UNIVERSITY OF GENOA

PhD in Digital Humanities - DIGITAL TECHNOLOGIES, ARTS, LANGUAGES, CULTURES AND COMMUNICATION

Cycle XXXV

Video Augmentation in Education: in-context support for learners through prerequisite graphs

Thesis Supervisors:

Prof. Ilaria Torre
Prof. Gugliemo Trentin

Candidate:

Ilenia Galluccio

Year 2022/2023

Abstract

The field of education is experiencing a massive digitisation process that has been ongoing for the past decade. The role played by distance learning and Video-Based Learning, which is even more reinforced by the pandemic crisis, has become an established reality. However, the typical features of video consumption, such as sequential viewing and viewing time proportional to duration, often lead to sub-optimal conditions for the use of video lessons in the process of acquisition, retrieval and consolidation of learning contents. Video augmentation can prove to be an effective support to learners, allowing a more flexible exploration of contents, a better understanding of concepts and relationships between concepts and an optimization of time required for video consumption at different stages of the learning process. This thesis focuses therefore on the study of methods for: 1) enhancing video capabilities through video augmentation features; 2) extracting concept and relationships from video materials; 3) developing intelligent user interfaces based on the knowledge extracted. The main research goal is to understand to what extent video augmentation can improve the learning experience. This research goal inspired the design of EDURELL Framework, within which two applications were developed to enable the testing of augmented methods and their provision. The novelty of this work lies in using the knowledge within the video, without exploiting external materials, to exploit its educational potential. The enhancement of the user interface takes place through various support features among which in particular a map that progressively highlights the prerequisite relationships between the concepts as they are explained, i.e., following the advancement of the video. The proposed approach has been designed following a user-centered iterative approach and the results in terms of effect and impact on video comprehension and learning experience make a contribution to the research in this field.

Contents

List of Figures

Chapter 1

Introduction

Reflecting on certain keywords and exploring their etymology can be a fascinating exercise that helps us recognize how their inherent meaning has remained consistent from the outset. To introduce this work, we can examine the etymology of two terms "video" and "education".

What is a video? From Latin "vidēre", literally "to see", "video is the system used to analyse, process, record and, if necessary, transmit images, whether still or moving and whether or not accompanied by sound" [1]. As a medium, video originated in the 1920s with the first experiments on the transmission of moving images, at the dawn of television [103]. This invention was certainly disruptive and led to socio-cultural changes worldwide, breaking through not only in national popular culture, but also with the application of video in art [104] and not least in education. It is quite common to think that in Italy, for example, it was the effect of national television that contributed to the spread of the Italian language in the post-war period. Prior to that time, dialects still predominated in the fragmented and diverse linguistic landscape of the Belpaese [97].

What is education? From Latin "educĕre", literally "to draw out", "the word education means, in its most general sense, the activity of educating, i.e. helping the development of a person's abilities and aptitudes, refining his or her sensibilities, correcting his or her behaviour and transmitting cultural, aesthetic and moral elements to him or her"

---

[1]https://www.treccani.it/vocabolario/video/

[2]. Philosophers, psychologists, educationalists and researchers in general have tried to answer this question over the centuries. There is a phrase that is much abused in quotations in the field of education whose attribution is even doubtful, from Plutarch and Socrates to William Butler Yeats: "Education is the kindling of a flame, not the filling of a vessel". Beyond the doubts as to who really uttered this quote, we are interested in emphasising how it has always been clear that in education, filling the learner with content, following the so-called "scholastic transmissivism" [3], is not always the best way, but that stimulating "the flame" of hunger for knowledge and curiosity is a good alternative to engage learners. From these introductory words, an interesting reflection the Latin etymology of the two terms inspires us, wanting in fact to juxtapose the meaning of education and video we would have that literally they mean "to draw out from seeing". In a broader sense this is what we can call one of the main motivations for proposing this work, that is, by enrolling ourselves in the field of "video education" or "video-based learning" and "video augmentation" (which we will see in detail in Chapter 2) to try to understand how much can be drawn out of videos both in terms of extracting knowledge and in terms of how much videos can draw out knowledge from learners.

1.1   Research Context

The spread of digitisation in education [44] could be associated with the new millennium with the pervasiveness of the Internet and networking. In fact, as we saw earlier, the use of "multimedia" support has been happening since the early 1900s [96, 103]. What the internet has done is to make the phenomenon global, to create interconnection and exchange [43]. Although the experience of learning through video was an well-established reality, suffice it to think MOOCs (Massive Open

---

[2]https://www.treccani.it/vocabolario/educazione$_r$es $- d246dfc4 - dfef - 11eb - 94e0 - 00271042e8d9$/

[3]https://www.gazzettafilosofica.net/2020-1/aprile/la-maieutica-contro-il-trasmisivismo-scolastico-la-prospettiva-di-danilo-dolci/

Online Courses) [13] and the many specialized e-learning platforms, the last three years in which the world globally suffered the effects of the Covid-19 pandemic crisis saw a boost in the spread of educational video, as at some point it was realized that the only way to carry on education of all levels was to take advantage of the digital support [30, 73, 92, 138]. Hence the rapid growth of the phenomenon of "distance learning" [99] and "video-based learning" [135]. Not wanting to focus our attention only on specialized platforms, we could just mention how much each of us is subjected to audio-visual stimuli on a daily basis through social networks (Facebook, Instagram, TikTok etc) [127] and through Youtube. People look for tutorials to understand the "know how" in a wide variety of disciplines and fields. It is no coincidence that TikTok [71], offering only videos, is currently one of the most popular social networks among young people, which with its algorithms based on user preferences manages to keep millions of people glued. In fact, numerous researches are also proliferating on even the possible negative effects related to addictions to it [130]. It is therefore necessary to exploit this attractive and almost hypnotic power that videos can generate by transforming it into a phenomenon that can stimulate knowledge, curiosity and learning about new content. What makes TikTok attractive is the brevity of its videos and an ad hoc construction that engages the user [4]. In fact, one of the limitations in the use of videos in education is that very often they turn out to be unengaging, when we are talking about very long and overly discursive videos [161]. Therefore, the use of interactivity can be considered the key to increasing the potential that the video already possesses but often just needs to be enhanced [52, 62]. All this, of course, must be done without losing sight of the importance of in-depth study; we must not, of course, bow to the "hit and run" culture of one-minute videos, because the task of education is to provide the tools to open the mind and to stimulate reasoning, but to adapt content while maintaining its dignity to a more accessible mode of consumption.

