2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #33775

How Students Search Video Captions to Learn: An Analysis of Search Terms and Behavioral Timing Data

Mr. Zhilin Zhang, University of Illinois at Urbana-Champaign

Zhilin Zhang is a 5-year BS-MS student in Computer Science at the University of Illinois at Urbana-Champaign (UIUC), co-advised by Professor Lawrence Angrave and Professor Karrie Karahalios. His research interests are in Human-Computer Interaction and Learning Sciences. He studies, designs, and builds intelligent systems to support scalable and accessible teaching and learning through a computational lens.

Ms. Bhavya Bhavya, University of Illinois at Urbana-Champaign

Bhavya is a Ph.D. student in Computer Science at the University of Illinois at Urbana-Champaign advised by Dr. Chengxiang Zhai. Her research interests are in novel applications of text mining, machine learning, and human-machine collaboration, particularly for improving education and health care.

Prof. Lawrence Angrave, University of Illinois at Urbana-Champaign

Lawrence Angrave is an award winning Fellow and Teaching Professor at the department of computer science at the University of Illinois at Urbana-Champaign (UIUC). His interests include (but are not limited to) joyful teaching, empirically-sound educational research, campus and online courses, computer science, engaging underrepresented students, improving accessibility and creating novel methods that encourage new learning opportunities and foster vibrant learning communities.

Mr. Ruihua Sui, University of Illinois at Urbana-Champaign

Ruihua Sui is a senior student in Mathematics and Computer Science at the University of Illinois at Urbana-Champaign. He is interested in software development, and was nominated for Illinois Innovation Prize and received ICCP James N. Snyder Memorial Award in 2020 because of his contribution to the educational software ClassTranscribe.

Mr. Rob Kooper, University of Illinois at Urbana-Champaign

Rob Kooper is a lead research programmer at the software directorate at the National Center for Super computing Applications. He is interested in enabling scientists to do research work using software developed with the help of NCSA as well as teaching good software principles during this process. He is interested in software deployment and scaling software deployments from small research projects to larger installations with many users.

Mr. Chirantan Mahipal, University of Illinois at Urbana-Champaign

I'm a Computer Science grad student at University of Illinois, Urbana-Champaign, working under the mentorship of Prof. Lawrence Angrave. Prior to this, I was working as a Research Fellow at Microsoft Research in the Technology for Emerging Markets (TEM) group.

Prof. Yun Huang, University of Illinois at Urbana-Champaign

Dr. Yun Huang is faculty in the School of Information Sciences at the University of Illinois at Urbana-Champaign. Her expertise is in the area of social computing, human-computer interaction, Internet of Things, and human-AI interaction. In her work, she designs, implements and evaluates social computing systems that can engage community members to co-create new services for better community wellbeing.

How Students Search Video Captions to Learn: An Analysis of Search Terms and Behavioral Timing Data

Abstract

Engineering students used ClassTranscribe, an accessible video player, in multiple engineering courses to view course videos and search for video content. The tool collected detailed timestamped student behavioral data from 1,894 students across 25 engineering courses that included what individual students searched for and when. A previous analysis, published in ASEE 2020 [1], found that using ClassTranscribe caption search significantly predicted improvement in final exam scores in a computer science course. In this paper we present how students used the search functionality based on a more detailed analysis of the log data. ClassTranscribe automatically created captions and transcripts for all lecture videos using an Azure speech-to-text system that was supplemented with crowd-sourced editing to fix captioning errors. The search functionality used the timestamped caption data to find specific video moments both within the current video or across the entire course. The number of search activities per person ranged from zero to 186 events. An in-depth analysis of the students (N=167) who performed 1,022 searches was conducted to gain insight into student search needs and behaviors. Based on the total number of searches performed, students were grouped into "Infrequent Searcher" (< 18 searches) and "Frequent Searcher" (18 to 110 searches) using clustering algorithms. The search queries used by each group were found to follow the Zipf's Law and were categorized into STEM-related terms, course logistics and others. Our study reports on students' search context, behaviors, strategies, and optimizations. Using Universal Design for Learning as a foundation, we discuss the implications for educators, designers, and developers who are interested in providing new learning pathways to support and enhance video-based learning environments.

1. Introduction

This paper presents the findings from behavioral analyses of students' caption search activities that were recorded in the user activity logs of ClassTranscribe [1, 2, 39], an enhanced video-based learning environment that provides multiple learning pathways for students to engage and learn. Previous studies have shown that caption search was reported as useful for learning by students [3] and was correlated with improved student exam scores in a sophomore-level computer science class [1]. In this study, we investigate how students use caption search in video-based learning and discuss implications for improving video-based learning environments based on Universal Design for Learning (UDL) principles.

This paper is organized as follows: The background section provides a broader context of UDL, and caption-based video searching and indexing, gives a brief overview of ClassTranscribe with

comparison to other similar tools and studies, and presents the research questions for this work. The methods section presents the data collection and analysis process. The result section shows the findings on search usage, user types, and search behaviors. Finally, we conclude the paper with the discussion section where we summarize the findings and discuss implications for future research.

2. Background

2.1 Universal Design for Learning

Universal Design for Learning (UDL) [33, 34, 35] is a conceptual framework and associated set of educational principles and practices designed to improve learning outcomes for all students. The UDL framework advocates creating a curriculum and educational environment that offers: i) Multiple means of expression to provide learners alternatives for demonstrating what they know; ii) Multiple means of representation to give learners various ways of acquiring information and knowledge; and iii) Multiple means of engagement to tap into learners' interests, challenge them appropriately, and motivate them to learn.

In this study, we researched how students interacted with ClassTranscribe, a UDL-based online learning tool and presented caption-based video searching and indexing as a new learning pathway. We studied students' behaviors in interacting with this learning pathway and proposed ways to support and enhance students' learning experiences. Previous research [36, 37, 38] has shown that UDL can provide opportunities for enhancing learning experience for students with disabilities and promote an inclusive learning environment. Our study adds to that body of research by discussing how a searchable video system can not only help learners with physical or cognitive disabilities, and learners who are non-native English speakers, but ultimately provide an inclusive enhanced video-based learning experience *for all learners*.

2.2 ClassTranscribe

This research presents findings from students' usage and behaviors data analysis in using ClassTranscribe [1, 2, 39], a UDL-inspired tool that can automatically generate text-searchable captions to lecture videos uploaded by instructors in a web interface that includes accessibility support (e.g., support for users who use a screen reader), or require a low-distraction interface. ClassTranscribe uses Automatic Speech Recognition (ASR) to generate captions for lecture videos and indexes them to facilitate keyword-based search. Students can search by keywords across all lecture videos of the same course and retrieve the relevant videos and moments in the videos. They can directly jump to a specific moment by clicking on the search result. Figures 1 and 2 provide sample screenshots of the video and searching interface of ClassTranscribe.

