Acta Univ. Sapientiae Informatica **14,** 2 (2022) 273–301

DOI: 10.2478/ausi-2022-0016

Investigating the AlgoRythmics YouTube channel: the Comment Term Frequency Comparison social media analytics method

Pálma Rozália OSZTIÁN
Sapientia Hungarian University of Transylvania
Târgu Mureș, Romania,
Department of Mathematics-Informatics
email: osztian.palma@ms.sapientia.ro
ORCID:0000-0001-6917-8307

Zoltán KÁTAI
Sapientia Hungarian University of Transylvania
Târgu Mureș, Romania,
Department of Mathematics-Informatics
email: katai_zoltan@ms.sapientia.ro
ORCID:0000-0003-2343-3629

Ágnes SÁNTHA
Sapientia Hungarian University of Transylvania
Târgu Mureș, Romania,
Department of Applied Social Sciences
email: santhaagnes@ms.sapientia.ro
ORCID:0000-0002-5065-557X

Erika OSZTIÁN
Sapientia Hungarian University of Transylvania
Târgu Mureș, Romania,
Department of Mathematics-Informatics
email: osztian@ms.sapientia.ro
ORCID:0000-0002-6019-5677

Abstract. In this paper we investigate the comments from the AlgoRythmics YouTube channel using the Comment Term Frequency Comparison social media analytics method. Comment Term Frequency Comparison can be a useful tool to understand how a social media platform, such as a Youtube channel is being discussed by users and to identify opportunities

Computing Classification System 1998: K.3
Mathematics Subject Classification 2010: 68W40
Key words and phrases: social media, YouTube, comments, time series graph, subtopic word frequency analysis, gender differences analysis, sentiment analysis

to engage with the audience. Understanding viewer opinions and reactions to a video, identifying trends and patterns in the way people are discussing a particular topic, and measuring the effectiveness of a video in achieving its intended goals is one of the most important points of view for a channel to develop. Youtube comment analytics can be a valuable tool looking to understand how the AlgoRythmics channel videos are being received by viewers and to identify opportunities for improvement. Our study focuses on the importance of user feedback based on ten algorithm visualization videos from the AlgoRythmics channel. In order to find evidence how our channel works and new ideas to improve we used the so-called comment term frequency comparison social media analytics method to investigate the main characteristics of user feedback. We analyzed the comments using both Youtube Studio Analytics and Mozdeh Big Data Analysis tool.

1 Introduction

In the past 10 years, the AlgoRythmics YouTube channel[1] has become a remarkable informatics teaching tool all over the world. The channel has gathered 7,714,576 visualizations, 93,482 likes and 5,314 comments since the first videos had been posted. Its success undoubtedly lies within the fact that these videos visualize basic informatics algorithms in unique ways. Seven sorting algorithms are illustrated by folk dance performances, three searching algorithms by flamenco and classical ballet choreographies. Lately, the project was moved to the theater stage. As such, the latest video displays an escape room scene sequence in order to address Greedy and Dynamic programming algorithm strategies. AlgoRythmics videos and the web-based learning environment based on these have provided the background for a number of high standard research on education sciences. The aim of this paper is the study of the channel based on the Comment Term Frequency Comparison social media analytics method, with the purpose of exploring and identifying further research agendas. The method has recently been published by Thelwall [20] and used for the thematic analysis of YouTube video comments.

2 Previous work

In this section we succinctly present the history of AlgoRythmics project, channel, environment, and the related previous research which is based on the

[1] https://www.youtube.com/@AlgoRythmics

book *Algorythmics: technologically and artistically enhanced computer science education* by Kátai Zoltán [12].

2.1 AlgoRythmics: an award-winning project

A 2013 report by the joint Informatics Europe & ACM Europe Working Group on Informatics Education [21], maintains that for nations to succeed in the pursuit of technological innovation, the basics of informatics – the science behind Information Technology (IT) – must be fundamentally and generally understood by all. Today's highly competitive job market dictates that students display a firm grasp of the key concepts that underlie informatics. Furthermore, computational thinking (CT) is an ability that should be developed by all, and the skills that form the basis of informatics are not only teachable but should form an integral part of the primary curriculum and especially that of secondary schools.

The Informatics Europe "2013 Best Practices in Education Award" – set up to promote Informatics Education in Primary and Secondary Schools – gave that year's award jointly to Sapientia Hungarian University of Transylvania, Romania, and the Warsaw School of Computer Science, Poland. The evaluation committee praised the originality of the Multi-Sensory Informatics Education proposal (by Zoltán Kátai, László Tóth and Alpár Károly Adorjáni). They also commended their blend of algorithm-learning with sensory experience as an innovative teaching experiment, and its inter-cultural character. In the following, we briefly detail the preliminaries of this prestigious award and the achievements of the research group in the next period.

2.2 The AlgoRythmics research group, YouTube channel and learning environment

It was during the 2003–2007 period at Sapientia Hungarian University of Transylvania that the AlgoRythmics project was initiated. One area that the author, Zoltán Kátai, addressed was the multi-sensory approach of CS education, focusing initially on investigation of the part played by the senses in education. The study included undergraduate student Alpár Károly Adorjáni, who developed a software tool which allowed the mutual inclusion of vision, hearing, touch and subsequently the kinaesthetic sense within the teaching and learning process of computer algorithms. The later inclusion of Tóth László saw the introduction of dance in the programme following which, the research group began a collaboration with a professional folk-dance group (Maros Artistic Ensemble). This collaboration resulted in the creation of six folk dance choreographies illustrating sorting algorithms. The videos were posted on the AlgoRythmics YouTube channel in 2011 (Kátai & Tóth, 2011). An AlgoRythmics web application followed, allowing the association of interactive computer animations with algorithmic dance performances. With the later addition of Erika Osztián and Géza Károly Vekov, the project was given a new impetus and the AlgoRythmics collection grew to include four new dance choreographies (Kátai, Osztián, Osztián, & Vekov, 2018). Further collaborations broadened the repertoire with new algorithms and dance styles to include flamenco (The András Lóránt Company) and ballet (Cluj-Napoca Hungarian State Opera). Assisted by three undergraduates, the AlgoRythmics web application was eventually re-designed (Kátai, Osztián, Osztián, Nagy, & Cosma, 2020). Recently, a new colleague, Sántha Ágnes, joined the research group, who represents a new point of view, as she approaches the addressed research questions from the side of social sciences.

The six sorting algorithm visualizations posted on YouTube in 2011 were: selection-sort performed with a local Gypsy folk dance, insertion-sort performed with a Romanian folk dance, bubble-sort with a Hungarian (Csángó) folk dance, shell-sort with a further Hungarian (Székely) folk dance, merge-sort with a Transylvanian-saxon (German) folk dance, and last, quick-sort with a Hungarian (Küküllőmenti legényes) folk dance. The algorithm-dance associations were established based on the common defining characteristics. Analysis revealed that all these choreographies entailed some binding elements: two elements' comparison (all algorithms); two elements' swapping (all algorithms); current sequence division into two sub-sequences (quick-sort and merge-sort); current sequence partitioning in two sub-sequences (quick-sort); merging of two neighbor sub-sequences (merge-sort). Consultation with the choreographers ensured that the appropriate dance-steps were matched to these key-operations of the algorithms. The seventh video posted in 2018 was Heap-sort with Hungarian (Mezőségi) folk dance. The last three visualizations focused on searching algorithms. For the linear and binary search algorithms it was decided to choose a more international type of music and choreography by adopting the style of the well-known southern Spanish flamenco. In addition, a classic ballet choreography was chosen to illustrate the recursive version of the classic backtracking algorithm's four-queen variant to solve the famous "eight queen puzzle".

