Instructional Videos and Others on Youtube
Similarities and Differences in Comments

Hugo Silva and Isabel Azevedo
Games, Interaction and Learning Technologies (GILT) Research Group, Instituto Superior de Engenharia do Porto, P. Porto, Porto, Portugal

Keywords: Informal Learning, Instructional Videos, YouTube.

Abstract: YouTube is a video sharing platform and its resources have been used for formal and informal learning. Users can add comments, as well as sign that they like a given video. The work described in this paper is mainly devoted to the comments provided by users and how they differ (or not) depending on the type of videos: those that are used to support learning versus those that does not. An application was developed to collect data available on the YouTube platform. The analysis of comments extracted from YouTube was performed using natural language processing techniques and differences in writing were also analyzed. Two major groups of videos were considered: technical, or instructional videos, and non-technical videos. The former usually have an educational nature and are watched by people that aim to improve their knowledge and skills, while the others are more devoted to entertainment. The similarities and differences found between these different types of videos are discussed.

1 INTRODUCTION

Formal and informal learning are not straightforward concepts. Different dimensions have been used to distinguish them. The physical space where learning occurs is one of them, where the main distinction is related to the use of in- or out-of-school learning environments (Ramey-Gassert, 1997), e.g. a classroom.

Marsick and Watkins compared formal and informal learning: “Formal learning is typically institutionally sponsored, classroom-based, and highly structured. Informal learning, a category that includes incidental learning, may occur in institutions, but it is not typically classroom-based or highly structured, and control of learning rests primarily in the hands of the learner. (…) Informal learning can be deliberately encouraged by an organization or it can take place despite an environment not highly conducive to learning.” (Marsick and Watkins, 1990).

Another often used characteristic is who mainly manages the experience, e.g. the learner or a teacher. Smith explores informal learning as an administrative notion (Smith, 2008).

Incidental learning, on the other hand, is defined as “a byproduct of some other activity, such as task accomplishment, interpersonal interaction, sensing the organizational culture, trial-and-error experimentation, or even formal learning” (Marsick and Watkins, 1990).

As many professionals need to continuously acquire new skills to keep up with their usual activities, the number of online resources, such as instructional videos, forums, technical blogs, FAQ websites and others, has been increasing.

YouTube videos are available to a great number of people, including self-learners. They have been used for formal and informal learning. Watching an instructional video demands an intention to learn about a given subject. Thus, incidental learning may occur secondarily, but for others topics.

YouTube resources have been utilized in academic contexts as part of the adopted pedagogical strategies. Tan and Pearce described their experience in using YouTube videos in a ten-week introductory sociology course (Tan and Pearce, 2011). The students’ insights on the use of YouTube were generally positive but the learners emphasised the importance of watching the videos in the classroom with teachers’ support and wide discussion.

Azevedo et al. reported about 7,600,000 results obtained with ‘tutorial’ as search term (Azevedo et

Silva, H. and Azevedo, I.
Instructional Videos and Others on Youtube - Similarities and Differences in Comments.
DOI: 10.5220/0006333104180425
In *Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017) - Volume 1*, pages 418-425
ISBN: 978-989-758-239-4

al., 2014). As of December 2016, there are about 169,000,000 results for the same search specification, including one entitled “Java Programming Tutorial - 1 - Installing the JDK” with more than 4,600,000 views. This video has more than 5,000 comments and one of them is “…What is Java actually used for, in terms of programming? ...”. It had some responses, including this one: “As far as I know: Desktop apps/games and android apps”.

Another comment on a video entitled “REST API concepts and examples”, starts this way: “Your content is absolutely excellent. I very much appreciate your videos and am learning a lot from them. I just wanted to give you some tips, as a web developer myself, but someone who spent 5 years as a teacher. Your videos are very well structured, however I feel that you, at some points, are speaking too fast. …”.

Despite the large number of papers about YouTube platform, little is known about the comments provided to educational videos freely available on YouTube. The main research question of this work is: What kind of information do these videos’ comments give when compared to others, and how can this information be retrieved?

In this research, we analyzed the following characteristics: the use of emoticons, linguistic characteristics, number of response to comments, among others.