---

[4]https://medium.com/dataseries/how-tiktok-is-addictive-1e53dec10867

1.2   Goal of the Thesis

The aim of the thesis work is to investigate and propose approaches to enhance Video-Based Learning (we will use the acronym VBL from this point on) [135] and in particular interactivity applied to it, which in the literature is expressed with different formulations, "interactive video" [62], "augmented video" [74], "hypervideo"[28, 94, 168], mainly seeking to enhance the implicit knowledge already present within the video by providing additional tools that can help capture it, both from the perspective of the instructor/teacher and from the learner's point of view.

Specifically, the knowledge extracted from videos concerns the concepts explained in the video and their prerequisite relations, intended as precedence relations that specify what concepts are necessary in order to understand further concepts. The key aim was thus to study how to make this extracted knowledge usable by learners, through the design of a video augmentation platform.

The thesis is part of a larger project funded by the Department of Computer Science, Bioengineering, Robotics and Systems Engineering of the University of Genoa [5], entitled EDURELL - EduResource knowledge graph extraction and in-context support to learners. The EDURELL project is focused on the study of methods for extracting concepts and relationships from video materials and the development of intelligent user interfaces based on the knowledge extracted and represented as a temporal semantic knowledge graph, based on the EDURELL Data Model [6]. The EDURELL project came about as a natural continuation of a previous project that involved the extraction of prerequisite maps from textbooks called PRET - Prerequisite Enriched Terminology, within the Technology-Enhanced Learning and Digital humanities (telDh) Programme [6–8].

The idea is therefore to take advantage of all the work already done on annotating and extracting prerequisite maps and identifying relevant

---

[5]https://dibris.unige.it/en
[6]https://teldh.github.io/edurell/

concepts within textbooks, to apply this procedure to videos [1, 2]. Of course, videos, being multimedia, possess a number of constituent elements that make for a composite media as opposed to a written text, with the visual component becoming an integral part of the narrative and which must be taken into account. While textbooks may be more easily "navigable" [22, 23], since by their nature they present a pre-set structure given by tables of contents, page numbering, summaries, list of tables, codes, figures etc., videos do not always present this structuring. This structure is often lacking in university lecture videos, if we refer only to the immense amount of videos produced already only in the 20/22 biennium in the university environment with the adoption of distance learning and lecture recording [118]. Therefore, a system that allows these raw videos to be processed and enriched with additional features suitable for navigating it more easily and providing support to students can be an added value in the current VBL landscape. For this reason, within this thesis work the realisation of a video augmentation platform for learners through the enhancement of the resources already present within the video itself is among the main goals. Suffice it to say that thanks to the Web, external resources such as Wikipedia [7] and DB-pedia [8], many platforms make use of this capability to enhance content, but if for domains where there is a large presence of well-structured materials, there are more niche or specific disciplines that may suffer instead from a lack of external support.

Publications

A number of publications were derived as part of the work done within the frameworks just mentioned. Parts of this thesis are, therefore, based on (or might refer to) the followings:

- Alzetta, C., Galluccio, I., Koceva, F., Passalacqua, S., Torre, I. (2020). Digging Into Prerequisite Annotation. In iTextbooks@ AIED (pp. 29-34).

---

[7]https://www.wikipedia.org/
[8]https://www.dbpedia.org/

- Coccoli, M., Galluccio, I., Torre, I.,(2022). Using Visual Feedbacks in an Augmented Video-based Learning Tool. Proceedings of the 28th International DMS Conference on Visualization and Visual Languages - DMSVIVA 2022 (pp. 77-85)

- Coccoli, M., Galluccio, I., Torre, I., Amenduni, F., Cattaneo, A., Clarke, C. (2022, June). Advanced Visual Interfaces for Augmented Video. In Proceedings of the 2022 International Conference on Advanced Visual Interfaces (pp. 1-3).

- Torre, I., Galluccio, I., Coccoli, M. (2022, June). Video augmentation to support video-based learning. In Proceedings of the 2022 International Conference on Advanced Visual Interfaces (pp. 1-5).

- Mirenda, L., Galluccio, I., (2022). Semantic video annotation tools for educational purposes. Workshop on Advanced Visual Interfaces for Augmented Video. In conjunction with the International Conference on Advanced Visual Interfaces. June 7th, 2022, Rome, Italy.

- Coccoli, M., Torre, I., Galluccio, I. (2023). User Experience Evaluation of Edurell Interface for Video Augmentation. Multimedia tools and applications. Paper under review - second round revision.

1.3   Guide to the Thesis

Given the introductory chapter, the thesis will be structured as follows:

- **Chapter 2**, will explore the literature review on the introductory topic of VBL. Subsequently, the concept of interactive video will be introduced and some of the most significant experiences of hypervideo will be reviewed, namely the basis of the state of the art from which to design our framework. Finally, methods for the automatic extraction of knowledge from videos will be dealt with.

- **Chapter 3**, will focus on the collection of all the issues, previously mentioned in chapter 2 and further elaborated upon, to arrive at

the formulation of the problem statement, the research questions and the methodology used to answer them.

- **Chapter 4**, will present the EDURELL framework, with a description of the design process, the architecture of the tools implemented, namely the video augmentation for education and video annotation applications.

- **Chapter 5**, will describe all the experiments carried out to validate the methods. For the video augmentation application for learners in particular, the heuristic evaluation with experts and the subsequent evaluations with users will be discussed. For the video annotation application, we will briefly look at the validation with metrics for calculating the accuracy of the automatic methods.

- **Chapter 6**, will conclude this thesis project with the final discussion, limitations and possible future scenarios and developments.

Chapter 2

State of the Art

In the following chapter, we offer the reader an overview of the state of the art in specific areas that will be directly touched upon within this thesis project. In particular, starting from the more introductory concept of VBL and its applications in various educational fields, we will introduce the concept of visual feedback and interactivity to support the educational experience through video exploitation. We will therefore review the most significant experiences of video augmentation that constituted the baselines for our project, also analysing their limits and shortcomings. It is important to emphasise that of all the knowledge that can be extracted from the content of the videos for the purpose of augmentation, it was chosen to use the extraction of prerequisite concept maps and to offer the core of support through them. For this reason, in this chapter, we will introduce the topic of concept maps and prerequisite concept maps. Finally, we will explore the topic of automatic methods for extracting knowledge from videos, that we address in the project.