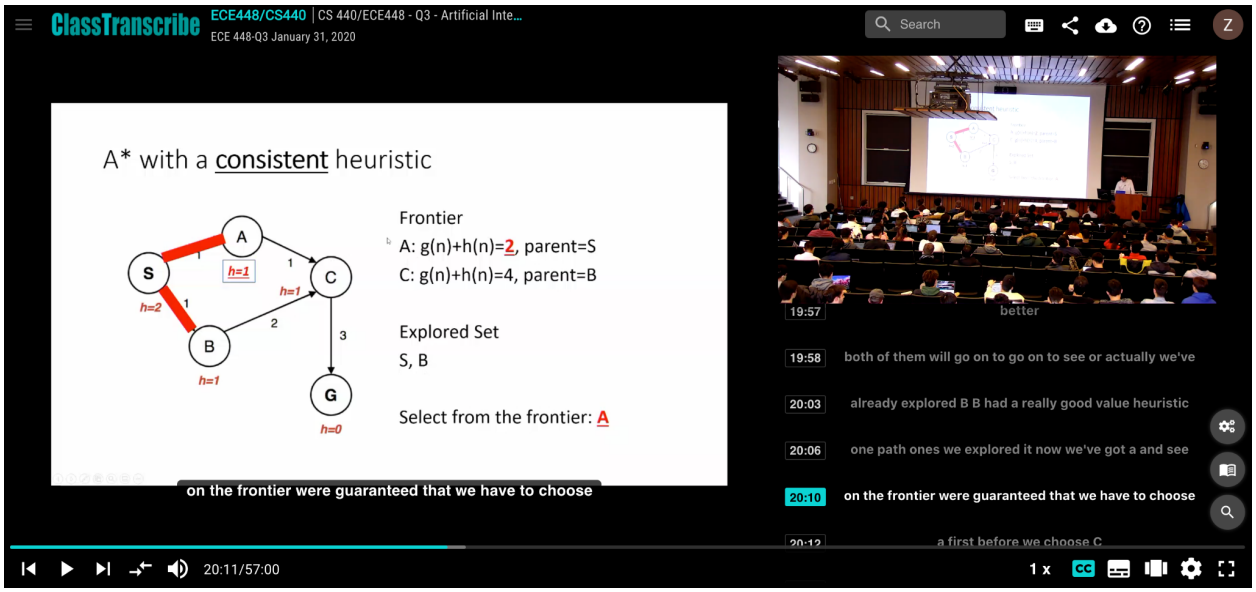
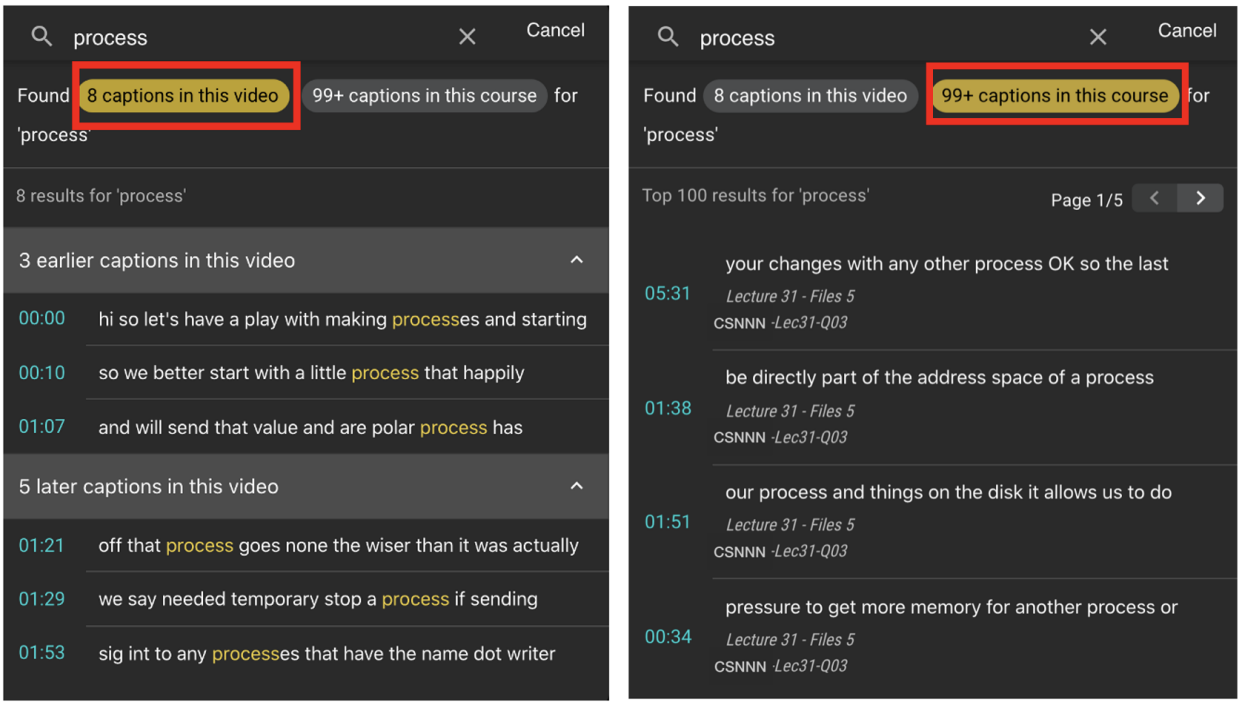


Fig 1. Video Interface of ClassTranscribe. Closed captions can be turned on and configured. Captions on the side are provided with spacing to facilitate reading by users with dyslexia.



(a) (b)

Fig 2. The Searching Interface of ClassTranscribe; captions appear on the right side of a lecture video. (a) matching captions found in the *same* lecture video as the one being currently viewed. (b) matching captions found in *other* lecture videos.

Table 1 summarizes a mapping of ClassTranscribe features to the UDL framework. The authors' opinion is that the caption-based video search functionality should be mapped into the "Provide options for Perception" item in the "Provide multiple means of Representation" category because it offers a new way to "customize the display of information" by enabling students to search and gather information they need.

UDL Guideline	**UDL Guideline item(s)**	**ClassTranscribe Feature**
Provide multiple means of Engagement	Minimize threats and distractions	- Distraction/stress-free learning interface
	Provide options for Sustaining Effort & Persistence	- Student personal usage analytic reports based on interaction with the platform
Provide multiple means of Representation	Provide options for Perception Offer ways of customizing the display of information Offer alternatives for auditory information Offer alternatives for visual information	- Captions & transcriptions available - Caption-based video search to filter caption results - Multiple cameras angles - Configurable playback interface, caption font size and color, background color, background transparency - Can generate textbook from video in multiple formats (epub, pdf, html, etc) from videos
	Provide options for Language & Symbols	- Multiple languages for captions/transcriptions
Provide multiple means of Action & Expression	Provide options for Physical Action	Accessible design - Multiple keyboard shortcuts - Support screen readers for users who are blind or have low-vision - Continue watching a partially completed video
	Provide options for Expression & Communication	- Create shareable link for specific video moments

Table 1. Features of ClassTranscribe classified using UDL.
(Guideline items reproduced from udlguidelines.cast.org)

2.3 Caption-Based Video Searching and Indexing

A previous study found an enhanced video-based learning environment (i.e., with embedded note-taking, supplemental resources, and practice questions) significantly improved recall test scores over a common video environment limited to play, pause, rewind, and forward operations [5]. Researchers have also experimented with video-lectures in a traditional mathematics course (Calculus). The students' feedback and usage analysis showed that video resources are considered useful and are correlated with improved academic results [6]. Based on the positive learning outcomes from video-based learning environments, researchers have explored new features in video-based learning environments and are creating new learning pathways for students. One of these new features is caption-based video searching and indexing.

Education researchers have shown that the introduction of searchable video lectures in a sophomore-level system programming course, which complemented the equivalent online book content, saw an increase in course performance for all students [2]. Similarly, in a freshman-level introduction to electronics course, translated searchable class videos led to improved course scores among students that used them [1]. With the demonstrated improved performance in CS and ECE courses, we were curious to better understand how students used the search functionality as part of their authentic university course experience (as opposed to a simulated learning environment). This study is based on student usage data from 25 engineering courses to explore how students used video-based index search, and provide insights so educators can learn about student behaviors and learning outcomes related to their use of the search and indexing features.

A similar tool, Indexed Captioned Searchable Videos (ICS), has been built and studied by researchers at the University of Houston. They conducted a student survey that found students considered indexing and search features to be helpful [3]. Another study [7] explored the design space of video navigation. The study created a prototype video tool, lectureScape, and used it in a simulated learning environment. The tool provided several UI navigation mechanisms including visual and keyword and the study participants performed pre-specified navigation tasks to search and watch a 15-minute video.

To the best of our knowledge, this is the first study to analyze students' usage and behaviors in caption-based video searching and indexing in multiple engineering courses and in an authentic engineering setting.

2.4 Research Questions

Using the event data logged by the video system, our goals were to better understand students' interactions with this video-based learning and propose methods to support and enhance it for a more effective learning experience. We asked the following research questions about searching captions in educational videos.

