of genres. Using the "Browse" option, videos were filtered by category and then the "Most Discussed" videos of the preceding month were accessed. From this the 3rd, 5th, 7th and 9th most commented videos in each category were selected to obtain videos with a large number of comments. A problem was encountered with the "Education" category, which did not allow for videos to be filtered by "Most Discussed". For this reason, the videos selected from the "Education" category were 3rd, 5th, 7th and 9th on the "Most Watched" list. This resulted in a corpus containing 66,637 comments drawn from 60 videos. Using highly commented videos meant that we had a rich source of comments. It also meant that we were dealing with popular and/or controversial videos which may have a lot of certain types of comments. A bias towards heavily commented videos should not, however, mean that we miss *types* of comments. To test this we ran a coding reliability test (described below) on less popular videos which revealed no new types of comments other than those which occurred in the classification scheme presented in section 4.

As in any communication environment, YouTube comments attract a broad spectrum of opinions, some of which can be controversial, racist or offensive. Many videos, particularly under "Sport" and "News and Politics" attracted a lot of profanities and abusive content, whilst others had content veering into areas of racial provocation or fetishes. However, no comments were removed on the grounds of content and we did not exclude any video on the grounds of copyright violation.

The videos encompassed a broad range of interests, styles and addressor characteristics. Some were posted by large organisations like the ITN News agency, Stanford University and the BBC's TopGear programme, whereas others appeared to originate with private individuals. Some were constructed in a way that suggested careful attention had been paid to stylistic features and the editing process, whereas others (especially in categories like "People and Blogs") had been roughly filmed using web cameras and appeared to have undergone little or no editing. Additionally, whilst some videos focussed on major news stories and sporting events like the World Cup, others focussed on much smaller scale domestic topics and individual experiences, such as a tribute to a much-loved pet. The majority of these "Most Discussed" videos were in English, the only exceptions being an

Italian singer in the Comedy category and a Vietnamese performer at a charity concert under "Non-Profits and Activism".

All comments for the selected videos were collected on the 18th July 2011 and stored separately in case comments or videos were removed.

3.1 Devising Categories

In order to generate initial ideas for comment categories, we compared and contrasted the comment categories in the studies by Jansen et al (2009) and Park et al (2008). This indicated significant areas of overlap around concepts like opinions, description and information seeking and supply.

The Jansen et al study (2009) presented a variety of rather specific categories relating to marketing and branding on Twitter. It contributed categories like "order" or "consume" that might not map directly onto the YouTube environment, but could be considered equivalent to concepts like "watching videos" or "request posting of another video" that would come under Site Processes. It also contained useful ideas relating to recommendation (users might well recommend other videos) and anticipation of future products (videos).

The Jansen study also indicated that there might be value in subdividing some of the broader categories. For example, within the broad heading of "Give Information" it could be useful to separate an initial direct answer to a question from a supplementary response that confirmed or disputed the initial direct answer. It might also be useful to draw a distinction between positive and negative impressions and opinions.

The Park et al study (2008) presented much more general categories, which were useful in indicating areas where categories could be linked or separated from each other. It confirmed that giving and eliciting opinions could be united under "Opinion" and "Ask for information" and "Give information" could be covered by "Information". However, it also indicated that there was value in keeping certain of the proposed categories separate rather than combining them. For the schema, it was decided that immediate impressions should be treated as conceptually different to more considered opinions, and that comments relating to description of the video's content should be categorised separately from supplementary information from other sources and authors.

The possible categories identified were tested iteratively against the corpus. The comments were examined to see if they could be allocated to one or more of the categories, although the exact types were not recorded for each comment because of the impracticality of constantly revising them to reflect changes in the category list[4].

Over time, it became clear that certain categories could be combined or omitted altogether because they were uncommon or absent in the corpus. At the same time, new ideas were incorporated into the list. Through analysis of "unclassifiable" comments, other ideas were devised and incorporated into the list.

This eventually led to the production of the category list presented in Table 2 below. In Table 2 we present 10 broad categories, 58 subcategories and an operational definition of how comments are to be classified within each subcategory. For simple definitions some clarification is provided within

[4] A second test where we classified a new set of comments to provide a quantitative account of differences in the type of comments left for different genres of video is reported in (Madden, 2011).

Table 2; for more complex definitions and subcategories which are more interesting a fuller discussion is given in section 4. Categories are not mutually exclusive as comments may contain text that falls under more than one heading. We do not present a prescriptive use of our classification scheme, i.e. say that comments should be classified under only one major category or in multiple categories as such decisions have to be informed by the underlying research questions behind the use of this classification scheme.