2.1 Video-Based Learning

The use of video as a medium for learning is called, in the literature, Video-Based Learning (VBL) [135]. This definition now seems to be well established. Wanting to go to a higher pyramid level, we could

include video as multimedia. The origins of Multimedia Learning theory can be attributed to Mayer's early work formalizing the discipline in his 2005 handbook entitled "Introduction to Multimedia Learning" [101]. The author, starting from the assumption that *"People can learn more deeply from words and pictures than from words alone"* clarifies the theoretical scaffolding on which multimedia learning theory is based and, above all, clarifies first and foremost the meaning of multimedia [101]. Mayer defines the term multimedia as a mixture of words and images, such words can be either in written or oral form just as the image can manifest itself as an illustration, photo, animation or video. The educational stage through multimedia occurs when the individual is able to construct knowledge through words and images [132]. Hence, the use of audiovisual media in education has a long history throughout the 20th century. Suffice it to say that as early as 1968, the two researchers Maclean and Roderick wrote an article on the use of television in education in Britain [96]. However, technological progress and the spread of the web have certainly given a boost to the use of video as an educational medium. The experience of MOOCs (Massive Open Online Courses), which gave distance education and VBL a more structured connotation, is now more than twenty years old and already well documented in numerous reviews from the 2010s [60, 173, 174]. VBL [135] is therefore currently a more than consolidated reality, which has however gained even more centrality when, due to the Covid 19 pandemic involving the 2020/2022 period, distance learning and video support have had an "almost obligatory" proliferation at a global level [73, 145] considering the "ante-covid" phase as one in which the blended approach was favoured [46, 67]. Recent studies have worked on the detection of the benefits of using VBL in many domains [31, 172], in different educational levels [4, 148], for an improvement of learning outputs [109], engagement and in particular cognitive engagement [155]. The usage of videos in the educational field, due to the pandemic exponential growth, has been analysed using large amounts of data to analyse student feedbacks. Last researches show that video-based learning positively fits into

the student's perception, also taking into account the variables of gender and digital inequality [118]. In the work cited above [155], Taskin et. al., exploiting the so-called Nudge Theory [83] from the world of behavioural economics and cognitive psychology, combine the benefit of nudging within the context of VBL to analyse an improvement in cognitive engagement, not just stimulating activity but also favouring elaboration. Although the advantages of VBL can be recognised by various studies, there are a number of issues relating to the characteristics of video itself, in the era we are living the use of content is no longer seen as "passive" (as it was for television), the new generations [57], accustomed to the use of YouTube, Netflix and other similar platforms, want to have an active role, as a so-called "prosumer" and not only "consumer" [24, 111] for this reason the linear viewing of audiovisual content can be seen as limiting even with respect to the engagement. Numerous approaches have been experimented in research to overcome this limitation, such as the possibility of building environments for collaborative annotation [33], the presence of quizzes to verify ongoing learning [81], the adoption of interactive annotation to encourage soft skills learning [107].

2.1.1 Features, Advantages and Limits of VBL in Education

In [135] the use of video support in education is analysed through a systematic review covering 2008 to 2019. In more than a decade, technologies have evolved considerably, what emerges is that VBL offers numerous possibilities with regard to the student engagement, for improving learning outcomes, as a feedback tool and for teacher development too. In order to be effective, the literature shows that an educational video should take into account the Cognitive Load for the learner, consider student engagement, and encourage active learning [20], as displayed in the pyramid (figure 2.1). Another underdeveloped strand is also to what extent videos can stimulate questions with respect to the topics discussed and formulate hypotheses, in a vision that is therefore not only
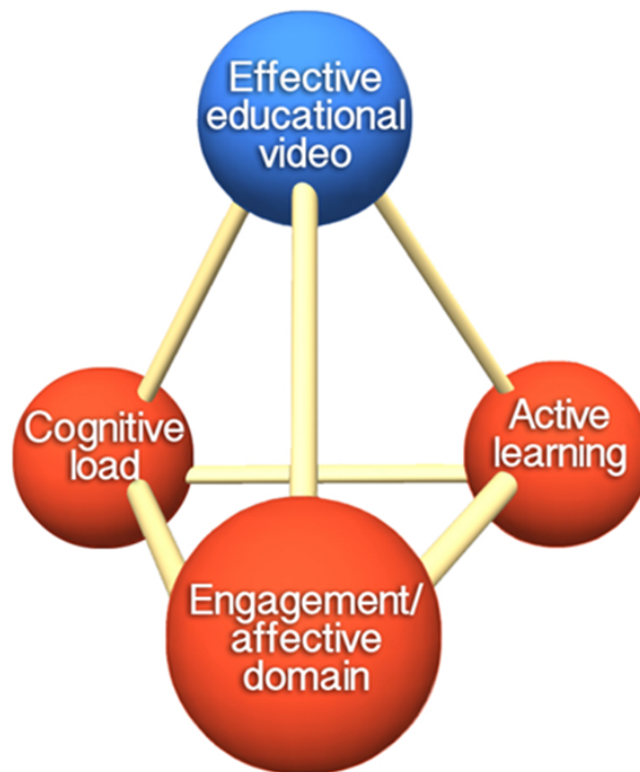
Figure 2.1: Educational video pyramid (from [20])

passive in an evaluative manner but in a more critical perspective [69]. Among the most debated issues for VBL we will look at some of them in more detail, namely the design of videos according to the Cognitive Load Theory, the problem of student dropouts especially with regard to the enrolment in online courses (such as MOOCs) and also the so-called visual clutter concerning the exposure of the individual to visual interfaces and the reduction of it in terms of usability.

The Cognitive Load and Overload of VBL

The Cognitive Load Theory has been formulated by Sweller [149] [150] and based on the assumption that human memory is composed of several components: 1) sensory memory, which is stimulated from the outside (audio/visual stimuli); 2) working memory, which processes information and has a limited capacity; 3) long-term memory, which has unlimited capacity and allows the storage of information selected from the working

Load Reducing Method	Description
Off-loading	presenting words as narration
Segmenting	segmenting the information in small chunks
Pretraining	providing instruction concerning the system components
Weeding	reducing extraneous information that does not contribute to the learning objective
Signaling	using of text or symbols to highlight important information.
Aligning	inserting printed words in correspondence with graphic elements
Eliminating redundancy	eliminating repetitions between oral and written words
Synchronizing	associating narration and corresponding animations
Individualizing	ensuring that the learner has all the necessary skills to acquire all the information

Figure 2.2: Cognitive Load reducing methods

memory [20].

Mayer and Moreno, in 2003, study the phenomenon of Cognitive Load applied to multimedia learning, highlighting the possibility of the risk of what they call "cognitive overload" which occurs when the total intended processing exceeds the learner's cognitive capacity' [102]. The possibility of this occurring in the case of videos is therefore quite common, considering that with poor design of an educational video, the student may be subjected to an excessive amount of stimuli such that the desired cognitive attainment is not achieved. For this reason, Mayer and Moreno identify various strategies to reduce the risk of Cognitive

Load [102], that are summarised in figure 2.2.