RQ1: What do searchers search for using caption-based video search?
RQ2: What collective behaviors do searchers exhibit before, during and after search?
RQ3: Is caption-search used to review previously viewed content or to find new content?

The interaction of each student with the ClassTranscribe website was recorded by the server as a time series of event data stored in a database. This included data that represented student actions e.g., starting a search or playing and pausing a video. Each event included a student identifier, a timestamp and details about the specific event. To address the above research questions we aggregated by student identifier the event data and analyzed the event sequence and event types (e.g., search, seeking, video play) performed by each student. This is described in the following section.

3. Methods

3.1 Event Types

Every user interaction (event) on the ClassTranscribe website was logged by the server for later analysis. The event types logged by the system are listed and described in Table 2. The *filtertrans* event is logged when a user performed keyword-based search on captions. From the search results, when a user clicked on a found caption in the same lecture video, the seeking event was triggered and the tool automatically adjusted (scrubbed) the playback position to jump to the corresponding video moment (Figure 2). Once the scrubbing completed, the *seeked* event was generated. The *seeking* and *seeked* events could also be generated when the user manually adjusted the video playback position. When the user navigated to another lecture video (e.g., by clicking on a caption line in the search results), the *changevideo* event was generated. Thus, *seeking*, *seeked*, and *changevideo* events after a search indicated that the learner clicked on a caption matching the search query.

Event Type	**Event Description**
timeupdate	Watched another 15 seconds of video
play	Played a video
seeking	Adjusted to a video playback position
seeked	Finished loading to a video playback position
pause	Paused the video
userinactive	Switched focus away to another webpage
changevideo	Changed to a different video
filtertrans	Updated search filter of the transcription interface
edittrans	Edited the transcript; updated a caption line of text

Table 2. The Event types logged by ClassTranscribe.

3.2 Mining System Logs

To answer our research questions, we mined the system event logs. Log mining is commonly used to analyze search behavior [31, 32]. We defined **search behavior** in terms of frequency of searches, types of search queries, frequency and types of other events (Table 2) around search performed by students. Previous work [28] used a similar definition of search behavior for educational video search, but our analysis was more detailed (e.g., we also studied the types of search queries).

Figure 3 shows an overview of the study design. With approval from the Institutional Review Board (IRB), we collected anonymized student interaction data logged by ClassTranscribe. The data included interaction events from 1,894 students across 25 engineering courses during 09/2019 to 07/2020. For each user interaction with the tool, the event logs contain information about the user (anonymized UserID), the Event Type (Table 2), the timestamp of the interaction, the lecture video being viewed at the time of interaction (Video ID), and additional details relevant to the interaction (e.g., search query in case of *filtertrans* event).

Event logs were extracted from the system's Postgres database using a SQL query and subsequent data processing and analysis was performed in Python using the Pandas, numpy and scipy modules. Search data was grouped by view using an anonymous identifier. Widely used statistical techniques (e.g., Chi-square tests [12]), data mining techniques (e.g., k-means clustering [29]) and qualitative analysis techniques (e.g., manual coding using a grounded approach [30]) were used for further analysis as described in the following section.

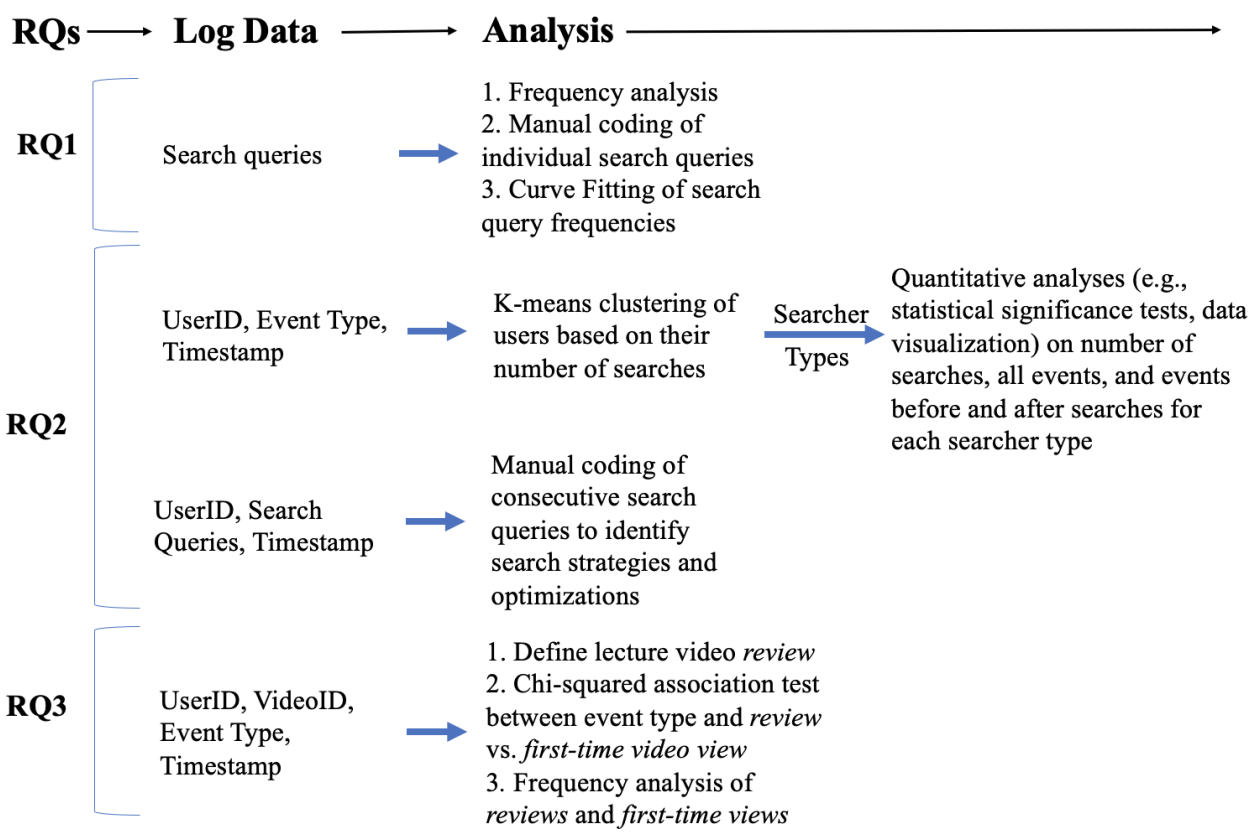


Fig 3. Overview of the analytical architecture

4. Results

4.1 What do searchers search for using caption-based video search? (RQ1)

4.1.1 Number of Searches

Among the 1,894 students, the number of search activities per person ranged from 0 to 186 events. A time-series analysis of the students (N=167) who performed 1,022 searches found search was used throughout the semester with peak usage occurring in the middle of the semester. One possible reason for the relatively low 9% (167/1894) usage rate of the searching function was that students were unaware that the search functionality existed. This is supported by survey results previously reported in [1], where only 40% (96/242) responded affirmatively to "Were you aware of ClassTranscribe's ability to search for course content using text search?". Thus an estimated usage rate for the students who were aware of this functionality would be 22% (167/758). These results suggest user interface designers of digital learning aids should consider the importance of awareness of the feature when tools are used in authentic settings that extend beyond a simulated classroom experience.

4.1.2 Categories of Search Queries

A manual analysis of the search queries data found three major categories of search terms: *STEM-related keywords*, *Logistics*, and *Others* (see Table 3). Among the 589 unique search queries, 427 (72.5%) were *STEM-related keywords*, 45 (7.6%) were on logistics of the class, and 117 (19.9%) fell into the *Others* category. In terms of the percentage of total searches (1022), *STEM-related keywords* constituted 71.7%, *Logistics* constituted 9.5%, and *Others* constituted the remaining 18.8%. As the majority, *STEM-related keywords* were primarily domain words used in the class. We speculate that students were searching for these keywords to learn about specific concepts in the class. *STEM-related keywords* constituted 37 (74%) of the top 50 most frequently searched keywords. Another frequently searched keyword category was *Logistics*. We observed that students searched for keywords about academic integrity and course policies on collaboration for assignments throughout the semester. These topics are normally specified on the course website. However, a collaboration policy can vary significantly between courses, the same course in different semesters and even vary between instructors of different sections of the same course. Thus it is unsurprising that we found searches related to collaboration and cheating.