3.2 Testing Inter-Coder Reliability

The categories noted in Table 2 were derived iteratively and approximately 98% of the comments could be attributed to at least one of the categories, with the remainder assigned to the 'unclassifiable' category.

To test the robustness of the classification scheme and the clarity of the descriptions, two volunteers who were not involved in the study were asked to separately classify the first one hundred comments on a single video which was not used in developing the scheme. The comments were also classified by one author. Although the test demonstrated good classification power, it highlighted some areas of confusion in distinguishing between Information and Advice comments. Changes were made to the classification descriptions to better differentiate between these categories and address any other weaknesses in the scheme. These changes have been incorporated into the final scheme in Table 2.

This scheme was then subject to a new test of inter-coder reliability using the Kappa Statistic and using two new volunteers who were not associated with the development of the scheme and had not used the scheme before. We selected 100 new comments; 1 from each of 100 recently posted videos across the range of YouTube categories. Unlike the original data collection we randomly chose videos rather than focussing on popular videos. The volunteers were given the coding scheme as it appeared in Table 2 along with the textual descriptions that appear in section 4 for situations in which further details were required on how to classify a comment into categories. They were also provided with links to each video. The coders were instructed to categorise each comment into *any* category in which they felt the comment belonged. The interrater reliability for the raters was found to be Kappa = 0.608 ($p < .0.001$), 95% CI (0.51, 0.706) which is generally seen as a good level of agreement. A second Kappa, using only the top-level categories, resulted in a Kappa score of 0.671 ($p < .0.001$), 95% CI (0.57, 0.77) indicating that whilst the coders agreed more on the top-level categories there was some difference of opinion in how these categories should be used. We return to this in the discussions section.

Category	Subcategory	Description
Information	Request	A comment containing a request for an explanation of or further information about something: this may be to do with video content, video context or something unrelated to the video (section 4.1)
	Make comparison	A comment that compares one thing to another: this could be video content, a person in the video, something in the context of the video or something unrelated (section 4.1)
	Give (video content)	Any comment which provides factual information about something directly featured in the video: might be in response to a request or might be unsolicited (section 4.1)
	Give (video context)	Any comment which provides factual information about something related to the video but not directly featured in the video: in response to a request or unsolicited (section 4.1)
	Give (general)	Any comment that provides factual information on something. Might be in response to Information: Request or might be unsolicited (section 4.1)
Advice	Request	A comment in which a person asks for ideas or assistance (section 4.2)

	Give	A comment in which a person offers ideas or assistance (section 4.2)
	Recommend another video	A comment which signposts the reader towards another video on YouTube (section 4.2)
Impression	General	Short comment expressing immediate reaction (section 4.3)
	Positive	Short comment expressing immediate reaction and positive sentiment (section 4.3)
	Negative	Short comment expressing immediate reaction and negative sentiment (section 4.3)
Opinion	Request	Comment that asks a person/people to give their views on a video or topic (section 4.4)
	Give (general)	Comment expressing the commenter's view on a person, video or topic (section 4.4)
	Give (positive)	Comment expressing the commenter's view on a person, video or topic and expressing positive sentiment (section 4.4)
	Give (negative)	Comment expressing the commenter's view on a person, video or topic and expressing negative sentiment (section 4.4)
	Give (mixed)	Comment expressing the commenter's view on a person, video or topic and expressing a combination of positive and negative sentiments (section 4.4)
	Insult	A comment intended to hurt or offend someone (section 4.4)
	Compliment	A comment which says something nice to/about someone (section 4.4)
	Criticism	A comment that picks up on a fault with something or someone to do with the video or a previous comment (section 4.4)
	Tribute	Similar to "Compliment", but specifically with reference to a deceased person/character/ being (section 4.4)
	Speculate	A comment in which the writer hypothesises about the outcome of a future event/process (section 4.4)
Responses	Agree	A comment that expresses agreement with something in a previous comment or the video content (section 4.5)
	Confirm	Like "Agree" but with evidence / justification (section 4.5)
	Disagree	A comment that expresses disagreement with something in a previous comment or the video content (section 4.5)
	Challenge	Like "Disagree", but with evidence/justification (section 4.5)
Expression of personal feelings	General	A comment in which the writer expresses their personal feelings or emotional response (section 4.6)
	Positive	A comment in which the writer expresses their personal feelings or emotional response (specifically expressing positive sentiments) (section 4.6)
	Negative	A comment in which the writer expresses their personal feelings or emotional response (specifically expressing negative sentiments) (section 4.6)
General Conversation	Greetings	Any comment that fulfils the linguistic function of greeting: may be intended to initiate conversation (section 4.7)
	Thanking	A comment that thanks someone for doing or saying something (section 4.7)
	Joke	Comment that (is/intends to be) humorous (section 4.7)
	Apology	A comment which fulfils the linguistic function of apologising: so the writer will express sorrow/regret for something said/done (section 4.7)
	Reference to offsite communication	A comment referring to communications taking place outwith the YouTube site, e.g. social networking site, instant messaging, telephone, meet in person etc. (section 4.7)
	Status description	A comment that describes what the commenter is doing at the time of leaving the comment (section 4.7)
	Anecdote	A comment containing a story about a commenter's personal experience (section 4.7)