The rest of this paper is organized as follows. This introduction provides detailed information about the context and the research question of this work. The second section mainly examines the YouTube platform as an online social network and the allowed interactions. The next section discusses how data was extracted, characterizes the data and provides a detailed data analysis. The penultimate section summarises the results, discusses the most relevant contributions and includes some final remarks. Finally, the last section states future research in the field, mainly based on the limitations of this work.

2 YouTube

Online social networks are among the most visited websites. An online social network allows a direct interaction between users. Facebook, Twitter, Google Plus and YouTube are some examples of social networks. In most of these social networks, the users establish connections between them in a user-user interaction (Wattenhofer et al., 2012). In contrast, social networks like Facebook and Google Plus provide a different experience, with users sharing personal thoughts, URLs, videos and creating connections with other users (friendship).

The interaction between users is perhaps the utmost of social networks, since users make appreciations about contents published or shared by other users. Alternatively, users can express themselves also by commenting publications of other users.

YouTube is a somewhat different social network, in which the main goal is the video sharing. Created in 2005, YouTube raises about 137 million users per month (Weaver et al., 2012). YouTube platform is also accessible on mobile devices through an optimized website or apps. In YouTube, a user can create a channel, upload videos, comment and express his opinion about a video using the appreciation options (like or dislike). YouTube videos are embedded in other websites and are frequently shared in other digital social networks.

The usability and functionality of YouTube, which allows users to easily create a channel and post content that is shared almost instantaneously in the internet, turns YouTube into an attractive platform to content creators and media companies (Susarla et al., 2012). The content of the videos is key for the success of YouTube, as many videos become viral when shared in social networks (Weaver et al., 2012). A “most-viewed” category is another indicator of the popularity of YouTube videos.

In this social network, social interaction is different from others (e.g. Facebook). On YouTube, social interaction is more often a user-content-user interaction (Wattenhofer et al., 2012) and thus users do not establish a bidirectional connection with other users unlike Facebook. On YouTube, connections between users are based on channel subscriptions, which were created by other users. The most expected interactions are video appreciations (like or dislike) and commenting videos.

3 CASE STUDY

YouTube is a web platform where users can demonstrate a positive or negative opinion by clicking the like or dislike buttons, respectively, or can also express their opinion by writing a comment on the video.

For example, the popular video Psy - Gangnam Style (OFFICIALPSY, 2012) reached more than two billion of views and more than four million comments. In January 2016, this video had more than 13 million appreciations (88% likes and 12% dislikes). However, some information remains unknown:

- Does this happen in other types of videos?
- How do users comment videos?

With these questions in mind, our work aimed at study and compare two different types of videos and their users' behaviour. Two different categories were defined: technical videos and non-technical videos. The technical videos selected for this analysis were tutorials and learning video-classes related with programming contents. The non-technical videos consisted in movie trailers.

This work was divided in three main steps: data collection, pre-processing and analysis with natural language processing techniques.

3.1 Data Collection

An application to extract data from YouTube videos was developed (see Figure 1).

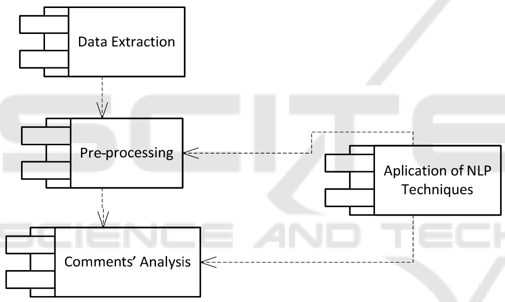

Figure 1: Solution components.

Data extraction represents the data retrieved from YouTube to be analyzed. This component is responsible for communicating with YouTube API and extract the videos and comments. Extracted comments are stored in a database.

The YouTube API (version 3, REST) was used to search for the videos according to the thematic defined and to collect the data. As this work was focused on analysing texts in English, it was intended to retrieve content and comments mainly in English. To achieve this, the parameter *relevanceLanguage* provided by the API was used with the value *en*.