The meta-study of algorithm visualization effectiveness presented by Hundhausen, Douglas, and Stasko [3] concludes that algorithm visualizations support effective learning when students are actively involved rather than passively viewing a visualization. Such involvement offers learners control over the algorithm animation process [15]. Algorithm orchestration can be regarded as a special version of the so-called interactive prediction method [2], students have the skills for predicting and even for performing the entire step sequence of the algorithm using an interactive learning environment. AlgoRythmics learning environments were developed based on this principle of genuine active involvement. To each dance choreography an interactive abstract computer animation was attached. The animation module of the first version of the web application worked in two modes. In the first, termed as white-box task, users need to predict a compare/swap operation sequence associated with the sorting algorithm for a random number sequence stored in a "white-array" (with visible numbers). The second mode, the Black-box task, requires that the user perform the same task, but this time on a random number sequence stored in a "black-array" (with hidden numbers). Recently, the environment was redesigned. The current version of the web application includes a complex module which guides students through five interactive learning steps "from dance to code". The environment takes the form of a "dance floor", where the "choreographer" is allowed to identify the level of user interaction. The "dancers" pick up the "rhythm of the displayed algorithm" and have the possibility to control it themselves. The code being built by the user, is then executed together with the animation.

2.3 AlgoRythmics research

Below we present a brief summary of our ten most important researches.

2.3.1 Computer science education using multiple senses

Our earlier research was focused on the principles of multi-sensory learning in order to support CS education. Prior research in this field concluded that: (i) the brain is organized in such way that it can elaborate information from different sensory channels in a cooperative fashion; (ii) visual, auditory, and reasoning processes show an interconnection; (iii) there are cognitive benefits to multisensory structured methods; (iv) more senses can give rise to a more efficient teaching-learning process. These findings are also supported by Stevens and Goldberg [18] who state that the core principles of brain-based learning include: our brains desire for multi-sensory input; learning involves the whole body. Other research calls attention to the fact that our senses reach our feelings, emotions and aesthetic sense, as well as our intellect, and that even multimedia digital content disregards “the extraordinary capacity of our brain to capture and process information from [all of] our senses”. Seeing, hearing and touching computer algorithms Computer algorithms are abstract series of operations, and a variety of instruments are used to make them distinguishable to learners, most commonly, visual representations. A further challenging step involves making algorithms perceptible for the auditory sensory. Connecting tactile senses to the learning environment can be even more challenging. This was the focus of our first research [10] that investigated the effectiveness of the multisensory method and tool we developed to support the teaching-learning process of elementary algorithms. Results confirmed our expectation that the proposed multi-sensory method can support students in analyzing and designing basic algorithms. Playing recursive scenarios the second research we performed focused on the process of teaching-learning in the recursive algorithm [6]. We have proposed that our multisensory method be extended with didactical role playing. Students were assisted in understanding how recursion works by being invited to play the running process of certain recursive functions and procedures. Enacting such recursive scenarios helps students imagine how recursion works. Unlike the spectator-students who get an overview of the whole process, the participating actor-students take advantage of kinaesthetic memory to learn specific moves associated with each recursive-scenario. Again, reported findings suggest that role playing could be an effective method to support students in in analyzing, designing, and implementing recursive procedures and functions. Dancing sorting algorithms our third research [7] analyzed how dance could be assimilated into the teaching and learning of sorting algorithms. First, students with an affinity for dancing were invited to collaborate in our project. They each assumed the roles of numbers from the sorting sequence, and wore the corresponding number on their costume. Later, to increase the artistic value of the choreographies, amateur dancers took the place of the students. Research results revealed that video recordings of such algorithmic dance performances have the potential to support students in assimilating basic sorting algorithms.

2.3.2 AlgoRythmics: science and art without ethnic borders

After the above described professional folk dance choreography representations were completed, in the framework of a fourth research, we directed our attention to the field of intercultural education. In an influential report of the Committee on Culture, Science and Education [22], the Council of Europe Parliamentary Assembly states that education shall promote understanding, tolerance and friendship among nations and ethnic groups, and all forms of artistic expression are tools in intercultural education. Accordingly, we have proposed to develop an online e-learning environment that can equally promote both intercultural education and CS education. The folk dance choreographies described above illustrate both basic CS concepts and cultural diversity within Transylvania (Romania). This combination of art and science also exemplifies how the concept of 'unity-in-diversity' can be employed in a science educational context. Interestingly, the study we performed [4] revealed possible complications CS teachers might be presented with when delivering scientific content in a culturally diverse context. Research results have shown that a student's cultural related concept concept and it may occur that feelings influence the way they relate to the scientific content. This phenomenon was reflected in a number of YouTube comments too.

2.3.3 Teaching "not blind learners" to program "blind computers"

In comparison with humans, computers are blind in many ways. Any algorithm visualization system may display information that has additional meaning for learners. In this type of learning environment, learners may encounter difficulties following a precise computer algorithm. For instance, sorting algorithm visualizations can expose the number-sequence to be sorted and as learners see the numbers, they naturally discern if two elements are in the correct order or not, and therefore are prone to skip the explicit comparison operation represented by the computer algorithm. The fifth study we performed [5] investigates this phenomenon and suggests solutions to circumvent possible side-effects that arise when substituting "blind computers" with "non-blind humans" in algorithm visualization learning environments. Research results revealed latent deficiencies that algorithm visualization systems might have: displaying information that has additional meanings for human viewers can obstruct them in following strict computer algorithms. Findings also showed that hiding applied in a wise manner leads to more effective algorithm visualization thanks to its higher epistemic fidelity.

2.3.4 Promoting algorithmic/computational thinking of learners focused on sciences and humanities

Since algorithmic and computational thinking are important skills for today's information based society, our sixth research addressed the topic of promoting these skills for all. One of the greatest challenges of such an initiative is dealing effectively with diversity. We focused on a specific aspect relating to diversity that of sciences- vs. humanities-oriented students. Since motivation plays a critical role in learning and forms the basis of any approach to active learning [19], the importance of diversity for learner motivation is a critically important aspect of this educational topic. The analysis of the following question was proposed: Is it possible to construct unified learning environments capable of promoting algorithmic and computational thinking inclusively? The two related studies we performed [8] intended to (1) present research results revealing that calibrated learning environments can potentially promote algorithmic and computational thinking of both sciences-oriented and humanities-oriented students; (2) analyze the precise motivational challenges that instructional designers might be confronted with when developing learning environments aiming to bridge the diversity gap and which targets both learner communities.