	Random off-topic exclamation	A comment that bears no relation to anything that precedes it: perhaps intended to be controversial and provoke a reaction or start off a new conversation. (section 4.7)
	Request personal information	A comment in which the writer asks someone for information about him/herself: name, occupation, physical characteristics etc. (section 4.7)
	Give personal information	A comment in which a person volunteers information about him/herself: name, occupation, physical characteristics etc. (section 4.7)
	Express plans/desires/wishes	A comment in which the writer identifies something s/he wants to do (section 4.7)
	Anticipate	Comments in which the writer discusses their (positive) expectations of a future event (section 4.7)
Site processes	Posting videos	A comment which talks about the action/process of posting videos (section 4.8)
	Suggest content for a future video	A comment which puts forward an idea for something that the uploader (or someone else) should put in a future YouTube video (section 4.8)
	Request posting of another video	A comment in which the viewer asks for another video to be posted (section 4.8)
	Watching videos	A comment which talks about the action/process of watching videos (section 4.8)
	Commenting process	A comment which talks about the action/process of leaving user comments or makes a general comment about the body of comments already attached to a video
	Rating features	A comment which talks about the action/process of rating a video, using the like/dislike buttons
	Sharing videos	A comment which talks about the process of sharing a video using the buttons underneath the video stream
	Profiles and subscription	A comment that talks about the processes of subscribing to channels on YouTube or interacting with other users via their profiles or personal messages.
	Site policy	A comment which contains an observation or judgement about a feature of YouTube's site policy
	Site design	A comment which contains an observation or judgement about a feature of the YouTube site design
Video Content Description	Summary (paraphrase)	Comment that summarises something in the video content- paraphrases (section 4.9)
	Summary (direct quote)	Comment that directly quotes something in the video (section 4.9)
Non-response comments	Unclassifiable	A comment that doesn't fit any of the existing categories, and is therefore a diagnosis of exclusion.
	Spam	Comments that have been marked as spam by YouTube's automated systems or by users (section 4.9)
	Removed comments	Spaces where a comment previously existed but was removed by the commenter/moderator. These comments are demarcated by the message "This comment has been removed". This could be because the author changed their mind and wanted to remove their remarks from public view, or because the poster of the video disagreed with the content or viewpoint expressed in the comment, and so deleted it.
	Repetition	A comment that has (deliberately or accidentally) been published multiple times. These comments are completely identical to those surrounding them, and have the same author.
	Non-English	Comments not in English
	Nonsense words/ punctuation	Comments that are not in a recognisable language, or which are composed entirely of random punctuation marks.

Table 2: Category List with Definitions

4 Comment Characteristics and Classification Notes

This section presents a more detailed account of the categories, including information on how categories are differentiated from each other. Examples from the comments we analysed are provided for illustration.

4.1 Information

Information comments are those that request or provide factual information about something in the video content, video context or a completely unrelated topic.

Information: Request comments will ask for details, confirmation or explanation of something, and containing questions such as "What is X?" "Where did X come from?" "How did you do X?" or "Is that X?". They may contain reference to specific features of the video, or deep links. Some will be targeted at specific people like the poster of the video or a previous commenter, via @tags, usernames or personal names, whilst others will be open and aimed at a general audience.

> *First of all great video.. Second, dude where the heck are you riding? Japan?*

Information: Make Comparison comments are characterised by the use of constructions like "X is bigger/smaller/better/worse than Y", "X reminds me of Y" or "X is nothing like Y". This category is included under Information, because such comments are providing information that relates to a comparison. This category is closely related to the Opinions category (see section 4.4), because to say, for example, that one thing is better than another involves a subjective value judgement.

Information: Give comments are those that provide details or explanations, either unsolicited or in response to an earlier "Information: Request". Comments that provide information often contain direct statements, like "X is…", "X comes from…" or "I did X by…". Sometimes, these comments will give justifications, and some may make reference to external sources by recommending other videos, providing search strings to use in search engines, or providing modified URLs to point to specific online resources.

For the purposes of this classification schema, the **Information: Give** category has been split into three parts, namely Video Content, Video Context and a General category. A comment giving information on **Video Content** will make a specific reference to an element of the video. This may be achieved with a deep link, by using a phrase like "the bit where X happened" or by referencing something in the video.