For each type of video, 150 videos and respective comments were collected. The search terms were "movie trailers" for non-technical videos and "javascript", "java programming" and "php web" for technical videos. The methods used were *Search.list* for searching contents, *Videos.list* to get video info, and *CommentThreads.list* to get comments' info. The API returned the data in JSON format.

For each video, the following information was kept: title, description, URL, video identification, number of likes, number of dislikes, number of visualizations, and channel identification.

3.2 Sample Characterization

In Table 1 and Table 2, the number of videos and the respective YouTube category for technical and non-technical collected videos is presented.

Table 1: Technical videos per category.

YouTube Category	Number of videos
Howto & Style	60
Education	45
Science & Technology	36
People & Blogs	7
Movies	1
Classics	1

Table 2: Non-technical videos per category.

YouTube Category	Number of videos
Film & Animation	87
Entertainment	50
People & Blogs	4
Comedy	3
Trailers	3
Shows	2
Music	1

For each video, the first 5000 comments, if available, were collected in descending order by date, as it has been used before (Cambria and White, 2014) and (Baccianella et al., 2010, p. 0). The data collected for each comment was: comment text, identification, date and time, author's name, channel identification and, if the comment is an answer to a previous comment, the comment parent identification.

Table 3 provides data about the number of comments and their answers. Notice that proportionally there are more comments provided as response under the technical videos category (17% of response-comments for technical videos and 14% for non-technical videos), a point that is examined in section 3.5.

Table 3: Number of comments per type of video.

	Technical videos	Non-technical videos
Number of comments	73,901	657,014
Number of main comments	61,486	565,370
Number of comments with responses	6,430	37,912
Number of comments provided as responses	12,415	91,644

Technical videos have an average of 91,4 characters per comment, while non-technical videos tend to have longer comments. An average of 116,6 characters per comment was measured in our sample.

Regarding the number of users represented in the comments, we have 41,430 people for instructional videos and 466,417 for the others.

According to the number of videos commented by each user (see Table 4 and Table 5), users seem to be prone to watch more than one instructional video, perhaps aiming to improve learning.

Table 4: Percentage of users and number of videos commented.

Number of commented videos	Percentage of users
1	80,38
2	12.05
3	3.65
4	1.56
5 or more	2.36

Table 5: Percentage of users and number of videos commented per category.

Number of commented videos	Percentage of users (technical videos)	Percentage of users (non-technical videos)
1	71.50	81.30
2	15.9	11.66
3	5.63	3.44
4	2.56	1.45
5 or more	4.41	2.15

On technical videos, 18.26% of comments have (single) question marks, as usual in questions, a number that decreases to 13.39% for non-technical video comments. It was also verified that 90% of the technical video comments with question mark (?) do not have this punctuation signal repeated. In non-technical videos, this percentage is 79.4%.

3.3 Pre-Processing

To be able to perform an analysis on the collected comments, it was necessary to prepare the comments considering the type of writing on YouTube. Writing comments in platforms like YouTube can be seen as computer mediate communication (CMC) (Hogenboom et al., 2015). It is very common in YouTube comments the use of abbreviations, repetition of characters and signal punctuation and the use of emoticons.

The use of repeated characters (chars) is often a way to reinforce an idea. For example, the use of the word "*looooove*" has clearly the intention of emphasizing the positive sentiment. However, that word does not exist in any dictionary or lexicon. The removal of repeated characters in this work uses the same approach used in (Lehnert and Ringle, 2014). Words with at least three equal consecutive characters were considered to have repeated characters. For those words, the repeated chars were removed by using the following approach: the tool removed all but two consecutive characters; then the new word was checked in a dictionary; the word was considered if the dictionary recognized it; otherwise, the last repeated char was removed. For example:

- gooooood → good (exists)
- loooooove → loove (does not exist) → love (exist)

Two percent of the comments for technical videos had repeated chars that were removed. In non-technical videos, the number of comments with repeated chars removed was four percent.

Repeated punctuation was also removed to achieve a better performance on analysis as described in (Liu, 2012). The punctuation signals identified were the full-stop (.), the question-mark (?) and the exclamation point (!). Two or more consecutive occurrences of a punctuation signal were considered as repeated punctuation (except for reticence - three consecutive full-stops). Repeated punctuation was removed in 13% of the comments of technical videos and 18% of the comments of non-technical videos. The most frequent repeated punctuation signals are identified in the Table 6.