Dropouts in VBL

One of the major issues with VBL, particularly for MOOCs, is the high levels of course dropouts [128]. By now, given the decades of experience with this type of distance learning, the amount of data is quite substantial to be able to carry out numerous analyses. Many studies have tried to investigate the causes, also trying through predictive models to prevent it [70, 88]. What emerges in this study [12], based on questionnaires suggested to students, is that beyond personal causes (such as lack of time or other life priorities) among the most frequently cited reasons for dropping out seems to be excessive video length, lack of feedback and support from instructors. The problem of dropout could be linked to the the topic of student engagement, which has held sway in the field of VBL for a long time. Numerous studies have been carried out to understand the criteria for student abandonment of video viewing, low engagement, etc. One of the most important and at the same time very simple and intuitive guidelines is that the video should be short. Studies have shown that the ideal length is 6-9 minutes.[19, 133]. At the same time, the use of supporting tools that can make navigation interactive can foster an active learning experience and encourage engagement.

Visual Clutter in VBL

Always connected to the topic of cognitive load, there is another issue that concerns the visualisation of screens, especially when we talk about videos, interactive videos and visual interfaces in general, namely visual clutter. In the literature, this is defined as an excess of information and elements on a screen that can create a chaotic visualisation that can distract the user, thus causing an incorrect pursuit of the intended goal [49]. For designers and information visualisation experts, it is therefore important to measure visual clutter to try to reduce it [134]. Visual clutter is to be kept in mind when designing visualisations such as graphs

and concept maps, in EmotionCues [66] the researchers have realised a visual analytics system to monitor students' emotions during lessons, and to alleviate visual clutter they propose a stream graph with different layers, similarly in [176] they remove components with a high number of connections from the graph by creating a "separated component area".

2.1.2 Visual Feedbacks to support VBL

The feedback is an information provided by an external agent about performance or understanding [64], its pedagogical validity is widely tested in various sectors of education [169]: 1) computer-assisted pronunciation training (CAPT) for delivering feedback in both auditory and visual modalities for a new language learning [18], 2) technology-mediated feedback for enhancing dynamics in higher education piano learning and performance [63], 3) in-video visual feedback to improve learner's attention in MOOCs [144], 4) real-time augmented visual feedback of airflow in voice education [98]. If in traditional education this feedback can be transmitted by a human agent, in the case of VBL it is important that the learner is not left alone during the use of contents. Intelligent visual interfaces (IVI) [100] can make up for this typical lack of solo viewing. Immediate feedback through intelligent interfaces technologies provides new strategies to overcome limits for people with functional disabilities [55], to support symmetrical collaborative educational writing among visually impaired and sighted users [167] or more simply to improve the level of activity of studying a specific discipline [80]. Technologies for video augmentation and hypervideo (HV) [94] can be included in the aforementioned IVI ecosystem.

2.2 Interactive Videos in Education

Although the dissemination of video in everyday life and also in education is a widespread practice and partially formalised by the definition of VBL [135], when we refer to interactive and supportive modalities to complement the typical consumption of video content there is still

no standard vocabulary unambiguously shared by the entire scientific community. Nevertheless, in the literature a lot of studies regarding interactive videos [62] in order to improve usability and navigate educational contents with more control and flexibility have been conducted [91]; in a broader sense interactivity in the world of education has been classified by numerous scholars, starting with Moore [108] who had identified three levels, in [9] will later be added further levels that can be summarised as follows:

- Learner-Content Interaction;

- Learner-Learner Interaction;

- Learner-Instructor Interaction;

- Learner-Interface or Technology Interaction.

The work of Schoeffmann et al. [139] tries to classify video interaction into seven types: Video Annotation, Video Browsing, Video Navigation, Video Editing, Video Recommendation, Video Retrieval and Video Summarization. If, therefore, the interactivity represents the 'big picture', different ways to refers to interactive video are used in the literature: for example video augmentation [156], video enhanced learning [52] or hypervideo [175].

2.2.1   Hypervideo from a Technical Perspective

The definition of hypervideo (HV) was not formalised from the outset [94], but as early as 2004 Zahn et. al. identify the HV as a *"combination of digital video and hypertext, which draws largely upon audiovisual media as central parts of their structure. They consist of interconnected video scenes containing 'dynamic' hyperlinks that are available during the course of the video scenes and that refer to further information elements (such as texts, photos and graphics)"* [175]. Currently, to define HV we can refer to a series of affordances for interactivity; Cattaneo et. al [25–27, 122, 136] identify a number of characteristics of HVs, emerging from the analysis of the existing literature, they are:

- **Dinamism**, a feature that provides a moving image view, already typical of video, but which in HV is especially crucial for the learning phase of, for example, procedures, practical knowledge, etc [26].

- **Control features**, refer to the possibility of watching a video in a non linear manner [26].

- **Hyperlinks**, that are clickable markers to provide external or internal reference to additional information [26].

- **Individual video annotation**, gives the possibility to add personal notes [26].

- **Collaborative video annotation**, when the system allows to receive feedback from other people or even from the system too [26].

- **Quiz**, that provide immediate feedback on content learning [26].

Improving the navigation experience has been the subject of numerous projects, such as data-enhanced transcript search and keyword summary, automatic display of relevant frames, a visual summary representing points with high learner activity [75], non linear consumption of videos using personal fragment navigation [162], exploration of e-learning contents via small screens [125].

2.2.2   Hypervideo from a Domain Perspective

The use of video, as seen in section 2.1, and HV are widely adopted in different domains and levels of education, but there is more literature and, therefore, application of these tools in scientific disciplines, so-called STEM and in medical and vocational disciplines. In health care, the work of Meixer et. al. [105, 106] focuses on both medical and physiotherapy training. In particular, the physiotherapy application consists of the implementation of a dual screen tool (handheld and TV) in order to view designed HVs to help patients correctly perform their exercises, with the aim of adapt their training to their daily needs and training progress. Within the STEM disciplines, many experiences

are found in literature: mathematics, statistics [153], physics. For example, in [42] the assessment of interactive video to enhance learning has been tested on applied physics. Applications of HV in mathematics can be found in numerous researches, in the works of the early 2000s, an issue related to the communication of mathematics seems to emerge mainly, so [28] published a work focusing on the communication and learning of mathematics through HV, while a few years later they focus on the special aspects of visualisation [29]. A recent study (2022) confirms that the use of HV in particular in the study of mathematics improves student performance [38]. Another interesting aspect of HV application in different domains is related to 'know-how', i.e. the vocational field. What emerges from the studies mainly by Cattaneo et. al [25, 27, 122] is that students are able through the use of HV to have a greater focus of attention and in particular to make this link between theoretical knowledge and practical knowledge.