> *hes in china or some asian country because he looks at the road (5:13) and its that asian writing*

Comments relating to **Video Context** will generally make reference to something related to the video, but not directly featured in the video. For instance, the below example is taken from a video featuring a road test of a Vauxhall car in the UK, and this comment provides contextual information about equivalent vehicle models in other countries.

> *The old VXR was equivalent to the Pontiac G8 but this newer VXR is equivalent to Buick Regal.*

The **General** category encompasses comments which provide factual information about something that is unrelated to the video, or where it is impossible to attribute the comment to either the content or context categories.

4.2 Advice

Advice comments are those in which people request or give assistance or suggestions on what to do in a particular situation. They are specifically requesting or giving the particular judgements of the individual commenter, unlike Information comments, which focus on objective facts. This is reflected in the language employed: Information comments tend to contain definite statements like "It is" whereas Advice comments tend to contain modal verbs like "could", "should" or "would".

Advice: Request comments will generally include questions like "What should I do to achieve X?" or "Where can I find/buy X?". Some will be targeted at specific people such as the poster of the video or a previous commenter, using @tags, usernames or personal names, whilst others will be open and aimed at a general audience:

> *Hey guys (: I have very pale skin and rather pale lips.. I can't wear red lipstick because they're too strong. What colour of lipstick should I wear? Love your videos x*

Advice: Give comments are rarely unsolicited, usually appearing in response to an earlier Advice: Request. They tend to make use of constructions like "You could do X" or "Why not try X" or "I think you should…".

> *Bubz you should wear less make up :)*
> *I mean, you can wear it if you want to because it's your life but it sort of gives off the wrong impression. Like you think being beautiful is all that matters. You are also sorta giving the wrong impression to men as well :*
> *And you are a really good guru, I just think you should embrace your natural self more :)*

Advice: Give comments differ from Opinion comments because in Opinion comments the addressor is expressing an opinion on a situation, person or topic, whereas in Advice the addressor is recommending actions that can be taken to accommodate, change or remedy a situation.

Advice: Recommend another video is included under the category of Advice, because such comments can be interpreted as one person giving advice to another about which videos to watch. Unlike Information: Give comments, which use a recommendation as part of a comment, in this subcategory the main action is a recommendation.

4.3 Impression

Impression type comments express people's immediate reactions to what they have watched in a video or read in comments. They tend to be very short, sometimes even single words, and will generally be fragments as opposed to complete sentences. They will often contain exclamation marks and paralinguistic features like emoticons, as well as interjections like "Yay!", "Boo!" and "Wow!". As discussed by David Crystal (2008), these features are employed to convey personal responses and imitate features of spoken conversation.

This Impression category is split into three subcategories: general, positive and negative. This could potentially be useful for gauging the tone of a particular discussion.

Positive impressions include words and emoticons like "Awesome!", "Lol" and ":D", whereas negative ones can be recognised through the use of words and emoticons like "Ewww!", "Boring",

":/" or ":(" . The general strand encompasses terms like "wow" and "oh", as these could be positive or negative and it is not always possible to determine which from the context.

A difficulty that can be encountered is the use of slang, because terms like "sick" which have negative connotations can be used to convey a positive reaction. There may also be implicit sarcasm in the use of words like "great", but because of the short comment length, there may be few contextual clues to help determine whether the intended sentiment is positive or not.

4.4 Opinion

Opinion comments are those in which commenters request or give their points of view on a video, person, object or topic. They differ from Advice comments because they are not encouraging actions, and are unlike Information comments because they contain explicit and subjective value judgements. They can be distinguished from Impression comments because they are longer, are more likely to be complete sentences, and are more likely to contain some form of evidence or justification for the perspective offered by the commenter.

Opinion: Request comments ask people to volunteer their perspectives on a video, person, object or topic, and may thereby aim to initiate a discussion. They will employ constructions like "What do you think of X?", "Isn't X…?" or "Do you think X is…". Some will be aimed at a general audience, but they may also be targeted at specific individuals like the video poster or a previous commenter:

> *@izzacrazy do you really think she did better than Shakira?*

Opinion: Give comments express people's views on a video, person, object or topic, and are characterised by the use of opinion words like good, pretty, right, wrong, funny, happy, sad and boring. Some will employ verb constructions like "I think…" and "I believe…" "I love…" or "I hate…" as well as expressions like "In my opinion…". They might also express preferences using terms like "favourite" or "least favourite". As with the Impression category, Opinion: Give is split into three strands, General, Positive and Negative. The decision on which category to use is again determined by the use of words with particular positive or negative connotations.