One important inference of this research was that there are no unbridgeable differences between the ways humanities- and science-oriented students relate to algorithmic and computational thinking promoter e-learning tools. As such, correlations between both the performance results and motivational scores of the two learner groups were strong. Furthermore, the superiority of science-oriented students over their humanities-oriented colleagues progressively and significantly lessened as the groups progressed through their e-learning tasks. These studies ought to encourage curriculum developers and instructional designers to further examine the possibility of designing and developing unified algorithmic and computational thinking promoter learning environments for all students.

2.3.5 Schematic versus human movement enriched realistic algorithm visualization

Instructional dynamic visualizations are classified as either schematic or realistic representations [14]. Recent decades have seen some relevant studies, in different fields of education, that have focused on the relative value of these types of visualizations [17]. Algorithm visualization, on the other hand, has been scantily researched from this viewpoint. Possibly because computer algorithms are inherently abstract entities, lacking any evident exemplification in the real world therefore, schematic animations often serve to illustrate computer algorithms. There are, however, representations that are closer to real-world visualizations, one example being the AlgoRythmics environment in which the videos illustrate basic computer algorithms with realistic dance choreographies. A distinct strength of these videos is that they illustrate the basic operations associated with algorithms by referring them to human movement. The idea that observing human movements can be cognitively beneficial is borne out by recent research [1]. As multiple external representations have complementary roles, some research undertook the study of learning environments that combine realistic and schematic dynamic visualizations.

In accordance with the prior research presented above, the AlgoRythmics dance choreographies can be seen as human movement effect (HME) enhanced realistic representations and the attached computer animations as schematic representations [13]. To get an insight into the AlgoRythmics environment from this perspective as well, our seventh research investigated the relative effectiveness of the following learning settings:(1) schematic + schematic, (2) HME-realistic + HME-realistic, (3) schematic + HME-realistic and (4) HME-realistic + schematic. The algorithm we selected to be analyzed was the Shell sort strategy and the visualizations were presented successively for all four groups.

2.3.6 Algorithm visualization environments: is there an optimal interactivity level?

An essential goal of redesigning the AlgoRythmics environment was the increase in user engagement in the visualization process. The new animation module includes the interactive prediction feature on three levels: : "no-interactivity" (viewers are passive), "half-interactivity" (the animation stops at predefined key moments so that users have to indicate the next movement or operation) and "full-interactivity" (users are required to coordinate the algorithm by predicting the entire operation sequence). The eighth study [16] we performed focused on the influence that varying degrees of interactivity have upon students' learning. One of the most significant conclusions of this investigation is that a universally optimal interactivity level cannot be achieved. Also, our findings suggest that the three levels of engagement may yield advantages and disadvantages. If we take the precept that every student deserves the opportunity to learn, and that everybody is different, then the most effective learning style should match the case of each student. Therefore, we are certain that all applied learning environments should entail different interactivity levels. If this were to be achieved, all participants would be able to access the best learning style, thereby achieving better results.

2.3.7 Improving the teaching-learning environment of AlgoRythmics by asking questions

A significant common characteristic of the above studies is that the authors analyzed only self-paced learning sessions. Additionally, the examined learning settings supplemented the viewing of the dance choreographies with the interactive orchestration of the algorithms. In a ninth study [11] we analyzed such AlgoRythmics learning settings where the principle of active involvement was implemented by questioning (both with and without teacher guidance). Our experience with AlgoRythmics videos (as CS teachers) also provided inspiration in this sense, since we observed that without teacher support the potential of these choreographies to serve as computational thinking promoter tools is not fully utilized. More specifically, we have proposed to establish whether dance choreography visualizations (created for given input sequences) are sufficiently expressive to support students who lack prior knowledge in computing to imagine the best and worst case behavior of algorithms. The attached questions, beside supporting and guiding students' thinking process, enhanced focusing on the visualizations' domain-relevant aspects,too.

2.3.8 Investigating young school students' computational thinking ability across grade levels

While previous AlgoRythmics research mainly focused on assessment of the computational thinking level of certain age groups, our tenth study [9] was prompted by the question of whether there is any detectable progress in students' computational thinking during their K-9 education? The exact research question we addressed was: are there detectable differences in how 3rd, 5th, 7th and 9th grade students from Arts and Theoretical schools (who have little or no explicit prior experience with computational thinking) relate to learning tasks in which a certain level of computational thinking is assumed? More detailed: (i) What is the pace of the potential computational thinking growth? (ii) How dependent is the rate and pace of computational thinking growth on the nature of the current curriculum (arts vs. theoretical schools)? (iii) How well can students of different grade levels assimilate a basic computer algorithm (linear search)? (iv) What signs are there of advanced computational thinking at different grade levels? (v) Can the rate and pace of computational thinking growth be said to depend on the gender of the learners?

3 The Comment Term Frequency Comparison (CTFC) social media analytics method

During the past decade mining social media for opinions expressed in public have become a standard practice in several fields like industry, academia, etc. Recently, the so-called Comment Term Frequency Comparison (CTFC) social media analytics method has been proposed to investigate YouTube comments in a specific topic [20]. In the following, we present this method and apply it to the analysis of the AlgoRythmics channel. The process starts with the "data gathering and filtering" (1) step including the following phases: Topic definition and delineation; Initial subtopic query set generation; Query testing and refinement; Video list generation; Video list checking; Comment downloading; Duplicate commenter removal; Comment pre-processing; Language filtering. As a next step a "time series graph" (2) is created to identify typical dates of comments as background information. The 3rd step implies a "subtopic word frequency analysis". During this phase of the investigation issues that are characteristic of each subtopic are identified. Next a "gender differences analysis" (4) is performed that seeks gender specific subjects. Step 5 is a "sentiment analysis" based on a word frequency comparison of comments with strong positive or negative sentiment. In the last step an overview network is generated (6) where vertices represent the subtopics and edges characterize the similarities between subtopics (measured in similar words in the corresponding comments). The suggested method (Thelwall, 2018) also includes two supplementary steps: pilot testing (step 0) and insight verification (step 7). The role of the pilot testing is to identify likely sources of off-topic information. The 7th step implies reading the comments containing the relevant terms. The role of this verification phase is to avoid "apparent insight" caused by spam or off-topic comments.

4 Method

YouTube comments can be analyzed in a number of ways to understand viewer engagement and sentiment. In order to investigate users' feedback and opinions we used two types of visualization tools: the so-called built in YouTube Studio analytics and the Mozdeh Big Data Analysis tool suggested by the authors of the CTFC method. We investigated the AlgoRythmics channel characteristics following the proposed 6 steps. Step 0 has no relevance in the case of our study. We used step 7 to strengthen our results from the previous steps.

4.1 Procedure and materials

In this study we analyzed the AlgoRythmics YouTube Channel comments in the case of ten algorithm visualization videos: 7 sorting algorithms and 3 searching algorithms. The details (algorithm type, video url, publish date) of these visualizations can be seen in Table 1.