The *Others* category consisted of more general search queries, which we further categorized into i) *search-functionality exploration*, ii) *direct navigation*, and iii) *indirect navigation* behaviors that we discuss here. Queries that explored the search functionality used words that were known to be in the transcript (e.g. “hello” and “i”). *Direct navigation* searching used a specific topic or course item. For example, some students searched for “L25" (lecture 25), “mp 7” (machine problem 7), “question eight”, and “csNNN” (course number anonymized) to directly navigate to the relevant point in the video. Lastly we defined *indirect navigation* where students recalled that the instructor uttered a key or unique phrase and used this as a proxy for a specific moment in a video search. Example search queries included, “essentially”, “actual”, “i thought i knew”, “i don't have the operator.”

Category	**Sub-category**	**Example queries and number of search occurrences**
STEM-related Keywords (72.7%)		malloc (31), process (18), intro to diodes (17), deadlock (16), Bresen (10)
Logistics (7.7%)		cheating (17), collaboration (10), plagiarism (9), academic (5), lab (3)
Others (19.6%)	Exploring search functionality	i (8), your (5), hello (5), little (4), everything (3)
	Direct Navigation	csNNN (5), L25 (2), mp seven (2), problem/question eight (2)
	Indirect Navigation	Essentially (1), actual (1), i thought i knew (1), i don't have the operator (1)

Table 3. Categories of the search queries with corresponding counts.

4.1.3 Zipf's Law of Search Keywords

Zipf's law is an empirical law proposed by linguist George Kingsley Zipf [8, 9], who observed that in a large corpus of written words, the frequency of any word was inversely proportional to its rank in a frequency-of-use table. This is often roughly described as "there will be a few common words and a large number of rare words," however, Zipf's law is a stronger empirical observation that the frequency of utterances follows a power law distribution. Previous research shows that search queries follow Zipf's law, (e.g., on domain-specific queries on Google [14]) . So, we aimed to investigate whether this law also holds on queries in educational videos.

As shown in Figure 4, we plotted the top 200 search keywords by its frequency rank and number of occurrences, and utilized the scipy package to generate the fitted curve following the Zipf's law (kx^{-s}), where x is the frequency rank of and k,s are parameters. The best fit was found at $k = 32.1,\ s = 0.58$. A two-tailed Kolmogorov–Smirnov test [13] between the actual and fitted curves indicated that the number of occurrences of keywords follows a Zipfian distribution, $D(592)=0.41$, $p<0.005$. We concluded that Zipf's Law also holds for video-based searching in an educational context. We expect this to hold true especially in case of searches within domain-specific courses (e.g., engineering) because the queries are generally coherent and coherence is an important requirement for Zipf's law [15]. For example, as described in section 4.1.2, *STEM-related keywords* constituted the majority of searches in our dataset.

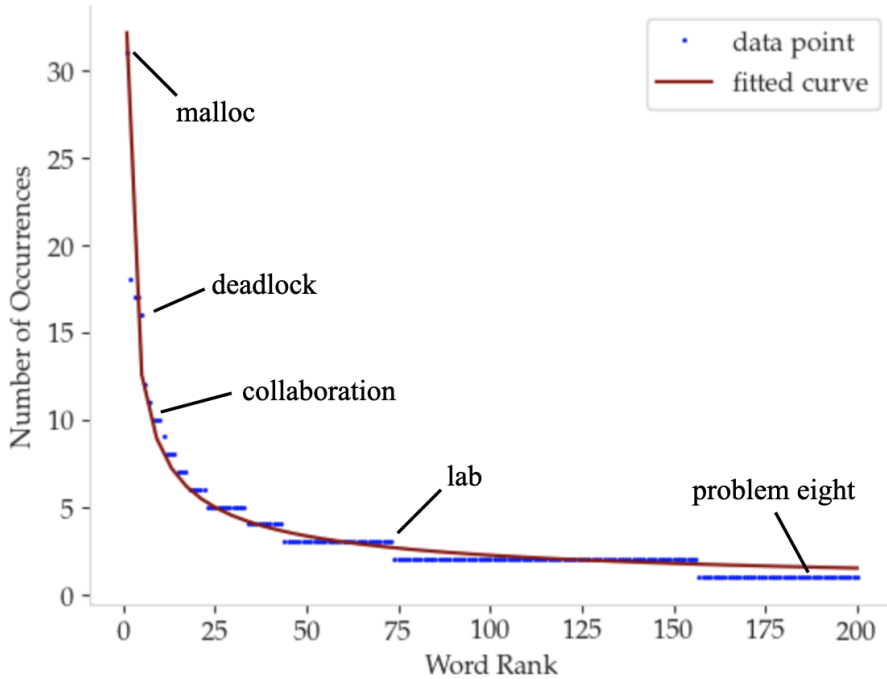


Fig 4. Zipf's Law in Search Keywords

4.2 What collective behaviors do searchers exhibit before, during and after search? (RQ2)

4.2.1 Classification of users based on search frequency

As stated in Section 4.1 above, the total number of searches per student has a large range of [0, 186]. One hypothesis is that more active users, i.e., users who have more events overall on the tool, performed more searches as well. To check this hypothesis, we plotted the number of searches vs. total number events for all users as shown in Figure 5. As we can see from the plot, this was not generally true; the students who performed no searches are spread across all activity levels. There is one outlier point with > 1 million activities and 186 searches, that is not shown in the plot. After removing the outlier, we computed the Pearson correlation coefficient and found no linear correlation $r(1893)=0.05$, $p<0.05$. This suggests that there can be variances in search usage among students with similar activity levels. To further study these variances, we performed k-means clustering [29] of users based on their number of searches as described below.

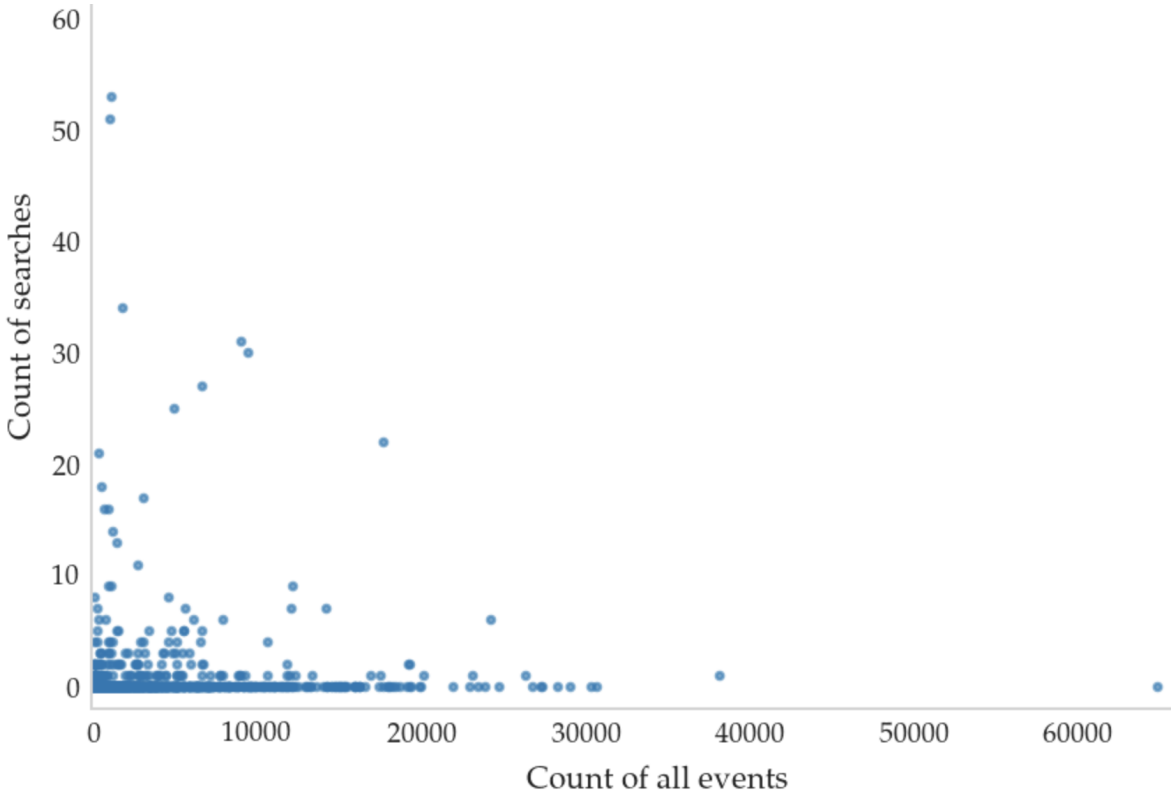


Fig 5. Total searches vs. the total logged records. Each dot represents one student.