> *this is my favorite song of yours!!!!! BEAUTIFUL*

> *wow i really hate your songs*

Under the general banner of Opinion comments, there are also a small number of more specific comment types.

Compliments focus on a particular person and say something positive about them, like "You are really pretty/talented/clever"

> *You guys are far and away the best makeup artists on YouTube... You and your sis are so talented; please keep helping us ladies out!*

Tributes pay reference to a deceased person (or creature, as in the example below):

> *HAMISH NOOOOO. R.I.P. U WERE SO FREAKIN CUTE! RIP from Germany*

By contrast, **Insults** are akin to Opinion: Give: Negative, but are targeted directly on an individual or group of people with the intent of causing hurt or offence:

I'm not even going to bother with a come back for you. There is no need. You've obviously got the brain capacity of pond life.

Criticisms are also similar to Opinion: Give: Negative, but pick fault with a specific part of a video or comment, or a specific characteristic of a person. These are not always entirely negative: they may be qualified with advice on how to improve upon or repair a perceived fault:

I think that James May should just stick to the challenges with the other guys. I don't really like how he reviewed this car, I hardly laughed and didn't really have that much fun watching it. Useful information though.

The other type of comment that is contained under Opinion is the **Speculate** type, in which a person hypothesises about the outcome of a future event. They will generally employ future tenses and constructions like "I think X will happen" or "I bet X is not going to happen". Many comments of this type will make reference to a specific event related to the video topic, for example a sporting event or television talent show.

i bet holland will win!

4.5 Responses to Previous Comments

This is a placeholder category, containing a number of subcategories where commenters either support or dispute information or opinions found in previous comments. It is identified as a freestanding category because it crosscuts Opinion and Information. There are four subcategories within this category, two for supportive comments and two for disputing comments.

Agree comments are those in which the commenter says that they share the opinion of a previous commenter or the video poster, or accept that what someone has said is factually correct. They will express positive sentiments, and may use phrases like "I agree", "That's right" or "Exactly". They may also reference something from the video or preceding comment to demonstrate what it is that they're in agreement with. **Confirm** comments are essentially the same as Agree comments, but will also contain some supplementary evidence or information to back this up.

yea its true i seen that alot

Disagree comments are those in which the commenter says they don't agree with something that has gone before them. They express negative sentiments, and may say things like "that's wrong", "It's not" or "Nonsense!" (or more colourful variations on this theme). **Challenge** comments are like Disagree ones, but will either provide evidence to back this up, or demand that the previous commenter or video poster provides evidence to justify their own point of view.

No he doesn't always choose the right one he chose the left one for germany vs England

4.6 Expression of Personal Feelings

These are comments in which the writer describes their personal feelings or emotional responses to the video content, video topic or something said in a previous comment. They are lengthier and more detailed than Impressions, and will make use of adjectives of emotion like "happy" "sad" or "angry". They will often employ constructions like "I feel", "I love" or "I hate", and may also reference physical expressions of emotion, e.g. "That made me cry" or "That made me laugh". As for Opinion

and Impression comments, this category is subdivided into three strands, General, Positive and Negative.

> *Seriously you two are STUNNING ! And I get soooo happy when I see you've posted a new video! Love your videos :) xo*

4.7 General Conversation

This category encompasses a range of comment types that fulfil particular purposes in initiating and maintaining conversations. Some of this conversation might relate to the video's topic and content, but other examples will have nothing to do with the video to which they are attached.

General Conversations: Greetings are comments like "Hello", "How are you?" or comments about the weather, which are intended to initiate conversation, by demonstrating that the commenter is present and available for communication, and thereby opening up a psychological communication channel.

General Conversations: Status descriptions are generally short expressions of what the poster is doing at the time of writing, and will resemble status updates on social networking sites like Twitter and Facebook. They will tend to be focussed around action verbs like "I'm watching" or "I'm reading", or may resemble the Expression of Personal Feelings discussed in section 4.6. The difference with a status description is that the poster describes their emotional state and not their emotional response. So e.g. "I'm happy" comes under status description, because the influencing factor is not made clear, whereas a comment such as "Watching this video makes me really happy" comes under Expression of Personal Feelings because it states why the poster feels that way.

Other comments under the General Conversation heading that have a social interaction/networking element to them are **General Conversations: Give Personal Information** (in which commenters describe themselves by providing details about themselves like their name, occupation or physical characteristics) or **General Conversations: Request Personal Information** (in which they ask other commenters or the video poster for such personal details). These comments could support the formation of online relationships/friendships and act as conversation starters. They could also provide useful information for researchers to use to investigate content authorship and contextualise other comments made by the same user.