Algorithm type	URL	Publish date
Bubble-sort	lyZQPjUT5B4	March 29, 2011
Insertion-sort	Ns4TPTC8whw	March 29, 2011
Selection-sort	Ns4TPTC8whw	April 03, 2011
Merge-sort	XaqR3G_NVoo	April 24, 2011
Quick-sort	ywWBy6J5gz8	May 03, 2011
Shell-sort	CmPA7zE8mx0	April 03, 2011
Heap-sort	Xw2D9aJRBY4	January 30, 2018
Linear-search	-PuqKbu9K3U	January 30, 2018
Binary-search	iP897Z5Nerk	January 30, 2018
N Queens problem	R8bM6pxlrLY	January 30, 2018

Table 1: YouTube videos

We chose English as the default language to analyze comments, therefore our conclusions were formulated according to this. The detailed view of our study and method procedure can be seen on the following diagram (Figure 1).

Taking in consideration that YouTube Studio provides many opportunities to analyze user specific feedback, we found this tool very effective to visualize channel analytic. This was used in the first two steps of our method: data gathering (step 1) and the time series graph (step 2). Within YouTube Studio, there are several options available for analyzing a channel's performance, including traffic source, audience data, subscribers, videos and comments on videos grouped by several characteristics as well. Therefore, we used the AlgoRythmics Channel Analytics dashboard in order to analyze and visualize all YouTube comments. It is important to mention that despite the fact that the YouTube Studio Analytics platform can generate many interesting and meaningful diagrams, these are analyzed only in a superficial way considering the following categories: number of comments, algorithm types, specific date of publishing the comment. In order to provide concrete information based on subtopic word frequency analysis (step 3), gender differences (step 4), sentiment analysis (step 5), network analysis (step 6) and verification (step 7)

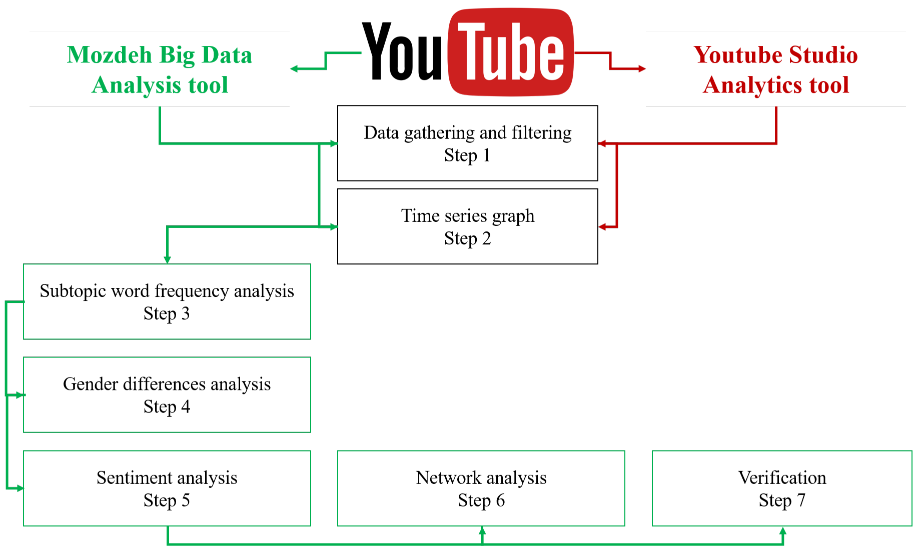

Figure 1: Method steps – workflow

as well we introduced the other effective tool, the Mozdeh Big Data Analysis which can complement the previously mentioned YouTube Studio tool regarding comment based analytics. Mozdeh is a free Windows Software which is able to collect and analyze comments on YouTube and Twitter. Based on Terlumun et al. Mozdeh carries out analysis on sentiment to detect and codify subjective content and estimate the strength of positive and negative sentiment in the texts. Since the key concept of our research topic is the AlgoRythmics channel in which case we planned to analyze comments belonging to 10 algorithm visualization videos, we adjusted the CTFC method to this context in order to investigate user engagement and sentiment.

4.2 Results and discussion

In the interest of drawing conclusions and edifications we were interested in the overall picture of user feedback including all seven steps shown in Figure 1.

4.2.1 Data gathering and filtering

As a first step we collected data from the AlgoRythmics YouTube channel using both tools. In the case of YouTube Studio Analytics dashboard there is a built-in technique which shows the number of all comments (5,314 comments in total) published to the channel. Due to the fact that we only analyzed 10 specific algorithm visualization videos, we added a filter criteria to define the number of comments belonging to these algorithms. This resulted in 5,234 comments which included texts in any language. Hereinafter, all YouTube Studio specific results are built on these 5,234 comments.

On the other hand, using Mozdeh a total of 3,136 comments were collected from YouTube since we used English as the main language. Despite the fact that Mozdeh already filters YouTube comments based on the main language it can often happen that there are many mixed posts which include more than one language. This software is described as working poorly if there are multiple languages in the data. In order to avoid this problem Mozdeh gives the opportunity to filter by language which can exclude all texts not containing any terms that are common in the selected language (in our case English) and rare in others. This filter can be added by using a query with the following predefined common English words: *had he his it that the to was with you are my their this thank for she who why where when how love hate awesome great amazing more*. Using this filter criteria we came to 2,019 matching comments and hiding duplicates resulted in 1,922 comments in total. Hereinafter, all Mozdeh specific results are built on these 1,922 comments.

4.2.2 Time series graph

First, we analyzed all comments using the YouTube Studio Analysis. We categorized results based on different time series such as day, month and year. We were curious to find out which day, month and year resulted in more and less comments.

When we examined the number of comments taking in consideration the daily timeline, we came to interesting results. First, the top number of comments are mostly assigned to the Quick-sort algorithm. Another particular result is that the top seven days with the highest number of comments are published during the weekend (from Friday to Sunday). It can be assumed that users tend to watch YouTube videos and to add more comments during the weekend.

Going forward, analyzing by month the number of published comments we came to the conclusion that users are more likely to watch these videos and add comments during the period between November-December and April (Figure 2). A possible explanation could be that this time interval is often considered as a pre-exam period when most of the students want to understand the specific educational material, such as sorting and searching algorithms. Naturally, during the summer, the number of comments decreased more than likely because of the fact that students are on vacation. This can encourage us and many other learning platforms to publish their videos around late autumn and late spring.