2.2.3   Hypervideo from an Instructional Perspective

Within the educational field, the HV tool is often associated with the purpose of learning and student oriented, nevertheless there are also positive implications from an instructional point of view and thus seen from the perspective of the teacher. Already in the construction of the educational pathway, teachers can benefit from the features of HV through the principle of nonlinear linkage for the change from linear learning culture towards a self-directed learning. During the creation of video fragments teachers can construct educational content in which the possibility of hyperlinks, for example, can provide interdisciplinary tools and insights from the web, moving towards an approach that students already normally adopt since the Internet has become pervasive in society [168]. Another advantage of using HV platforms is not only in the support phase of content creation, but also in the subsequent phase, i.e. the analysis of student behaviour in order to be able to track engagement, participation, and common problems or misunderstandings. Within these instruments instructors can collect information on their

students and improve their teaching [147]. HV can support different instructional scenarios, as extensively discussed in [27] [26] and promote, for instance, different types of approaches at the level of educational strategies that can relate to active learning [141].

2.3 Hypervideo Features and Platforms

There are numerous platforms developed in the field of HV and video augmentation, oriented towards the educational sector. One of the first attempts to realise a HV platform took place back in the mid-1990s with HyperCafe [137].

In this section, we will specifically look at a comparison with some platforms developed in recent years, in order to gather a set of desiderata that are indispensable for a quality HV platform. The literature analysis, which in its various forms was presented in this state of the art, started with a search on "Google Scholar" for combined keywords such as "video augmentation" and "education" or "hypervideo" and "education". For a more recent overview, we decided to focus on the four-year period from 2018 to 2022. The only exception outside this time span is the LectureScape project [75], which is an authoritative example of a video augmentation platform and a starting point and baseline for a comparison with more recent attempts. The search yielded around 80 results for the "video augmentation" and around 250 for "hypervideo", that were skimmed in various ways: by disambiguating the search, which often in the case of video augmentation leads to results relating to augmented reality projects in the educational field, eliminating reviews, non-English or Italian language articles or articles from conferences/meetings of recognised quality, and articles that did not present a platform development project or a prototype from scratch, eliminating articles in which keywords are not the main theme of the paper. The result is the presentation of these followings projects that represent the most significant starting point for us. As mentioned in the previous section, there are features that can characterise a HV or video augmen-

tation platform, which emerge from the literature and which, in this work, we group and categorise in this way:

1. **Interactive timeline**:

   - *toolbar*, a typical or augmented timeline. The starting point is a toolbar with options to play, pause, adjust volume, etc.

   - *user traces*, by tracking user activity, already viewed parts of videos are highlighted on timelines.

   - *segmentation*, a segmentation in clickable chapters.

   - *bookmarks* the addition of bookmarks with a view to customisation by the user.

   - *hotspot*, clickable points on the timeline. Their function can have various purposes: interaction peaks, markers, concept occurrence, etc.

2. **Enhanced search**:

   - *interactive transcript*, a transcript of the video speech, possibly interactive and clickable.

   - *concept search*, a concept-based search of the individual topic within the video by means of: highlights within the transcript, clickable word cloud, hotspots on the video frame, search bar etc.

   - *hyperlinks*, within the interface, they can have various functions: referring to external links, to other videos in a learning path, to specific points in the video, etc.

3. **Enhanced cognitive tools**:

   - *summarization* automatic or manual, it is the result of processing with respect to the content of the video.

   - *concept maps*, automatic or manual, it offers a schematisation and hierarchization of the video content.

- *annotation*, it allows customisation and the addition of information by the user.

4. **Exchange options**:

   - *notes*, it allows to take notes on video content, in a individual or collaborative way.

   - *forum*, it allows to exchange comments between users.

   - *quiz*, it allows to test and receive feedback about the video content understanding.

5. **Overlays**:

   - *augmented reality*, it allows to add an interactive experience that combines the real world and computer-generated content.

   - *360° visualization*, it allows to add another possibility of immersivity with a 360° live action video.

6. **Navigation affordances**:

   - *index or table of content*, a typical textbook-like subject index, in this case the contents of the video.

   - *fragment navigation*, clickable portions of video to allow non-linear navigation of the video.

   - *custom path*, it allows the user to choose the preferred learning path of the course, which is not necessarily linear.

7. **Extra**:

   - *external materials*, through hyperlinks, markers, etc., extra content in relation to the video is clickable and available.

These affordances are the result of a comparative analysis of the main platforms, which we show in detail in the 2.3.1. In 2.5, we see how the tools analysed do or do not cover the main feature package typical of a HV.
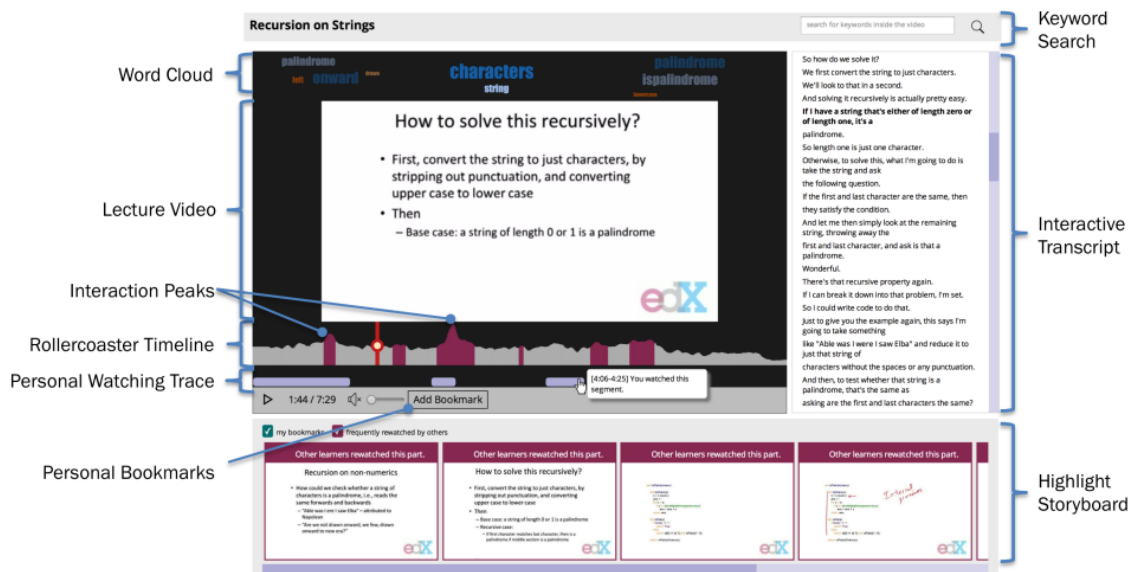
Figure 2.3: LectureScape, a prototype Web-based video interface

LectureScape

In the aforementioned project [75] an interactive approach techniques to augment a traditional web video player has been proposed.