Some comments under General Conversation will make reference to communications that take place via a medium/channel that is not part of the YouTube site (those which take place using personal messaging or other YouTube communication channels are contained under Site processes, section 4.8). **General Conversations: References to offsite communication** concern conversations that take place online (e.g. by email, instant messaging or social networks), by telephone, or in person if the people know each other and have physically met each other before. These may reference particular channels (e.g. "Remember what I said when we were in London?" or "Ask your mum to phone me") This category also encompasses requests to add people as friends on social networking sites outwith YouTube.

General Conversations: Thanking comments are simply those in which the commenter expresses their gratitude for something that has been posted in a previous comment (e.g. a positive opinion, positive impression or compliment) or that has been done before (e.g. the act of leaving a comment, the act of putting up a video). They will employ words and phrases like "Thank you", "Cheers" or "Ta".

General Conversations: Apologies are comments in which the commenter expresses regret or sorrow about something previously said or done, and will employ terms like “sorry”, “apologise” or “no offence meant”.

General Conversations: Jokes are comments that are, or are intended to be, humorous. It is difficult to determine exactly what constitutes a joke because humour is very subjective and context-specific. However, the presence of terms like “haha”, “lol” or “that’s (not) funny” in the comment itself or those surrounding it can provide some context.

General Conversations: Anecdotes are comments in which people tell stories about personal experiences and tend to employ the past tense, using constructions like “I went there once”, “I remember when I/we…” or “I saw…”. Unlike Information comments the information and descriptions put forward are unlikely to be externally verifiable. Comments which can be classed as anecdotes may also contain an Advice element, as people draw upon personal experience to illustrate why they recommend (or do not recommend) a course of action.

> *Beautiful video , You & Minty were perfect together and his memory is held in everyone's hearts on youtube . I know how you feel about the anxiety , I havent been out with my friends in over 2 years and I have dropped out of school and college and have been doing nothing for over a year and have spent everyday of it with Psycho i have panic attacks just walking to the yard and suffer stages of depression , i can understand completely what Minty was to you as that is what Psycho is to me , Rip x*

Commenters sometimes talk about things they plan or want to do in the future. These are grouped under the heading **General Conversation: Express plans/desires/wishes**. Unlike anecdotes, these refer to events that have not happened yet (and may never happen), and so they tend to employ future tenses and conditional forms like “I’m going to…” or “ I wish I could...”

> *But i'd LOVE to come to Texas and experience everything, especially your wonderful storms!*
> *:-)*

General Conversation: Anticipate type comments are those in which commenters say that they are looking forward to something that is scheduled to happen in the future, in contrast to Express Plans/Desires/Wishes comments, which refer to indefinite or hypothetical situations. Unlike a positive Express Personal Feelings comment an Anticipate comment pre-empts a future event, whereas Express Personal Feelings are reactions to existing stimuli. Anticipate comments will use expressions like “I’m looking forward to…” or “I’ll be really happy when…”

4.8 Site Processes

Many of these comment types are self-explanatory: comments about Watching Videos will use words like “watched” “viewed” or “saw”, whilst **Site Processes: Posting Videos** will contain specific references to the process of uploading content or the tools employed in so doing. If they have enjoyed a video, commenters may request another video, **Site Processes: Request Posting of Another Video**, either in general terms, or with reference to a specific topic. Commenters may also suggest content, **Site Processes: Suggest Content for a Future Video**, putting forward ideas for things that the uploader, or other YouTube users, should feature in their future output. There is overlap between these categories and the Information: Request and Advice categories.

Many of the other sub-categories reflect co-presence, people discussing things they can both see. In the case of YouTube this encompasses features of the Site Design, like the general layout, recent changes to the user interface, or new tools introduced for users. These often provoke discussion amongst regular users, which can spill over into the comments threads. Site Design comments may employ technical terminology or slang terms which are subculturally specific to YouTube users. Commenters may also make reference to YouTube's underlying Site Policy, like their rules on permitted content, or to sanctions like account blocking that are imposed when these rules are broken.

Commenters may also discuss the use of YouTube's Rating Features, either by making reference to the tools or by speaking of the number of "likes" or "dislikes" a video or comment has attracted. Some commenters also encourage readers to "like" their comment in order to make it appear in a more visible position at the top of the comments thread.

4.9 Video Content Description

These comments can either contain direct quotation of words and phrases used in the video, or paraphrased descriptions of the visual or audio content. They may also contain "deep links", which are hyperlinks in a time format like "03:56" that direct the reader to a specific point in the video.

at 31sec the dhl commercial board gets destroyed at 4 m height!!!!!!