Firstly, to determine the optimal number of clusters, we used the Elbow method [10]. Figure 6. shows the With-in-Sum-of-Squares (WSS) distance vs. number of clusters. Based on this plot, we selected 3 as the optimal number. Next, since k-means can sometimes give suboptimal results, we performed k-means 10 times with different centroid seeds and selected the results with the best WSS. Figure 7 shows the results of the k-means algorithm. From the results, we identified two main types of students based on their counts of searches: “Infrequent Searcher” (<=17 searches) and “Frequent Searcher” (18 to 110 searches). These thresholds are the midpoints between two (sorted) cluster centroids. There was only one student with >110 searches, who is identified as an outlier and excluded in further analyses. After removing the outlier, there were 166 students with 837 total searches. Further, since our focus is on students who performed searches, we used “Infrequent Searcher” for students with 1 to 17 searches and “Frequent Searcher” for students with 18 - 110 searches. Using this categorization, we found 155 Infrequent Searchers and 11 Frequent Searcher users.

To further investigate the differences between the groups, we additionally looked at their number of all activities performed and their durations on the tool. Here, duration refers to the time difference (number of days) between a user’s first logged activity and their last activity on the tool. A Mann-Whitney U test [11, 42] was used to compare these distributions for Frequent Searchers and Infrequent Searchers, i.e., the two extremes. We used this test because, unlike the student t-test, it does not require the assumption that the data are normally distributed. Based on

the test, we found the following: (1) the **duration** of Frequent Searchers (mean=157.2 days, std.=106.8) is significantly longer than the duration of Infrequent Searchers (mean=66.3 days,std.=70.6); z=3.22,p<0.005; (2) the **number of all activities** performed by Frequent Searchers (mean=5073.4,std.=5330.4) and Infrequent Searchers (mean=4349.8,std.=5839.1) is not found to be significantly different; z=0.9, p=0.36. Further, using a Chi-square test [12], there was a significant relationship between the *searcher type* (i.e., Infrequent Searchers or Frequent Searchers) and *Event Type* (i.e., Search or Not Search) , $\chi^2(1,N=4{,}574{,}297)=13{,}816.88$, $p<0.01$. This suggests that Frequent Searchers not only performed more searches but were more likely to perform Search over other Event Types and vice versa.

We note that the two groups had different sample sizes and the number of Infrequent Searchers is small. However, we used statistical tests (Mann-Whitney U and Chi-square tests) to compare them and found significant p-values <0.01 suggesting that they are different. Both the tests are shown to work well with unequal sample sizes [11, 40]. The Mann-Whitney U is also known to be robust for reasonably small sample sizes [11]. For the Chi-square test performed above, the small number of Infrequent Searchers was not a concern because the corresponding expected frequencies (cells in contingency table) were not small (>300).

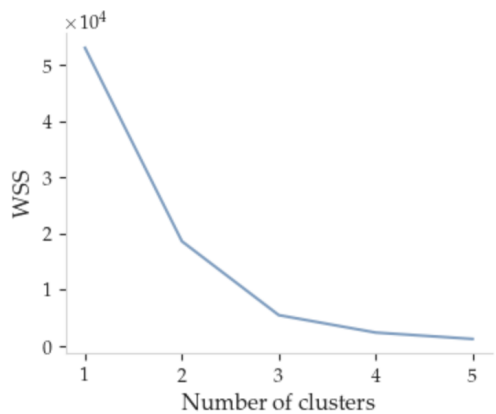


Fig 6. Determining the optimal number of clusters using the Elbow Method

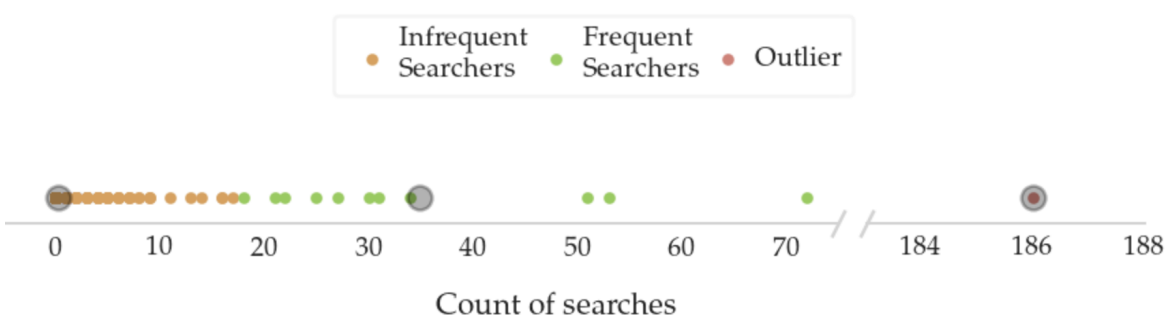


Fig 7. k-means clusters of students based on their counts of searches. Colored dots represent the students clustered into 3 clusters. Black bubbles are cluster centroids; cluster threshold boundaries (not shown) were equidistant between neighboring centroids.

4.2.2 Activities before and after searches

To investigate searchers' activities in terms of their interaction with the system immediately before and after a search event, we aggregated events surrounding the search activity separately for different groups of searchers (Figure 8). In this figure, "I" stands for Infrequent Searchers, and "F" stands for Frequent Searchers. The y-axis represents the event offset from the search activity event. For example, "+1" identifies the exact next event after the search, and "-2" identifies the second event before the search event. Each colored bar represents the sum of counts of a particular activity for a given group and their corresponding percentages are shown.