They introduce data-driven techniques that process, visualize, and summarize interaction data generated by learners watching the same educational video. On a methodological level, the motivations behind the implementation of the platform are based on a set of learner needs that can be summarised in these "watching scenarios":

- **Rewatch**: the user needs to watch the video again, especially to search for some specific definitions within it.

- **Textual search**: the user needs to 'jump' directly to when a specific concept is first mentioned.

- **Visual search**: the user wants to watch a specific video part associated with a concept explained within it.

- **Return**: the user wants to return to a previous point in the video, such as a specific slide.

- **Skim**: the user prefers to watch video content in a more summarised and non-linear manner.

These user needs and the related watching scenarios are addressed through the LectureScape platform (figure 2.3 shows its interface) where a series of video augmentation features assist the user in navigating the content. The features related to the first category in figure 2.5 are all present within the platform:

1. **Interactive timeline**:

   - *toolbar*, a "rollercoaster timeline" that expand the video timeline and visualize collective learners interactions.

   - *user traces*, track of which parts of the video they personally watched.

   - *segmentation*, segmentation in this case is designed as portions of videos already viewed by the user to keep track of his or her navigation.

   - *bookmarks*, to add and offer further personalisation to the user than the so-called "collective traces".

   - *hotspots*, "interaction peaks" as points in a video with significantly high playbutton click activity, to generate highlight frames of a clip.

2. **Enhanced search**:

   - *interactive transcript*, a transcript that runs simultaneously with the progress of the video.

   - *concept search*, the concept search is conceived as the keyword search within the transcript text.

   - *hyperlinks*, by clicking on them, students can choose topics of interest.

3. **Enhanced cognitive tools**:

- *summarization*, a word cloud displays automatically-extracted topics for the section of the video, providing a keyword-based summarization of the video.

4. **Navigation affordances**:

- *fragment navigation*, a highlight storyboad that displays the peak frames and allows the learner to visually scan the video's progress.

LectureScape is one of the most significant projects from a data-driven perspective. As early as 2014, in fact, the authors raised the issue of the biases that could arise from the massive use of data (in this case interaction data) in automatic mode, thus also proposing customisation as a tool that can compensate for the creation of visual feedback without the possibility of "human" revision. A hybrid approach is therefore also hoped for in iMOOC [131], in this case however starting from the reverse problem.

CAIV

Classroom augmented interactive video (CAIV) combines AR technology with the interactive video's affordances for personal and collaborative learning interactions [74]. Different types of interactions are foreseen within the platform, declined with different purposes whether it is student, class or instructor. Here we only consider interactions addressed to the individual student:

1. **Interactive timeline**:

- *toolbar*, a toolbar with all the typical features, in particular with captions which personalize video experience.
- *user traces*, this feature makes it possible to identify parts of videos not yet viewed or to review those of interest.

2. **Enhanced search**:

- *hyperlinks*, by clicking on them, students can choose topics of interest.

3. **Enhanced cognitive tools**:

   - *summarization*, a summary is proposed according to the students' selections, alternatively, students can also create their own summary from scratch.

   - *annotation*, mobile devices are used by students to add information to the common video projection screen.

4. **Enchange Options**:

   - *notes*, shared user notes and comments, on the scanned frame of the video, students are asked to insert their notes and comments.

   - *quiz*, by using their mobile devices, students are able to answer questions and receive ad-hoc personal feedback.

5. **Overlays**:

   - *augmented reality*, overlays on students' mobile devices displays AR application, these information can be adapted to the students' characteristics and preferences.

6. **Navigation affordances**:

   - *table of content*, global navigation that provides the option to access specific points in the video of special interest to learners.

   - *personal path*, students can choose a different learning path to the one on the shared screen.

7. **Extra**:

   - *external materials*, supplementary material for learners.

One of the most interesting issues that emerges from this project, and which is an open issue, is how very often the student in front of an online platform finds himself doing self-regulated study, despite the fact
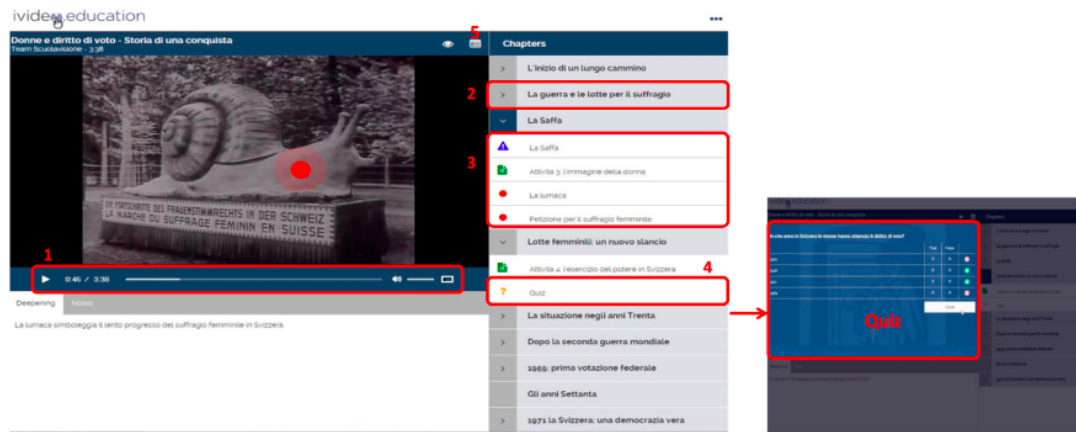
Figure 2.4: iVideo, interface and features

that there is also the possibility of using the classroom environment. Therefore, it is all very well to support the student in individual use, by means of visual feedback, but a "human" support component, as we shall also see in the iMOOC project in section continues to be of high value in the context of VBL.

iVideo

The iVideo.education hypervideo interface [26] (see figure 2.4), through a study of existing literature and referring to the design principles already mentioned in the sub-section 2.3.1, shows the main affordances of HV:

1. **Interactive timeline**:

   - *toolbar*, with all the options of a typical toolbar: play, pause, stop, adjust volume, etc.

   - *segmentation*, a segmentation into chapters, which can be clicked on by the user.

   - *hotspots*, called "markers" or "active points", they are clickable and apper during the video consumption to give anchor points to other resources.

2. **Enhanced search**:

- *hyperlinks*, hotspots that temporarily appear during video consumption to offer links to other (also external) resources.