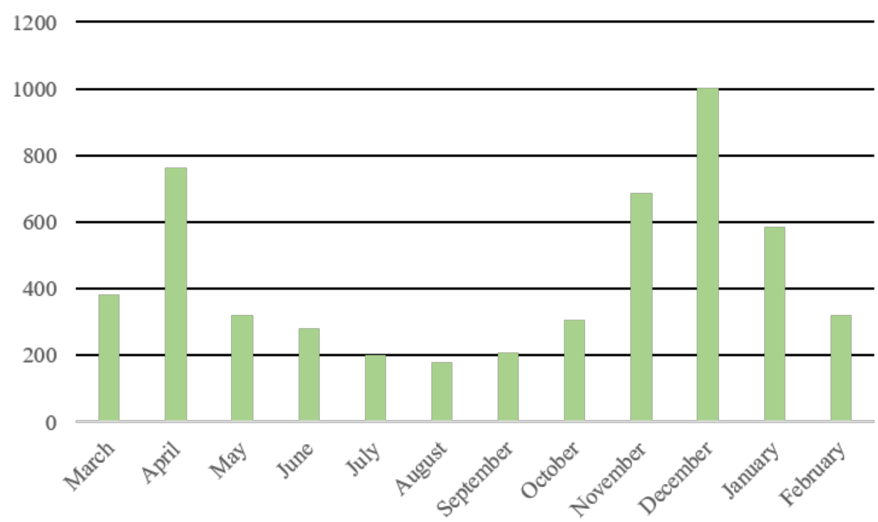

Figure 2: Number of comments by month

Analyzing the number of comments by year resulted in 2019 being the best year according to the number of published comments. This year the channel got 893 comments in total. Although, in order to find the reason behind this phenomenon requires further research a possible reason could be that in 2018 four new videos were uploaded to the AlgoRythmics channel which could bring and encourage current and new subscribers to follow the content of the channel. Between 2015-2018 a significant setback can be seen which can also be attributed to not publishing new videos since 2011.

As a next step we examined the number of comments by year based on the algorithm types as well (Figure 3). It is remarkable that the Quick-sort algorithm got the highest number of comments almost every year (2,433 comments in total) . The second most popular video is the Bubble-sort with a total of 1,184 comments. This result is not surprising since almost every education institute includes these two algorithms in the curriculum, Quick-sort being one of the most popular and hardest sorting algorithms.

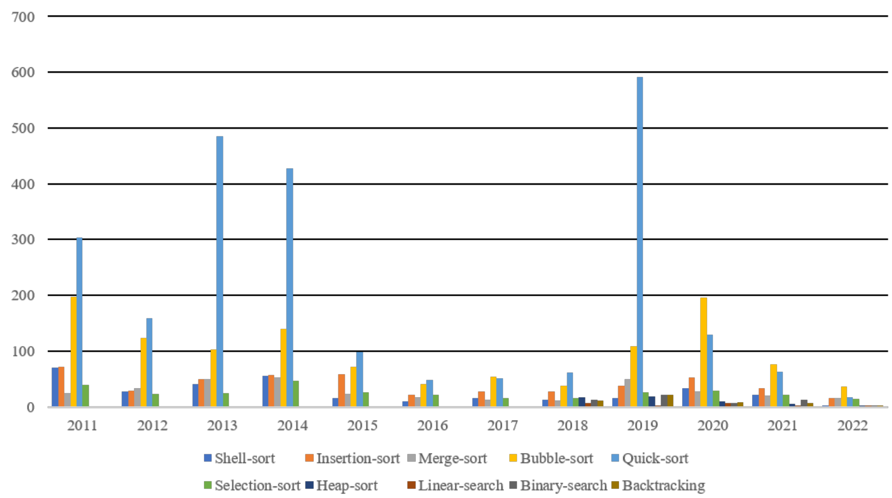

Figure 3: Comments by year and algorithm type

Although YouTube Studio is not able to show analytics grouped by comments and the age of users, an interesting fact we would like to mention is that the majority of the audience who are following the AlgoRythmics channel are people between the ages of 18-34. This fact can also enlighten the fact that we should take into account current curriculums, educational programs and exam periods when creating and publishing new videos because a large part of the audience may be students.

In accordance with this, we generated a time series graph using Mozdeh Analysis tool as well (Figure 4). As it can be seen, despite the fact that Mozdeh includes only comments in English we came to similar results. The highest number of comments were published between 2018–2020, while the fewest comments between the 2015–2018 time period.

4.2.3 Subtopic word frequency analysis

We continued the comment term frequency comparison with step 3, in which we extracted comments based on three different subtopics. We were interested in finding common words which can lead us to useful messages and instructive advice. First, we associated the following keywords to an informal educational context, namely studying for an exam: *exam, test* which resulted in 36 comments in total. Words most associated with this subtopic, compared to the remaining comments can be seen on Table 2. Words are listed in descending

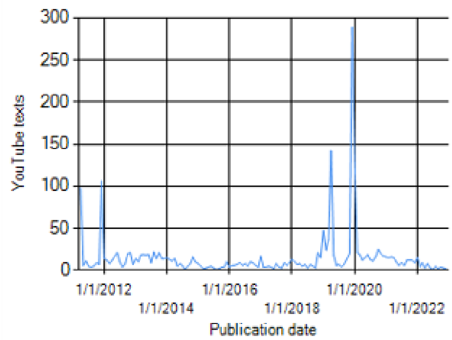

Figure 4: Time series graph by Mozdeh

order of statistical significance (Chi-squared test; Benjamini-Hochberg significance; $\alpha = 0.001$).

Word	Chisq	Sig
exam	1557.3	***
test	266.6	***
tomorrow	155.9	***
structure	130.5	***
brain	102.7	***
data	100.6	***
my	96.9	***
studying	92.6	***
during	61.1	***
cell	52.4	***

Table 2: Words associated to informal subtopic

As it can be seen from the Table, words such as brain, studying, exam, test are truly present in these comments. It is interesting to note that some of these specific words can also highlight the fact that students tend to study for their exam at the last minute and we believe that sometimes these videos give a great opportunity to understand the main concepts: **my data structure exam tomorrow**. Following step 7 has also proven the fact that these teaching materials are useful: *"Helped me during the exam"*.

Furthermore, we paid special attention to a formal educational context: studying within a school framework. We associated the following keywords to this subtopic: *class lecture teacher professor college university course* which resulted in 77 comments. Likewise, words associated with the formal subtopic can be seen on Table 3.

Word	**Chisq**	**Sig**
teacher	587.5	***
class	365.3	***
professor	242.9	***
course	218.5	***
university	194.2	***
my	136.6	***
college	121.4	***
us	106.1	***
link	102.6	***
lecture	97.3	***
sent	46.7	***
our	46.1	***
wish	32.2	***
in	30.1	***
student	26.9	***

Table 3: Words associated to formal subtopic

These words are underlining the fact that these algorithms are present in current curriculum and professors are teaching these during university courses and college classes. It is also important to mention that the word wish appears emphatically many times in these comments which can refer to a not proper, or not engaging enough education: *"I wish my professor were this entertaining"*. Analyzing these, we came to the conclusion that there are many exam-, algorithm- and education-specific words which can help us to improve the effectiveness of the channel.

Finally, in line with our previous research studies we were interested in the subtopic defined by the following keywords: *confuse confused confusing distracted distracting disturbed disturbingly*. In the spirit of step 7, the phase of verification we found many phrases in comments which can underline the presence of disturbing and confusing elements in videos:

- "That was fun to watch, but for educational purposes I think dancing people makes it much more confusing"; "Very entertaining. I must say I was a little distracted by the performance..."
- "I have to be serious, it confuses me more than it helps me"; "I can't understand , it's confusing"; "I'm even more confused"
- "Music is distracting on this one. Out of tune? Other than that ... love it."
- "I'm studying for a test, and this is somehow disturbingly helpful"

This could also justify the fact that users often want to use these visualizations to prepare themselves for an upcoming exam or test, therefore we should pay even more attention to design our videos as clearly as possible. It appears that dancing can be very entertaining but distracting as well.