Comments of this type have to be categorised with close reference to the video content. Often, a judgement will be required in order to decide whether the comment comes under this heading, or under the more general category of information. The key differentiating feature is that a Video Content Description will contain either a direct quote from the video or a description referring only to what is shown in the video, without adding outside information to contextualise it. So, for example, a comment referring to the fact that there is piano music playing in the background of a clip could come under Video Content Description, but another that named the specific piece of music would fall under Information: Video Content.

4.10 Non-response categories

Spam comments are those that convey unwanted and unsolicited messages like advertisements or links to dating or pornography sites. In this schema, Spam specifically refers to comments that have been demarcated as spam by YouTube's automated systems, or that have been flagged up by users. Such comments are hidden from view and replaced by the message "This comment has been flagged as spam". There is, however, an issue in relying upon this flagging to classify comments because sometimes people maliciously or anti-socially flag comments they disagree with as spam simply in order to cause them to be hidden from general view. This is particularly apparent on videos that precipitate heated discussions.

5 Discussion and conclusion

This study contributes a classification schema that can be used to classify user created comments left on YouTube. The benefit of a classification schema is that it allows researchers and practitioners to identify particular types or uses of comments and separate these out for analysis. These analyses can be qualitative, to investigate language use or communication, or quantitative, as in areas such as online advertising and sentiment analysis. Existing approaches to such problems often separate comments by the video to which the comments are attached or ignore differences between comments. Our approach presents a classification scheme that classifies comments according to the purpose of leaving a comment. This, of course, is only one way to classify comments and other schemas are possible. The value of a schema is in the uses to which it can be put. Using this schema

(Madden, 2011) showed how the schema could reveal interesting differences between commenting behaviour on videos from different genres.

The classification scheme we propose is a detailed scheme that covers impressions, advice, opinions, and comments based on the use of YouTube itself. It also covers unusual uses of YouTube: as a means of paying tributes to the deceased, expressing personal desires and making recommendations on how to behave. Many of the categories cover personal reactions and impressions and are highly subjective. As such these categories are highly amenable to research on opinion-mining and also to research on video indexing through textual annotations. Many comments can be classified based on the content of the comment alone; in other cases the video needs to be watched closely in order to properly classify the content.

The scheme possessed good inter-rater consistency. However, the differences between some categories are still subtle: for example the classes Impressions and Opinions (immediate reactions expressed as short textual bursts vs. more considered reactions expressed as fuller sentences) are close in nature. Similarly comments giving information and giving advice may appear similar. In the coding scheme we have provided guidance on how to differentiate these classes. Several categories, such as Advice or Opinions, have positive and negative sub-categories. Deciding on the orientation of a comment is not always simple and gives rise to a certain level of subjectivity in the classification process. As with any classification scheme the more it is used then the more we learn how to use it and how to improve it.

The quality and applicability of the list is dependent on how representative the corpus was. Our approach selected a range of videos from across the main YouTube categories. Although we also carried out informal searches for videos that might contain comments that lay outside our schema, further investigation using other videos and comments might reveal the existence of new categories that are missing from this list altogether, or that there are other subdivisions that could be introduced to create more specific categories.

In future work it might be helpful to investigate issues such as word frequency and more detailed linguistic analysis in order to identify if there are particular words that help demarcate individual categories. The guidance notes in their current form do indicate a number of standard words and phrases that may be helpful in classifying, but these are not yet extensive and nor do they incorporate variations. If certain words could be highlighted as category markers it would facilitate automated analysis and allow the examination of a larger corpus.

As Thelwall et al. note (2011), even though the minority of YouTube viewers leave a comment, the sheer number of users have generated a large repository of textual material that may be successfully mined for different purposes. Further ‘the site should not be treated as an undifferentiated mass but as a place that is used by different audiences in different ways’ (Thelwall et al., 2011, p15). The categorisation scheme we present here is one tool which may be used to understand these audiences, their purposes and their communicative acts.

6 References

Bell, J. (2005) *Doing your research project: A guide for first-time researchers in education*. Maidenhead: Open University Press

Bryman, A. (2008) *Social Research Methods*. Oxford: Oxford University Press

Canali, C., Colajanni, M. and Lancellotti, R. (2010) *Characteristics and evolution of content popularity and user relations in social networks.* IEEE Symposium on Computers and Communications. 750-756

Chatzopoulou, G., Sheng, C. and Faloutsos, M. (2010) *A First Step Towards Understanding Popularity in YouTube*. INFOCOM IEEE Conference on Computer Communications Workshops – San Diego, California, USA; 1-6

Cheng, X, Dale, C. and Liu, J. (2008). Statistics and Social Network of YouTube Videos 16th International Workshop on Quality of Service, 2008. IWQoS 2008. pp.229 – 238

Cosh, K.J. and Burns, R. and Daniel, T. (2008),"Content clouds: classifying content in Web 2.0", *Library Review*, 57 (9) pp. 722 – 729.