The most common emoticons presented in technical and non-technical videos are presented in Table 7 and Table 8.

Notice that there are 7,746 comments for technical resources with at least one of the five emoticons most used for technical videos as listed in Table 6 and 17,099 for non-technical resources with the top-5 emoticons listed in Table 8. Proportionally, it seems that the first group tends to have more positive emoticons.

Table 6: Repeated punctuation resume.

	Comments of technical videos	Comments of non-technical videos
Full-stop (.)	53.99%	44.59%
Exclamation mark (!)	30.40%	36.52%
Question mark (?)	13.52%	17.43%
Others (, or ;)	2.09%	1.46%

Table 7: Top 10 of emoticons on technical videos.

	Emoticon	# comments
1	:)	4,737
2	:D	1,481
3	:P	633
4	;)	476
5	xD	419
6	:(	414
7	XD	255
8	:/	166
9	:-)	136
10	:p	127

Table 8: Top 10 of emoticons on non-technical videos.

	Emoticon	# comments
1	:)	6,822
2	:D	4,760
3	<3	1,969
4	XD	1,814
5	xD	1,734
6	;)	1,640
7	:(	1,590
8	:P	1,390
9	:/	818
10	:3	497

By the analysis of these tables, it can be concluded that the top 2 emoticons are the same. However, the third most used emoticon on non-technical video comments (Table 8) is <3, used to represent a heart, which expresses an intense sentiment. This emoticon is not often seen in comments for technical contents.

3.4 Linguistic Analysis

To determine the frequency of the terms used in the comments, word clouds were generated. For technical videos, three types of clouds were produced: i) the most used words (Figure 2), ii) the most used adjectives (Figure 3), and iii) the most used verbs (Figure 4).

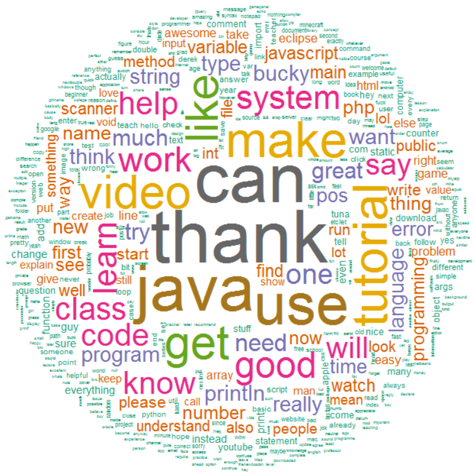

Figure 2: Words' frequency in technical video comments.

Figure 3: Adjectives' frequency in technical video comments.

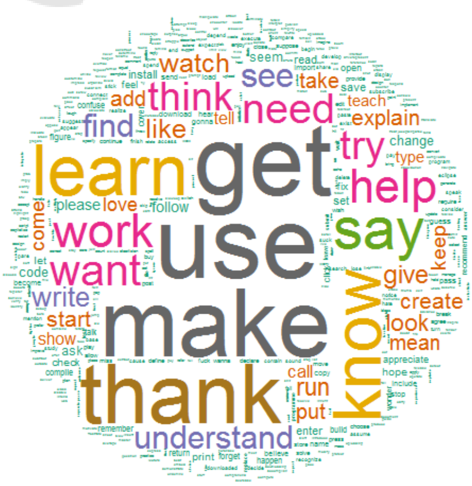

Figure 4: Verbs' frequency in technical video comments.

For non-technical videos, the same three types of clouds were produced. For the most used words (Figure 5), the word “movie” stands out. Figure 6 and Figure 7 present the most used adjectives and verbs, respectively. Despite not being highly used, the words “wait” and “see” appear in Figure 5 and Figure 7, which normally results from sentences as “can’t wait to see”.

Figure 5: Words’ frequency in non-technical video comments.

Figure 6: Adjectives’ frequency in non-technical video comments.

Figure 7: Verbs’ frequency in non-technical video comments.