4.2.4 Sentiment analysis

Furthermore, we used the Mozdeh Analysis tool in order to gather information about sentiment analysis. As it can be seen on Figure 5, the stronger a sentiment is, the more positive it could be.

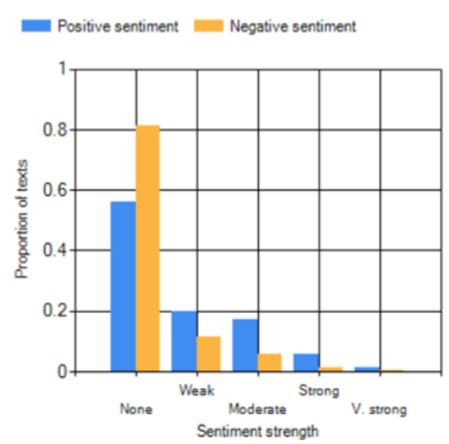

Figure 5: Sentiment analysis

In order to gain a deeper insight into this topic we extracted, as a first step, those comments that have dominantly positive sentiment values (moderate or strong positive sentiment and weak or no negative sentiment). Table 4 shows those words that occur most often in these comments compared to the remaining text. Words are listed in descending order of statistical significance (Chi-squared test; Benjamini-Hochberg significance; $\alpha = 0.001$).

Word	**Chisq**	**Sig**
awesome	231.9	***
love	134.3	***
great	129.1	***
amazing	119.3	***
wow	89.2	***
good	57.4	***
nice	52.9	***
please	47.3	***
beautiful	33.5	***
loved	33.5	***
brilliant	30.1	***
enjoyed	29.0	***
job	26.6	***
pretty	25.2	***
very	24.9	***

Table 4: Words with positive sentiment

Next, the dominantly negative comments were extracted (moderate or strong negative sentiment and weak or no positive sentiment). Table 5 contains the list of words associated with these comments. Again, words appear in the list according to their statistical significance (Chi-squared test; Benjamini-Hochberg significance; $\alpha = 0.001$).

As can be inferred from the above, AlgoRythmics channel is quite divisive when it comes to the opinions of the YouTube commenters. Since our goal is to improve AlgoRythmics visualizations, we were particularly interested in the direct or indirect suggestions found in the negative comments. According to step 7 of CTFC method we have read all the comments and have been able to identify a few areas that we would like to pay more attention to in the future.

In order to make the dance steps look more natural, in several cases we did not visualize the standard algorithm implementation. Some commentators found this confusing.

Word	Chisq	Sig
fuck	112.9	***
stupid	92.6	***
shit	83.6	***
worst	74.7	***
hate	55.2	***
case	44.0	***
nerd	35.4	***

Table 5: Words with negative sentiment

- Insertion sort: "This confused me because insert sort doesn't work the way this dance portrays it" ; "This video has been confusing for my students because it does not follow the typical implementation of insertion sort."
- Bubble sort: "Shame about the bug" + "Actually, that's not a bug, it's an optimization";
- Merge sort: "If you wanna do parallel processing with multithreading, now your threads can run simultaneously without waiting for another thread to rest…";
- Quick sort: "The pivot should be randomized... I'm really disappointed here...".

Several people complain that visualizations are too slow. This aspect is again attributable to the rhythm and style of folk music.

- "... it is veeery slow…";
- "I imagine the dance in the worst case scenario taking like 20 minutes";
- "This is a terrible implementation. If you are a Computer Science student remember to always use Jamaicans. They run much quicker."

Another possible reason for the negative opinions is highlighted in the comment below:

- "That was really boring, maybe if I didn't know how the algorithm works I would find it interesting"

In terms of positive comments, most highlighted the surprising, creative and ingenious art-science combination, the expressiveness of the visualizations, and the entertaining character of the illustrations.

- “Beautiful piece of art to explain computer science, this is so EPIC! Thanks!”; “Art + Science = Intelligent entertainment”; “I think I just found a new continent”
- “Best explanation I could imagine””; “It was so hard to learn it only with math and theories. So easy with these videos! Wonderful! Super mega nice work!”; “Epic, I totally understand it now!”; “Haha! 30 years after starting university as an IT student I finally understood quick-sort. Great visualization!”;
- “Entertaining and educational at the same time... also I find it incredibly amusing that the ‘pointers’ are the hats”; “I wish my professor were this entertaining”; “If my professors did this diddy for class I would not only be amused but I’ll never forget”.

4.2.5 Gender differences analysis

Going forward, we planned to investigate the differences between female and male users’ comments. The Mozdeh tool collected comments from a total of 1,855 users. Unfortunately, a large part of the users’ gender were unknown (Figure 6). Even so, we analyzed the presence of positive and negative sentiments in comments.

When we examined the sentiment analysis, we came to similar conclusions in both female and male users. As it can be seen on Figures 7 and 8 the stronger a sentiment is, the more positive it could be. As an overall sentiment the positive can be considered as dominant which also emerges from the average values: males Ap = 1.76%, An = 1.29%, females Ap = 1.87%, An = 1.30%.

Furthermore, taking in consideration the confidence intervals of females and males we came to the conclusion that female users are more likely to express positive emotions regarding the videos. Although, for positive sentiment, the female average is higher but the difference between males and females is statistically not significant since the confidence intervals for females (1.6869, 2.0582) and for males (1.6668, 1.8690) overlap.

In case of the negative sentiments, we came to similar results. For negative sentiment, the male average is lower and the difference is statistically not

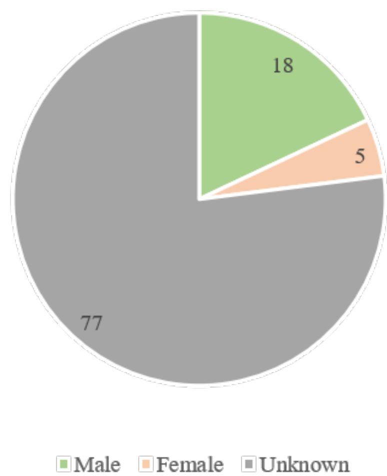

Figure 6: Distribution of genders by Mozdeh

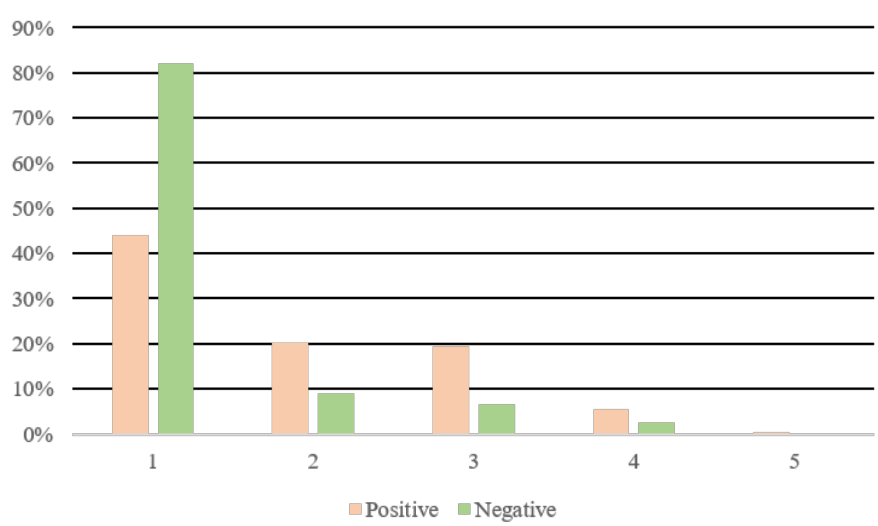

Figure 7: Sentiment analysis - male

significant since the confidence intervals for males (1.2241, 1.3719) and for females (1.1771, 1.4308) also overlap.