3. **Enhanced cognitive tools**:

  - *annotation*, in this project, annotation is conceived as individual or collaborative video annotation, which can also be mapped in the "notes" section.

4. **Exchange options**:

  - *notes*, individual and collaborative "notes" section allows the user to take notes on video content, visualize comments of other contributors and save them in PDF format.

  - *quiz*, multiple-choice questions to test understanding of the portion of the video and receive immediate feedback.

5. **Navigation affordances**:

  - *table of content*, the index is mappable with segmentation into chapters.

  - *personal path*, connected to the segmentation into chapters, allows the video to be easily navigated in a non-linear manner.

6. **Extra**:

  - *external materials*, supplementary material for learners.

The most significant issues that emerge from this experience relate to always open questions about to the use of interactivity. First and foremost, an evaluation of the true effectiveness for learning is needed, especially in different scenarios, i.e. the "learning contexts" that we will also have to face in our project. Secondly, a major open issue continues to be that of cognitive load, and here reference is also made to socio-cognitive load with respect to interactivity.

iMOOC

I-MOOC [131] is based on the HV technology and on teaching activities that are not bound to specific sequences of use. The aim of I-MOOC is therefore to offer participants autonomy in deciding when and how to use the content, through various levels of interactivity: with materials, peers and instructors. I-MOOC is a database and incorporates several videos, interactive and non-interactive, according to two levels of interactivity: (a) with each individual video and (b) with the database that hosts and manages the videos. The video lectures has been realised in different formats: live-action (video shooting), animation, mixed. In the first level of interactivity, the video lessons offer different types of interaction:

- video-lessons with annotations at key moments of the video.

- video-lessons with insights.

- immersive video lectures, realised with 360° live-action video.

The second level of interactivity concerns the different ways of navigating the contents, their organisation and visualisation, favouring a not necessarily linear view of the course.

Thus, in the schematisation carried out in figure 2.5, the features of iMOOC are summarised as follows:

1. **Interactive timeline**:

   - *toolbar*, with all the options of a typical toolbar: play, pause, stop, adjust volume, etc.

2. **Enhanced search**:

   - *hyperlinks*, links with supporting materials.

3. **Exchange options**:

   - *forum*, interactivity with tutors is ensured both by a direct channel (Ask the tutor) and by the presence in the thematic forums frequented by the participants.

4. **Overlays**:

   - *360° visualization*, immersive, 360° live-action video lectures.

5. **Navigation affordances**:

   - *personalized path*, there are different navigation modes: according to the main axes of the model; according to the temporal development of the model; by exploring the psychosocial context of the model's application.

The platform's limits mainly lie in the management of the forum and requests to tutors, which is why the ideal evolution of the platform assumes the need for the integration of automatic response systems, e.g. through chatbots.

HyperButton

HyperButton [76] is an interactive video platform that supports collaborative question answering via in-video buttons and hyperlinks. Compared to the other projects, HyperButton starts from a different principle more related to video annotation platforms, which due to the vastness of the existing literature will not be explored in detail in this thesis. The aspect that interests us, however, is the hypervideo features and visualization and how they support the consumption of educational video. HyperButton has three main design goals: a) enhancing interactivity of video learning content; b) enabling collaborative questions and answering; c) supporting video content design based on learner generated material. Within the video player interface learners can add questions and hyperlinks by clicking on the video contents. Question and hyperlink are added and the emoji are displayed on the screen. A prompt appears if learners click anywhere in the video, and learners can input a question and hyperlinks for specific spot in the video. Mapping this features on figure 2.5 HyperButton has the following affordances:

1. **Interactive timeline**:

- *toolbar*, a toolbar with typical options for play, pause, volume adjustment, etc.

- *hotspots*, through the use of emoji hotspots appear in various places on the screen.

2. **Enhanced search**:

- *hyperlinks*, through the system of hotspot hyperlinks could be displayed on the screen.

3. **Enhanced cognitive tools**:

- *annotation*, individual and collaborative annotations can be collected.

4. **Exchange options**:

- *quiz*, it allows learners to collaboratively generate questions and related learning resources for the lecture and share them.

5. **Extra**:

- *external materials*, links to external content or knowledge generated by the collaborative action.

With a view to assessing the issues arising from each individual project, focusing more on a visualisation problem, even in a collaborative platform with the possibility of video annotation by learners, the main problem also relates to visual clutter, i.e. if the number of emoji gets larger, given the absence of a filter, displaying all of them could generate visualization problems. There is also no quality control system for questions and hyperlinks.

2.3.1   A Comparative Analysis between Platforms

In this review of HV platforms, it was proposed to compare only a few of them that might have affordances of interest related to the visualisation part of HV features. As mentioned above, very often interaction types

merge into a single application. Current experiences, also on a commercial level [1], include a large customisation component on the part of the user with the possibility of annotating the video. In this comparative analysis, therefore, we start with Kim's platform [75], which, although a little older, certainly still represents a very up-to-date example of a complete interface with all the typical features of a HV platform. With iVideo [26], the "exchange options" component is represented more explicitly, unlike LectureScape, which provides the addition of bookmarks as a level of customisation. With CAIV [74] and iMOOC [131], the innovative components are related to the experiences of integrating augmented reality and 360° videos. In fact, the research trend at this particular time is in the direction of the integration of 360° videos [14, 158]. An element of interest in terms of visualisation is to be found in Hyperbutton [76], which is also the most recent experience mentioned in the previous section, which exploits a visualisation component typical of social networks, namely emojii, and makes it its own in order to stimulate the student with tools that are familiar to her/him even in contexts other than the educational sphere. In figure 2.5, we see schematically the comparative analysis with all the affordances that characterise the various platforms.

2.4 The Use of Concept Maps in Education and in VBL

A concept map can be regarded as an enhanced cognitive tool to support the learner throughout the learning process. Formally they are a node link diagram, consisting of nodes and arcs, where the nodes are the concepts and the arcs represent the type of relationship existing between the two concepts [140, 160]. Starting with the pedagogical theories of Ausebel [10, 11], who with "meaningful learning" and "advance organisers" lays the foundations for a type of learning that is not sterile memorisation and where the role of concept maps and hierarchi-

---

[1]https://www.annoto.net/

Category	Feature	Platform				
		1	2	3	4	5
Interactive timeline	Toolbar	■	■	■	■	■
	User Traces	■				
	Segmentation	■		■		
	Bookmarks	■				
	Hotspots	■				■
Enhanced search	Interactive transcript	■				
	Concept Search	■				
	Hyperlinks	■	■	■	■	■
Enhanced cognitive tools	Summarization	■	■			
	Concept maps					
	Annotation		■	■		■
Exchange options	Notes		■			
	Forum				■	
	Quiz			■		■
Overlays	Augmented Reality			■		
	360° visualization				■	
Navigation affordances	Index/Table of content		■	■		
	Fragment navigation	■				
	Personalized path		■	■		
Extra	External materials		■			■

ID	Platform
1	LectureScape
2	CAIV
3	iVideo
4	iMOOC
5	HyperButton

Figure 2.5: Comparison between platforms

cal, structured representations of knowledge can help the learner, Novak [115, 116] formalises concept maps as a tool to help actively construct knowledge.