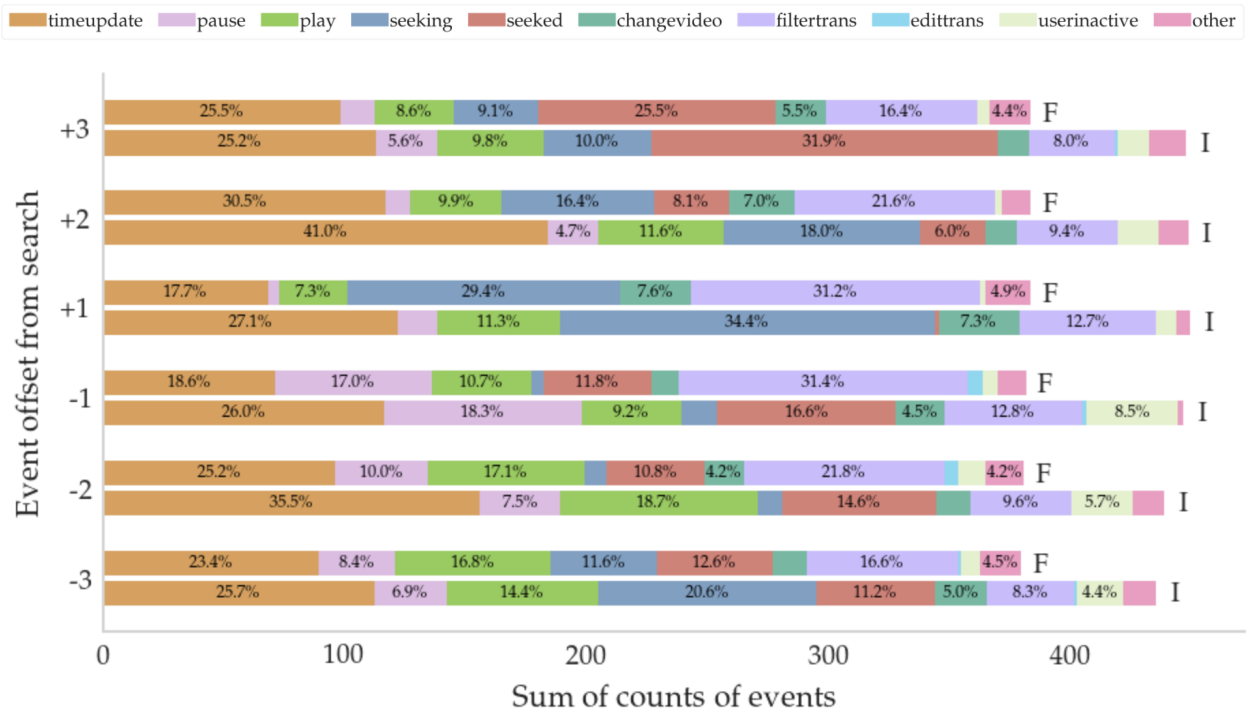


Fig 8. Distribution of events before and after a logged search event, aggregated for Infrequent Searchers (I) and Frequent Searchers (F).

By comparing the events surrounding the search activity, we found that Infrequent Searchers tended to continue watching (e.g., mostly performing events like *timeupdate, seeking, changevideo*), while Frequent Searchers tended to continue searching (i.e, *filtertrans*). The continous/consecutive search behavior is discussed in more detail in Section 4.2.3. We noticed that Infrequent Searchers were more likely to start searching from the state of *userinactive* (8.5% at -1 offset) compared to Frequent Searchers (under 4% at -1 offset), which means they switched to ClassTranscribe from another webpage to perform the search. Our hypothesis was that these

students were probably working on homework assignments and trying to look for information. However, we would need more information to support this.

We also note that overall, the number of activities performed collectively by Frequent Searchers around a search was only slightly fewer than that performed by Infrequent Searchers. This is expected because Frequent Searchers perform more searches, so there's overall more activity around search for Frequent Searchers. Although, at first glance, we might expect the number of activities collectively performed by all Infrequent Searchers to be much higher compared to that of all Frequent Searchers because 1) the number of Frequent Searchers is smaller than the number of Infrequent Searchers (11 vs 155); 2) Frequent Searchers and Infrequent Searchers generally performed the same total number of all activities over their entire durations on the tool. However, the high number of Frequent Searchers searches compared to those of Infrequent Searchers seemed to mostly nullify the potential higher counts due to those two statements.

Finally, we observed that *seeking* and *changevideo* activities were more likely to occur immediately after a search event compared to before for both types of searchers. For Infrequent Searchers, the *seeking* event comprises 34.4% of all activities immediately after the search compared to under 4% immediately prior i.e., about a third of searches, Infrequent Searchers clicked on a caption presented in the search results and seeked to a location in the same video suggesting they likely found relevant results.

4.2.3 Search Sequences and refinement strategies

By manually coding the search queries, four major types of caption-search behaviors were identified. They were1) *search and stop*, 2) *repeated search on the same query*, 3) *search again with minor modifications*, and 4) *search again with major modifications*. The *"search and stop*" behavior happened frequently for Infrequent Searchers users. We speculate that they either found the information they needed or failed and lost interest in searching. Conversely, Frequent Searchers users were more likely to repeatedly search for the same queries or modify their search queries and search again.

For *repeated search on the same query*, we suggest this was employed when students were performing a focused search on a major course concept. For query modifications, we found that users generally either changed their queries a small amount or changed their queries significantly but still under the same topic. An example sequence of *minor modifications* was "eight", "problem eight", "question eight" and, "eight". We speculate that this student recalled that the instructor explained the eighth problem and tried to retrieve that part, but they were unsure what exactly the instructor said or how it was transcribed. So they started with a general query "eight", and then tried "problem eight" and "question eight". However, they ultimately went back to searching for "eight", so we suspect that the student observed too many results by searching

“eight”, but too few results by searching “eight problem”. Another example of a sequence of searches with minor modifications was “Bresenham’s” “bresen” “Bresenham’s” “Bresenham’s” “Brese” “Bresenham’s” and finally, “Bresen”.

For searches with *major modifications*, we observed that one user searched for “doubling” “3t3” “cell” “mus musculus” and, “mouse”. We speculate that they were trying to retrieve relevant moments in the videos that discussed related bio-engineering concepts. On the other hand, students also optimized their search queries based on the auto-generated captions. For example, one user searched for “calloc” and then “kellog”; we speculate that they noticed that the term “calloc” was often mistranscribed as “kellog” so optimized their search queries accordingly.

Overall, in Section 4.2, we identified two types of users, Infrequent and Frequent Searchers who use the tool very differently. Infrequent Searchers tend to perform a large number of activities within a shorter period of time with a likeliness to perform other Event Types, whereas Frequent Searchers perform the same number of activities over a longer period with a likeliness to perform searches (*filtertrans*). Further, Frequent Searchers tended to continue searching around a single search event. Table 4 summarizes their main differences.

	IS (N=155)	**FS (N=11)**
Number of searches	*Significantly Lower* (mean=2.92,std.=3.03)	*Significantly Higher* (mean=34.91,std=15.98)
Duration (days) on tool	*Significantly Shorter* (mean=66.3, std.=70.6)	*Significantly Longer* (mean=157.2, std.=106.8)
Total number of activities	*No statistically significant difference found* (mean=4349.8,std.=5,839.1)	*No statistically significant difference found* (mean=5,073.4,std.=5330.4)
Likeliness of performing search vs. other Event Types	*More likely to perform other Event Types*	*More likely to perform search (filtertrans)*
Top 3 activities immediately before search	*timeupdate* (26%), *pause* (18.3%), *seeked* (16.6%)	*filtertrans* (31.4%), *timeupdate* (18.6%), *pause* (17%)
Top 3 activities immediately after search	*seeking* (34.4%), *timeupdate* (27.1%), *filtertrans* (12.1%)	*filtertrans* (31.2%), *seeking* (29.4%), *timeupdate* (17.7%)

Table 4. Summary of differences between Infrequent Searchers (IS) and Frequent Searchers (FS).

4.3. Is caption search used to review previously viewed content or to find new content? (RQ3)

We used the event log analysis to explore what video content was viewed after a student completed a search. We also asked if students were more likely to search when viewing new content or reviewing content.

To address the latter question we first defined a video review as the user performing any activity on a video at least one day after watching the same video. The one day threshold was justified as a reasonable time-period because a user may continue watching the same video over the duration of one day even if they take multiple breaks in between. Exploring alternative definitions and alternative review thresholds is out of scope for this study and a focus of future work.

Table 5 presents the number of searches and other activities that occurred during a review. We can see that 29.7% of the searches occur during a review compared to 19.4% of other activities. Using a Chi-square test [12], there was a significant relationship between the activity type (i.e., Search or Not Search) and review, $\chi^2(1,N=730,031)=57.0$, $p<0.001$. This suggests that searches were more likely to occur during a review. In terms of the number of users, 33.1% of all searchers performed at least one search during a review.

	Review	**Not Review**	**Total**
Search	249	588	837
Not Search	141,237	587,957	729,194
Total	141,486	588,545	730,031

Table 5. Contingency table for a Chi-square test of association between searches and video review.