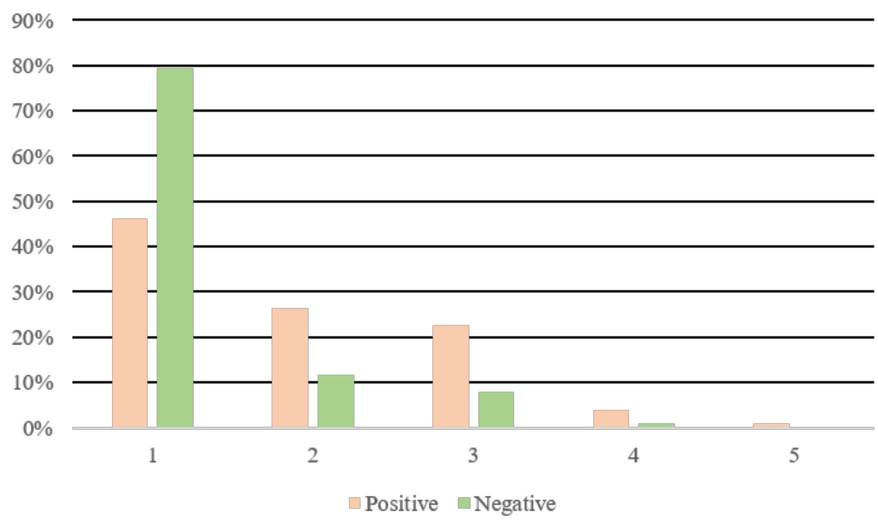

Figure 8: Sentiment analysis - female

As an interesting point of view, we would like to underline the fact that behind this dominant positive sentiment there are many words in comments which refer to positive feedback. Both females and males used expressive words such as *wow, original, glad, thank, helpful.*

4.2.6 Overview network

In order to get an overall picture, we generated the comment similarity network of the channel. Figure 9 shows the Maximal Spanning Tree of this network. It can be noticed that all nodes are leaves in the tree, except those representing algorithms Quick sort, Bubble sort, and Binary search. The grades of these nodes are 6, 3, and 2. In line with this the corresponding betweenness values are: 32, 15, and 8. Interestingly, these algorithms are the most "famous" searching and sorting strategies according to their google search (see Table 6). We searched for the given 10 algorithms as follows: *"name type" OR "name_type" OR "name-type" OR "nametype"* for example *"bubble sort" OR "bubble_sort" OR "bubble-sort" OR "bubblesort"*. This result suggests that quick-sort and binary-search are one of the "most-wanted" algorithms not only in case of our YouTube channel, but on the overall internet. Therefore, including such "famous" algorithms to the repertoire can lead to a numerous and active audience.

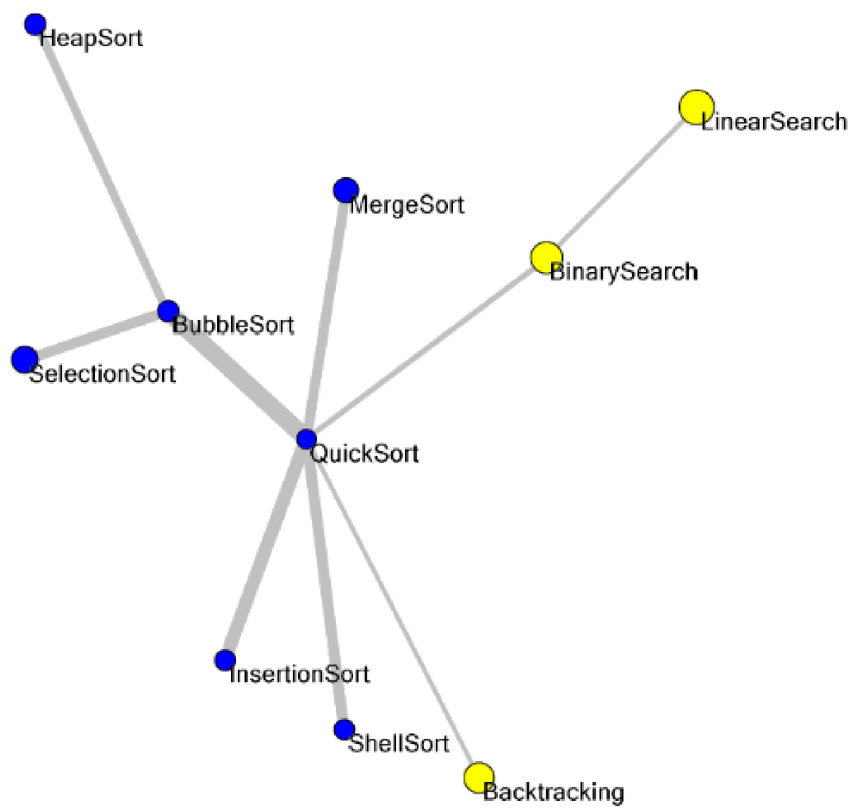

Figure 9: Network overview

5 The present and future

5.1 Moving from the dance floor to the stage

Summarizing our result, we could confirm that these videos got a lot of positive feedback from users and the overall picture contains more positive emotions than negative ones. Nonetheless, YouTube users have mentioned a lot of constructive criticism regarding the videos and the instructional material. One of the most important opinions could be that even though dance and music are very interesting and entertaining elements of the instruction material, sometimes these can result in a confusing and disturbing visualization.

Algorithm	Google results
Bubble-sort	1,700,000
Insertion-sort	1,880,000
Selection-sort	1,200,000
Merge-sort	1,740,000
Quick-sort	2,770,000
Shell-sort	1,170,000
Heap-sort	1,150,000
Linear-search	1,360,000
Binary-search	10,200,000
N Queens problem	431,000

Table 6: Google search results by algorithm

5.2 Replacing dance choreographies with theatrical scenes

In order to eliminate the disadvantages of the visualizations we decided to create our newest algorithm visualizations in a slightly different style: moving from the dance floor to the stage. Instead of dancing, acting will be considered as the main representation. This means that the presence of human movement will remain one of the most important elements of the algorithm visualization due to the "social belonging effect". On the other hand, the movement will be much easier to understand and hopefully clearer as well. Instead of dance including decorative elements users will be able to watch a simplified human movement.