3.5 Comments and Response-Comments

YouTube allows users to reply other users’ comments. In the collected data, 14% of response-comments were identified for non-technical resources and 17% for technical videos (see Table 4). The higher percentage of answers in technical videos might be an indicator of the behaviour of users in this type of videos. Videos with learning content are more prone to have comments with questions or doubts related to the video content, and in turn, to have more answers, replies and thankfulness.

Likewise, for technical videos an average of 1.9 responses was observed for comments with replies, while for non-technical videos this value is 2.4.

With the method *CommentThreads.list* of the API, it was possible to retrieve the number of positive appreciations (likes) of the comments. The percentage of comments with positive appreciations was 22.8% for technical videos and 22.3% for non-technical videos, which corresponds to 7.3 and 10.7 likes per comment, respectively.

Furthermore, as explained before, our analysis detected that technical videos have an average of 91.4 characters per comment and non-technical videos have an average of 116.6 characters/comment. In which regards to correction for repeated sequential characters, that was necessary for 4% of non-technical and 2% of technical video comments. However, the number of response-comments that needed this type of correction was close to 0. As the use of repeated characters tends to reinforce an idea, this value reveals that when users reply to comments, they tend to be more objective in their message. The same tendency was observed for repeated punctuation marks: 18% of repeated punctuation marks were found in non-technical video comments against 13% in technical video comments.

3.6 What Users Talk about

The comments collected have different characteristics in which regards to the type of video they correspond to. For technical videos, the most common technical terms related to the extracted videos were searched, and are related to programming languages. A set of technical terms related to programming languages was defined and the most frequent terms on comments were disclosed. Figure 8 shows the ten technical terms most frequent in technical video comments and their frequency.

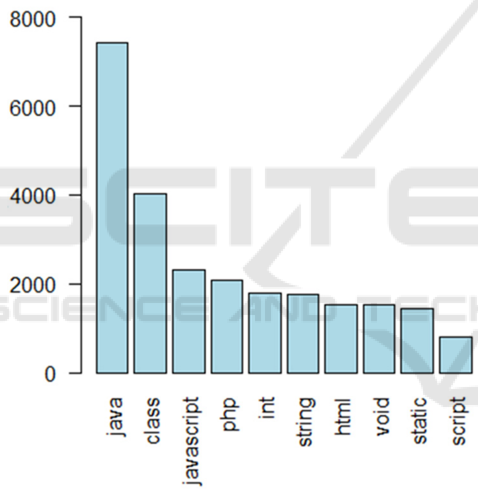

Figure 8: Frequency of technical terms.

The common use of terms like "class", "int", "string", "static" and "void" in comments suggest that users write code in comments to ask or answer questions.

We also investigated which topics related to the movies presented in the trailer videos are more frequent in the comments extracted by identifying the code of the respective IMDB (Internet Movie Database) movie. With this ID, and with the Open Movie Database (OMDB) API (Fritz, 2016), we extracted information about leading actors, director, writers and the movie title. Then these comments were analyzed to know how many comments mention these topics and which ones are more mentioned. Only 4.06% of the comments refer to topics collected from IMDB. Figure 9 shows the frequency of references to actors, directors, scriptwriters, and the title.

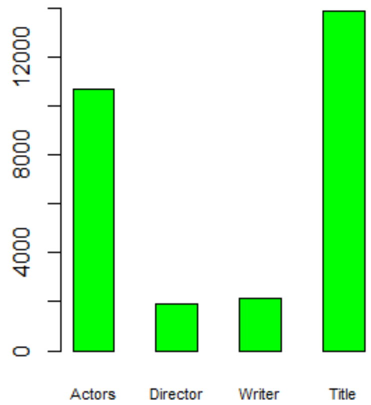

Figure 9: Frequency of IMDB topics.

The movie title and the name of the actors were the most referenced topics in these comments (a total of 13,861 references to the title; 10,667 references to the names of the actors; 2,157 references to scriptwriters; and 1,916 references to filmmakers).