We acknowledge that users may have attended the on-campus version of some of those lectures, and hence may actually be reviewing the material during their first time use of the tool for that content. However, this limitation does not diminish the above finding; that students were more likely to search for content when reviewing the same video content again.

To explore whether students generally navigated to a previously unwatched video from the search results, we computed the percentage of videos that students were watching for the first time on the tool when students navigated to the video after a search (i.e., a *changevideo* event). Out of the total 837 searches, 7.3% (61/837) led to a *changevideo* event. We found that 77% (47/61) of those searches led to videos being viewed for the first time and 67% (41/61) occurred during a Review (see Table 6).

	Review	**First-time View**
Review	11	6
First-time View	3	41

Table 6. Number of searches that led to a *changevideo* event. Rows represent searches that start from Review/First-time View and columns represent searches that go to Review/View

5. Discussion and Concluding Remarks

Our findings have both practical implications (e.g., improving search algorithms) and design implications for UDL for improving video-based learning as described below.

5.1 Challenges & Opportunities for Improving Non-Linear Navigation of Educational Videos

Our study shows that students used caption-based search as a method to navigate educational videos. Based on our results we suggest there are several challenges and opportunities to better support students in this task as listed below.

5.1.1 A Taxonomy of Search Queries

Taxonomies of search queries have been developed for other domains such as general web search [17], e-commerce [18], and educational queries on the web [19] but not for caption-based search for educational videos. In Section 4.1.2, we categorized the search queries that students used for retrieving content for video-based learning in an authentic classroom setting. This provides a better understanding of the students' informational needs and a useful classification for additional research in this area.

5.1.2 Searching for imperfect and inaccurate captions

In Section 4.2.3, we observed that errors existed in automatically generated captions. This is unsurprising as generating captions especially for STEM videos is challenging [16]. We found that students used strategies to modify their queries to match the erroneous words (e.g., search for "kellog"in place of "calloc") to overcome those errors. We suggest phoneme (audio) indexing (e.g. Soundex[43]) would facilitate successful matching for similarly sounding words and phrases that are misspelled or mistranscribed. Further it might be possible to train machine learning models to learn from student search strategies (e.g., based on phonetic similarity of consecutive search terms) and utilize them to correct errors in the lecture video captions.

Some users also corrected the transcripts using the *edittrans* function supported by the platform (refer to Figure 8), though the amount was only a small fraction of their activities (<4%). In other words, occasionally, when users were searching, they might observe a caption error and be motivated to correct it. Future education tools could use this opportunity to nudge students to correct those errors because this would immediately help them find better results. Learnersourcing [24, 25] uses a similar approach to motivate learners to improve educational content such that it benefits both individuals and their peers. For example, learnersourcing has been used for editing foreign language captions [21].

5.1.3 Search algorithms

Firstly, from sections 4.1.2 and 4.1.3, we found that the majority of the frequently searched terms were STEM-related. Thus, it would be productive for future work to focus efforts on improving the accuracy and utility of the tools for searching domain-specific keywords. However, there is also a long-tail of search queries. Similar to other search domains like general web [22], improving the accuracy for the long tail of all "rare" and unique queries may be challenging.

Secondly, to further study and improve accuracy of search algorithms, it is important to define what constitutes a successful search. The event sequences described in this study could be used to build a naive indicator of a successful within-video and within-course search activity. Currently, success is based on an increase in the number of *seeking* activities after search (Section 4.2.3) and we found that keyword-based search seems reasonably sufficient perhaps because students search for terms that are present in the lecture captions (e.g. STEM terms discussed in Section 4.1).

However identifying successful searches based on event logs will require additional studies. For example, the full caption text of the search results may have satisfied the student's question, or the subsequent duration of video viewed, i.e. dwell-time, (vs. the user performing additional searches or selecting alternative video moments) may be a useful measure of the value of the video clip(s) proposed by the system or selected by the user. Beyond clicks and dwell-time, researchers could also measure the utility of the search results as it is a better indicator of success [26]. For example, researchers have used questionnaires to manually measure information learned during educational video search in a lab-setting [28]. Further research is required for automatically measuring utility of search results in authentic settings.

Thirdly, we found that students used multiple consecutive search queries to potentially find a better and complete set of search results (Section 4.2.3). Previous work [28] showed that students who used fewer queries (keywords) during video search had better learning outcomes. Thus, a future research question is, "Is it possible to automatically suggest precise queries by training machine learning models on search log data and caption data?" For example, automated Query

Suggestion and Reformulation techniques for general web search [23] or K-12 educational search [20] could be adapted for caption-based educational video search. However, care must be taken as to not distract students or increase their cognitive burden [28].

5.1.4 Student-driven content

Our findings in Section 4.3 show the students used search to navigate to previously unwatched videos suggesting that search could be used as a means to discover new (unwatched) educational videos in a targeted way. Such non-linear navigation patterns (i.e., not strictly following the lecture video order defined in the course) using caption search suggests that the search log activity can potentially be used in two pedagogically interesting ways. Firstly, to automatically recommend relevant video segments to students. Secondly, to structure the course content in a way that students tend to find more useful compared to the traditional fixed course structure that is defined by the instructor. Using log data to augment educational videos has been shown to be useful to students [7]. Our exploration provides more insights on leveraging log data for video-based learning. For example, we found that students search when reviewing a video a second time (Section 4.3). Future studies could manually or automatically [27] identify student intent behind search queries, e.g., preparing for exams vs. working on assignments. Combining intent with the log data, it may be possible to automatically create exam helper modules that map difficult exam concepts (i.e., user search terms) to the corresponding video time segments that students found most useful.

5.2 Implications for UDL

Universal Design for Learning provides a framework and context to construct and deliver accessible and inclusive courses. Moreover, when implemented, UDL provides each student with multiple learning pathways that can span and intersect with multiple content modalities (e.g., video, captions, figures, audio and descriptive text). The caption-based video search function could potentially contribute to all three principles of UDL: “provide multiple means of engagement”, “provide multiple means of representation”, and “provide multiple means of action & expression.” We suggest that the caption-based video search function should be mapped to the “multiple representation” category of the UDL principles since it provides learners with a new learning pathway to retrieve content they need and customize the display of information. Analyses previously reported at ASEE [1] found search behaviors were predictive of improved exam scores. From the perspective of UDL - it is unsurprising that student search behaviors cannot be characterized by a single prototypical search behavior. Instead engineering students used "Infrequent Searchers" and "Frequent Searchers" behavioral patterns to find and review relevant content (Section 4.2). As the UDL framework has been discussed and adopted to support the needs of all students [41], our study shows that caption-based video search can still be improved to support varied navigational learning behaviors and needs in video-based

educational platforms. For future work, we will further investigate types of queries by frequent and infrequent searchers and provide design guidelines for caption-based video search. We encourage all instructors to adopt a text-searchable and accessible video platform, as part of a larger UDL-based approach to effective, accessible, and inclusive education.

Summary

In this paper, we studied and reported in detail the search-related activities of students in engineering courses using ClassTranscribe, an online web application that supports principles of Universal Design for Learning. By analyzing the system logs of student interactions with the tool, we studied the student behaviors during caption-based search. Our findings have both practical implications and implications for UDL for improving video-based learning, specifically by using caption-based search. We identified and fitted a Zipfian power law in search query terms ($k = 32.1,\ s = 0.58$), created a taxonomy and categorization of search queries and examined video and search-related actions prior to- and post- search events that varied between students categorized as Frequent Searchers and Infrequent Searchers. We examined differences in search and video-choice behaviors when students were watching new content versus reviewing previously-viewed content. A detailed understanding of search not only provided insights into student search-based interactions to find video content but also suggested how students are using searchable video-based content to learn in undergraduate engineering courses, and raised new research questions and ideas to improve the pedagogical utility of caption-search, video presentation and video-based learning. These results also demonstrated that students choose different learning pathways (infrequent vs. frequent vs. no searching) which provides empirical support for a UDL approach to course content design and delivery.