Until now, this project enjoyed many collaborations with folk-, ballet and flamenco dancers as well. This time we plan to initiate a collaboration with art students from the University of Arts of Târgu Mureș. Our idea is to recreate some of the dance choreography videos using acting. We also want to create visualizations where the specific numbers of the elements of the array cannot be seen, underlining the true working process of the "blind computer".

5.3 Algorithmic escape room

Besides recreating some of the sorting algorithm videos, one of our biggest plans is to create and record an algorithmic escape room video as well. This escape room will be identified as a video full of enthusiasm, excitement and collaboration where participants need to choose the correct strategy in order to escape. As a first step, this algorithm visualization will focus on the Greedy and Dynamic Programming strategy. This visualization will also be created in collaboration with art students. Our biggest question is whether visualizations containing simplified human movement can lead to better results or even better feedback than dance choreographies.

6 Limitations

There are some limitations to this study. First, the data was collected using both YouTube Studio Analytics and Mozdeh Big Data Analysis tool, which may not have captured all comments on the videos. Additionally, the tool may have missed comments that were deleted by the channel owner or removed by YouTube for violating the platform's community guidelines.

Furthermore, the study only analyzed the text of the comments and did not consider other aspects of the comments, such as the tone or the presence of emojis or other visual elements. These elements could potentially provide additional insights into the sentiment of the comments. Despite these limitations, this study provides valuable insights into the characteristics of YouTube comments and can serve as a starting point for further research on this topic.

7 Conclusions

In conclusion, the CTFC method is a valuable tool for understanding viewer reactions and opinions about a particular video or channel on a platform. By collecting and analyzing comments, we were able to gain insights into how our videos are being received and identify opportunities for improvement. Additionally, the use of Youtube comment analytics has the potential to provide valuable insights and improve the effectiveness of videos on the platform. However, it is important to consider the limitations of our study, such as the fact that Mozdeh only collected comments in English which brings a deficiency to the study.

Using this method, we were able to collect user feedback and to find some of the most important advice and opinion given to the videos. As an overall result, we could conclude that the AlgoRythmics channel made a positive impact on users. The dominant sentiment behind these emotions is positive which underlines the fact that the audience enjoys watching these videos. Besides positive feedback we also examined constructive opinions. Words and expressions such as *too slow, confusing, disturbing, error* encourage us to pay more attention to small details when creating and planning our videos.

As a next step, we will move from the dance floor to the stage where art students will represent the algorithm strategy using simplified human movement (acting) instead of dance. We really hope that this type of visualization will solve the inconvenience caused by destructive and decorative elements and it will also be exciting at the same time.

Funding

This study was supported by the Collegium Talentum Programme of Hungary and the Sapientia Foundation—Institute for Scientific Research.

References

[1] J. C. Castro-Alonso, P. Ayres, M. Wong, F. Paas, Learning symbols from permanent and transient visual presentations: Don't overplay the hand. *Computers & Education*, **116** (2018) 1–13. ⇒281

[2] S. R. Hansen, N. H. Narayanan, D. Schrimpsher, Helping learners visualize and comprehend algorithms. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, **2,** 1 (2000) ⇒277

[3] C. D. Hundhausen, S. A. Douglas, J. T. Stasko, A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, **13,** 3 (2002) 259–290. ⇒276

[4] Z. Katai, L. Toth, A. Adorjani, K.,Multi-sensory informatics education. *Informatics in Education*, **13,** 2 (2014) 225–240. ⇒279

[5] Z. Katai, Selective hiding for improved algorithmic visualization. In *Proceedings of the 2014 conference on Innovation & technology in computer science education* 2014, pp. 33–38. ⇒279

[6] Z. Katai, Multi-sensory method for teaching-learning recursion. *Computer Applications in Eneering Education*, **19,** 2 (2011) 234–243. ⇒278

[7] Z. Katai, L. Toth, Technologically and artistically enhanced multi-sensory computer-programming education. *Teaching and teacher education*, **26,** 2, (2010) 244–251. ⇒278

[8] Z. Katai, Promoting computational thinking of both sciences-and humanities-oriented students: an instructional and motivational design perspective. *Educational Technology Research and Development*, **68,** 5 (2020) 2239–2261. ⇒280

[9] Z. Katai, E. Osztian, B. Lorincz, Investigating the computational thinking ability of young school students across grade levels in two different types of Romanian educational institutions. *NAER: Journal of New Approaches in Educational Research*, **10,** 2 (2021) 214–233. ⇒282

[10] Z. Kátai, K. Juhász, A. K. Adorjáni, On the role of senses in education. *Computers & Education*, **51,** 4 (2008) 1707–1717. ⇒278

[11] Z. Kátai, E. Osztián, Improving AlgoRythmics Teaching-Learning Environment by Asking Questions. *International Journal of Instruction*, **14,** 2 (2021) 27–44. ⇒282

[12] Z. Kátai, *Algorythmics. Technologically and artistically enhanced computer science education.* Sapientia Books. Scientia Publishing House. 2021. ⇒275

[13] Z. Kátai, E. Osztián, Visualizing Algorithms: Schematic Computer Animations versus Realistic Dance Choreography Illustrations *Acta Polytechnica Hungarica* **19,** 1 (2022) 193–210. ⇒281

[14] M. L. Nugteren, H. K. Tabbers, K. Scheiter, F. Paas, Simultaneous and sequential presentation of realistic and schematic instructional dynamic visualizations. *Handbook of human centric visualization* 2014 pp. 605—622. Springer, New York, NY. ⇒280

[15] R. E. Mayer, P. Chandler, When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of educational psychology*, **93,** 2 (2001) 390–3ö7. ⇒277

[16] P. R. Osztián, Z. Kátai, E. Osztián, Algorithm Visualization Environments: Degree of interactivity as an influence on student-learning. *IEEE Frontiers in Education Conference (FIE)* 2020 pp. 1–8. IEEE ⇒281

[17] K. Scheiter, P. Gerjets, T. Huk, B. Imhof, Y. Kammerer, The effects of realism in learning with dynamic visualizations. *Learning and Instruction*, **19,** 6 (2009) 481-–494. ⇒280

[18] J. Stevens, D, Goldberg, *For the learners' sake: A practical guide to transform your classroom and school.* 2001, Tucson, Zephyr Press ⇒277

[19] A. Valle, R. G. Cabanach, S. Rodríguez, J. C. Nuñez, J. A. González-Pienda, P. Solano, P. Rosário, A motivational perspective on the self-regulated learning in higher education. *Global Issues in Education Research*, 2007. ⇒280

[20] M. Thelwall, Social media analytics for YouTube comments: Potential and limitations. *International Journal of Social Research Methodology*, **21,** 3 (2018) 303–316. ⇒274, 283

[21] ∗ ∗ ∗ Informatics education: Europe cannot afford to miss the boat. Informatics Europe & ACM Europe Working Group on Informatics Education (IE & ACM). (2013). ⇒275

[22] ∗ ∗ ∗ Cultural education: the promotion of cultural knowledge, creativity and intercultural understanding through education. Parliamentary Assembly Council of Europe, Committee on Culture, Science and Education. 2009. ⇒278

Received: December 20, 2022 • Revised: December 30, 2022