4 CONCLUSIONS

The findings of this study indicate that there is a sense of gratitude between viewers of instructional videos. Not only the word cloud includes the word “thank” but also users tend to express their appreciation using some emoticons such as “:-)”. Considering the top-5 emoticons used for the two categories of videos, 10,48% of comments for instructional videos with positive icons and 2,60% of comments with these positive symbols for the other videos.

Others verbs that are common for educational contents are “use”, “try”, “help”, “learn”. In addition to the high level of technical terms used in comments for technical videos, these data reinforce the idea that users debate contents and clarify doubts.

Users commenting just one video are less frequent for technical videos, which can express a strong interest in some topics that they aim to learn and, thus, users watch and comment more than one instructional video.

We also found that comments for instructional videos include more questions, which may be derived from the fact that users try to receive guidelines or some assistance, which are also in line with some words very common in comments for these videos, such as “help”, as seen before.

With this study, it was possible to verify that the comments of educational videos on YouTube are also a learning tool that complements the tutorship provided by the video. For those who seek knowledge on YouTube videos, the comments section should also be a source of educational resources. Also, educators who produce content for YouTube should be aware of the comments section.

5 LIMITATIONS AND FUTURE WORK

This research has some limitations that open future opportunities of research. First, considering that the kind of videos selected may have intrinsic and unknown characteristics that might have conditioned this study, the study must be replicated using other videos and comments.

The only criterion used to select the videos obtained with the search terms was the number of views. We did not examine the impact of video duration or their academic or not provenience.

Another factor that can have an impact in the results is the publication date or even some characteristics of their authors, for instance, their native language.

More experimentation and empirical research may lead to a better understanding on how people comment on instructional videos or even how these resources are used. This is important to improve instructional videos and enhance users’ experiences. However, issues related to privacy and ethics must always be considered when dealing with users’ observations even when they are publicly available.

REFERENCES

Azevedo, I., Carrapatoso, E., Carvalho, C., 2014. Supporting learning through tagging systems, in: Jovanovic, J., Chiong, R. (Eds.), Technological and Social Environments for Interactive Learning. Informing Science Press, Santa Rosa, CA, USA.

Baccianella, S., Esuli, A., Sebastiani, F., 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining, in: Proceedings of the Seventh Conference on International Language Resources and Evaluation. pp. 2200–2204.

Cambria, E., White, B., 2014. Jumping NLP curves: a review of natural language processing research. IEEE Computational Intelligence Magazine 9, 48–57.

Fritz, B., 2016. OMDb API: The Open Movie Database [WWW Document]. URL http://www.omdbapi.com.

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., De Jong, F., Kaymak, U., 2015. Exploiting Emoticons in Polarity Classification of Text. Journal of Web Engineering 14, 22–40.

Lehnert, W.G., Ringle, M.H. (Eds.), 2014. Strategies for natural language processing. Psychology Press.

Liu, B., 2012. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies 5, 1–167.

Marsick, V.J., Watkins, K., 1990. Informal and Incidental Learning in the Workplace. Routledge, London and New York.

OFFICIALPSY, 2012. Psy - Gangnam Style [WWW Document]. URL https://www.youtube.com/watch?v=9bZkp7q19f0.

Ramey-Gassert, L., 1997. Learning science beyond the classroom. The Elementary School Journal 433–450.

Smith, M.K., 2008. Informal learning [WWW Document]. The encyclopaedia of informal education. URL http://infed.org/mobi/informal-learning-theory-practice-and-experience.

Susarla, A., Oh, J.-H., Tan, Y., 2012. Social networks and the diffusion of user-generated content: Evidence from YouTube. Information Systems Research 23, 23–41.

Tan, E., Pearce, N., 2011. Open education videos in the classroom: exploring the opportunities and barriers to the use of YouTube in teaching introductory sociology, in: Proceedings of the 18th International Conference of the Association for Learning Technology. University of Leeds, UK.

Wattenhofer, M., Wattenhofer, R., Zhu, Z., 2012. The YouTube Social Network, in: Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM 2012).

Weaver, A.J., Zelenkauskaite, A., Samson, L., 2012. The (non) violent world of YouTube: Content trends in web video. Journal of Communication 62, 1065–1083.