Limitations

The results presented in this study include event data from COVID-19 affected semesters and non-COVID19 semesters prior to 2020. The data are from authentic learning environments of engineering courses under non-laboratory controlled conditions. Our current analysis is limited to event logs of higher education students in a subset of undergraduate engineering courses at one university in the U.S. using a single web tool. Although event logs provide a detailed view into student interactions with ClassTranscribe, the student's intent, motivation, learning context, prior-knowledge, or satisfaction with a search outcome was not measured.

Acknowledgements

We wish to acknowledge UIUC IT staff, NCSA, the College of Engineering, current and former undergraduate and graduate students, Director of Digital Media John Tubbs, and Prof. Hasegawa-Johnson, who have contributed to the development, support and direction of the

ClassTranscribe project. We thank the UIUC IRB review board, the Social Spaces Group led by Prof. Karrie Karahalios, and the CITL Data Analytics Group for their input and support.

The research reported here was supported by a Microsoft Corporation gift to the University of Illinois as part of the 2019 and 2020 Lighthouse Accessibility Microsoft-Illinois partnership, an award from Center for Innovative Teaching and Learning, and the Institute of Education Sciences, U.S. Department of Education through Grant R305A180211 to the Board of Trustees of the University of Illinois. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

[1] Angrave, L., Jensen, K., Zhang, Z., Mahipal, C., Mussulman, D., Schmitz, C. D., ... & Kooper, R. (2020). Improving student accessibility, equity, course performance, and lab skills: How introduction of ClassTranscribe is changing engineering education at the University of Illinois. In ASEE annual conference & exposition.
[2] Angrave, L., Zhang, Z., Henricks, G., & Mahipal, C. (2020, February). Who Benefits? Positive Learner Outcomes from Behavioral Analytics of Online Lecture Video Viewing Using ClassTranscribe. In Proceedings of the 51st ACM Technical Symposium on Computer Science Education (pp. 1193-1199).
[3] Tuna, T., Subhlok, J., Barker, L., Shah, S., Johnson, O., & Hovey, C. (2017). Indexed captioned searchable videos: A learning companion for STEM coursework. Journal of Science Education and Technology, 26(1), 82-99.
[4] Rose, D. H., & Meyer, A. (2002). Teaching every student in the digital age: Universal design for learning. Association for Supervision and Curriculum Development, 1703 N. Beauregard St., Alexandria, VA 22311-1714
[5] Delen, E., Liew, J. and Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. Computers & Education, 78, 312-320.
[6] Defranceschi, A., & Ronchetti, M. (2011). Video-lectures in a traditional mathematics course on iTunes U: Students' feedback. 14th International Conference on Interactive Collaborative Learning, 219-224.
[7] Kim, J., Guo, P. J., Cai, C. J., Li, S. W., Gajos, K. Z., and Miller, R. C. (2014). Data-driven interaction techniques for improving navigation of educational videos. Proceedings of the 27th annual ACM symposium on User interface software and technology, 563-572.
[8] Zipf, G. K. (1935). The psychology of language. Houghton-Mifflin.
[9] Zipf, G. K. (1949). Human behavior and the principle of least effort. Addison-Wesley.
[10] Thorndike, R. L. (1953). Who belongs in the family? Psychometrika, 18(4), 267-276.
[11] Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, 50-60.

[12] Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302), 157-175.
[13] Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association, 46(253), 68-78.
[14] Lillo, F., & Ruggieri, S. (2019). Estimating the Total Volume of Queries to Google. The World Wide Web Conference, 1051-1060.
[15] Cristelli, M., Batty, M., & Pietronero, L. (2012). There is more than a power law in Zipf. Scientific reports, 2(1), 1-7.
[16] Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J. P., & Duerstock, B. S. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. IEEE Transactions on Learning Technologies, 6(4), 299-311.
[17] Broder, A. (2002). A taxonomy of web search. ACM SIGIR Forum, 36(2), 3-10.
[18] Sondhi, P., Sharma, M., Kolari, P., and Zhai, C. (2018). A Taxonomy of Queries for E-commerce Search. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 1245-1248.
[19] Maabreh, M., Al‐Kabi, M. and Alsmadi, I. (2012). Query classification and study of university students' search trends. Program, 46(2), 220-241.
[20] Vidinli, I. B., and Ozcan, R. (2016). New query suggestion framework and algorithms: A case study for an educational search engine. Information processing & management, 52(5), 733-752.
[21] Culbertson, G., Shen, S., Andersen, E., & Jung, M. (2017). Have your cake and eat it too: Foreign language learning with a crowdsourced video captioning system. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (286-296).
[22] Jiang, D., Pei, J., & Li, H. (2013). Mining search and browse logs for web search: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 4(4), 1-37.
[23] Azad, H. K., & Deepak, A. (2019). Query expansion techniques for information retrieval: a survey. Information Processing & Management, 56(5), 1698-1735.
[24] Kim, J. (2016). Organic Crowdsourcing Systems. AAAI Spring Symposia, 268-271.
[25] Kim, J. (2015). Learnersourcing: improving learning with collective learner activity. Ph.D. dissertation, Massachusetts Institute of Technology.
[26] Azzopardi, L., Thomas, P., and Craswell, N. (2018). Measuring the utility of search engine result pages: an information foraging based measure. The 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, 605-614.
[27] Singh, J. and Anand, A. (2020). Model agnostic interpretability of rankers via intent modelling. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 618-628.

[28] Lei, P. L., Sun, C. T., Lin, S. S., and Huang, T. K. (2015). Effect of metacognitive strategies and verbal-imagery cognitive style on biology-based video search and learning performance. Computers & Education, 87, 326-339.
[29] Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, 5(4), 83-124.
[30] Corbin, J. and Strauss, A. (2014). Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications.
[31] Jiang, D., Pei, J., and Li, H. (2013). Mining search and browse logs for web search: A survey. ACM Transactions on Intelligent Systems and Technology (TIST), 4(4), 1-37.
[32] Luo, X., Wang, J., Shen, Q., Wang, J., and Qi, Q. (2017). User behavior analysis based on user interest by web log mining. In 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), 1-5.
[33] Rose, D. H., & Meyer, A. (2002). Teaching every student in the digital age: Universal design for learning. Association for Supervision and Curriculum Development, 1703 N. Beauregard St., Alexandria, VA 22311-1714 (Product no. 101042: $22.95 ASCD members; $26.95 nonmembers).
[34] Rose, D. (2000). Universal design for learning. Journal of Special Education Technology, 15(3), 45-49.
[35] Hall, T. E., Meyer, A., & Rose, D. H. (Eds.). (2012). Universal design for learning in the classroom: Practical applications. Guilford Press.
[36] Izzo, M. V., & Bauer, W. M. (2015). Universal design for learning: enhancing achievement and employment of STEM students with disabilities. Universal Access in the Information Society, 14(1), 17-27.
[37] Black, R. D., Weinberg, L. A., & Brodwin, M. G. (2015). Universal design for learning and instruction: Perspectives of students with disabilities in higher education. Exceptionality Education International, 25(2).
[38] Griful-Freixenet, J., Struyven, K., Verstichele, M., & Andries, C. (2017). Higher education students with disabilities speaking out: perceived barriers and opportunities of the Universal Design for Learning framework. Disability & Society, 32(10), 1627-1649.
[39] Mahipal, C., Angrave, L., Xie, Y., Chatterjee, B., Wang, H., & Qian, Z. (2019, June). “What did I just miss?!” Presenting ClassTranscribe, an automated live-captioning and text-searchable lecture video system, and related pedagogical best-practices. In ASEE Annual Conference and Exposition, Conference Proceedings.
[40] Sheskin, D. J. (2020). Handbook of parametric and nonparametric statistical procedures. crc Press.
[41] Ralabate, P. K. (2011). Universal design for learning: Meeting the needs of all students. The ASHA Leader, 16(10), 14-17.
[42] Corder, G. W., & Foreman, D. I. (2014). Nonparametric statistics: A step-by-step approach. John Wiley & Sons.
[43] National Archives and Records Administration. (2007). The Soundex Indexing System.