Crystal, D. (2008) *txtng: the gr8 db8*. Oxford: Oxford University Press.

Dotsika, F. (2009) "Uniting formal and informal descriptive power: Reconciling ontologies with folksonomies" *International Journal of Information Management* 29. pp.407–415

Gill, P., Arlitt, M., Li, Z. and Mahanti, A. (2007) *YouTube Traffic Characterization: A View From the Edge.* Proceedings of the 7th ACM SIGCOMM conference on Internet measurement –San Diego, CA, USA. 15-28

Heckner, M. and Wolff, C. (2009) *Towards Social Information Seeking and Interaction on the Web* http://epub.uni-regensburg.de/6761/1/PreprintISI2009HecknerWolffTowards_Social_Information_Seeking.pdf [Accessed 01/05/11]

Hutchby, I. (2001) *Texts, technology and affordances.* Sociology: The Journal of the British Sociological Association; Vol. 35, 441-456

Jansen, B.J., Zhang, M., Sobel, K. and Chowdury, A. (2009) *Twitter Power: Tweets as Electronic Word of Mouth.* Journal of the American Society for Information Science and Technology, 60, 11, 2169-2188

Jones, G. M. and Schiefflin, B. B. (2009). *Talking Text and Talking Back: "My BFF Jill" from Boob Tube to YouTube*. Journal of Computer-Mediated Communication, 14, 4, 1050-1079

Keshet, Y. (2011),"Classification systems in the light of sociology of knowledge", *Journal of Documentation*, Vol. 67 (1) pp.144 – 158.

Knautz, K. and Stock, W.G. (2011),"Collective indexing of emotions in videos", *Journal of Documentation*, 67(6) pp.975 – 994.

Kousha, K., Thelwall, M., and Abdoli, M. (2012). "The role of online videos in research communication: a content analysis of YouTube videos cited in academic publications." *Journal of the American Society for Information Science and Technology*

Keenan, A. and Shiri, A. (2009) *Sociability and Social Interaction on Social Networking Websites.* Library Review, 58, 6, 438-450

Lange, P. G. (2007). “Publicly Private and Privately Public: Social Networking on YouTube” *Journal of Computer-Mediated Communication*, 13, 1, 361–380,

Lee, Y.-J., Shim, J.-M., Cho, H.-G. and Woo, G. (2010) *Detecting and Visualizing the Dispute Structure of the Replying Comments in the Internet Forum Sites.* International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery- Huangshan, China. 456-463

Mai, J.E. (2011),"The modernity of classification", *Journal of Documentation*, 67(4) pp.710-730

Madden, A. (2011). *The Community Aspect of YouTube: Characterising User Comments*. Unpublished Master’s dissertation. 2011.

Paolillo, J.C. (2008) *Structure and Network in the YouTube Core*. Proceedings of the 41st Annual International Conference on System Sciences – Waikoloa, Hawaii, USA. 156-165

Park, J., Fukuhara, T., Ohnukai, I., Takeda, H. and Lee, S. (2008) *Web Content Summarization Using Social Bookmarking Service*. Proceeding of the 10th ACM workshop on Web information and data management- Napa Valley, CA, USA. 103-110

Potthast, M. (2009) *Measuring the Descriptiveness of Web Comments*. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval - Boston, MA, USA. 724-725

Potthast, M. and Becker, S (2010) Opinion Summarization of Web Comments. *Proceedings of the 32nd European Conference on Information Retrieval, ECIR 2010- Milton Keynes, UK. 668-669*

Potthast, M., Stein, B. and Becker, S. (2010) *Towards Comment-based Cross-Media Retrieval.*19th International Conference on World Wide Web- Raleigh, NC, USA.

Rowe, M. and Angeletou, S. and Alani, H. (2011) “Anticipating Discussion Activity on Community Forums” *IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing.*

Siersdorfer, S. and Chelaru, S. and Nejdl, W and San Pedro, J. (2010) “How useful are your comments?: analyzing and predicting youtube comments and comment ratings,” in *Proceedings of 19th International WWW Conference*, Raleigh, North Carolina, April 26-30.

Silverman, D. (2005). *Doing Qualitative Research: A Practical Handbook* (Second Edition). London: SAGE.

Thelwall, M. Sud, P. and Vis, F. (2011). “Commenting on YouTube videos: from Guatemalan Rock to El Big Bang.” *Journal of the American Society for Information Science and Technology*. 6. 3. 616–629.