Concept maps [61], have been applied in various scientific domains [39, 47, 129, 159] and in different levels of education (from primary school to university) [15, 17]. Even in the context of students with special needs and specific learning disorders [35, 51], concept maps can provide valuable support. Several projects propose interactive solutions [45], from the time-anchoring and editable concept map to facilitate video browsing and provide metacognitive support [90] to semi-automathic concept map generation [65].

2.4.1 Prerequisite Concept Maps

Prerequisite concept maps [164, 165] (in the figure 2.6 we can see an example) can be regarded as a particular type of concept map where the dependency relationship between the two nodes, i.e. the concepts, is represented by directional arcs in which one concept is prerequisite to another. From a cognitive point of view, therefore, a prerequisite concept represents the notion y that the learner should know in order to be able to learn concept x [1]. These properties are identified in the literature [1]: (1) binary relation: it involves pairs of concepts; (2) antireflexive relation: concept x cannot be a prerequisite of itself; (3) transitive relation: if $x \prec y$ and $y \prec z$, than $x \prec z$; (4) antisymmetric relation: if concept $x \prec y$, than $y \prec x$ must not hold.

Based on these characteristics, therefore, prerequisite concept maps are defined as acyclic [1, 8]. There is a vast literature on prerequisite relations that, starting with the theoretical studies of Gagné, who introduces the concept of learning hierarchies [53, 54] later become of crucial importance in the field of instructional design of educational content, such as learning objects [58, 59], and in educational technologies in general for structuring and organising knowledge. In fact, a number of practical applications can exploit the potential of prerequisite relations [8] such as the construction of domain ontologies [37, 77], sequential structuring
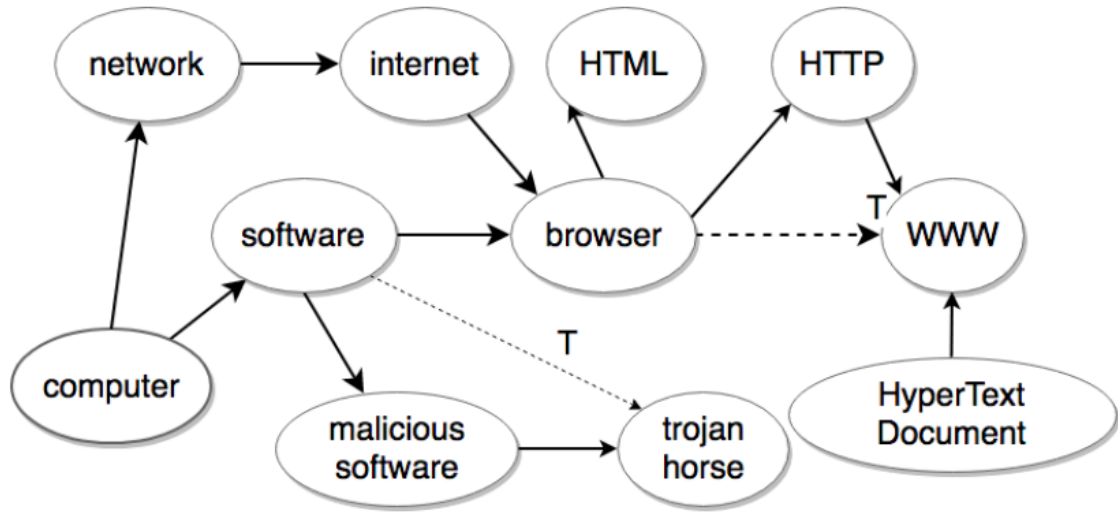
Figure 2.6: Prerequisite concept maps example

of courses and MOOCs [3, 89, 166] , automatic content creation [95], learning path recommendation [2, 154].

The creation of prerequisite maps follows annotation protocols and guidelines [7, 8, 50, 86], and as for the more generic concept maps it is an operation that requires a high degree of specialisation on the part of the annotator, also becoming a very time-consuming and demanding activity especially when it concerns the annotation of entire manuals and educational books or in the case of video courses and recorded lectures. For this reason, in section 2.5 we introduce the topic of automatic knowledge extraction from educational videos and in sub-section 2.5.2 we specifically explore the automatic extraction of prerequisite maps.

2.5 Automatic Knowledge Extraction from Videos

Considering that the creation of video content is already a time-consuming task in itself, the creation of additional media, such as video fragments, concept maps, transcriptions, etc. is even more of an onerous task for the teacher. In this sense, therefore, methods of automatic knowledge extraction from educational videos can provide crucial support for in-

structors. In this section, we mainly look at techniques for video segmentation and the creation of prerequisite concept maps.

2.5.1 Video Segmentation

The topic of non-linear viewing of a video is extensively dealt with in a very interesting study by the researchers Verma et. al.[161]. To favour a more dynamic viewing of videos, the giant YouTube also allows manual segmentation into "chapters" when the author publishes a video.[2] This manual operation, however, can be time-consuming, which is why working on automatic segmentation techniques is one of the research strands of absolute interest in recent years. The main goal of video segmentation is to obtain meaningful and self-consistent shots of video [79]. These issues have been of interest for more than two decades now, but with the refinement of machine learning techniques, the automation of the process and the quality of the final product is constantly evolving [56]. In the literature numerous methods and approaches are used: fragmentation based on slide detection [48, 72, 93, 113, 143], segmentation by automatically analyzing the synchronized slides exploiting the Optical Character Recognition (OCR) technique [34], modeling transition time and multimodal cues [142, 178], analysis of the change of concept in domain knowledge graphs [41]. In [161] NoVoExp technique generates multimodal fragments, grouping sentences (from video transcript) based on semantic similarity in the BERT embedding space. Exploiting the textual part of the video to carry out a segmentation is a type of approach that has been declined in various ways and numerous works: in [32] they built a model that extracts textual and acoustic features from speech and uses them to identify topical segment boundaries of the lecture video; in [5] they integrate Natural Language Processing (NLP) tasks to extract key linguistic features exist within the video; Soares et. al. have been worked on a method for automatic segmentation of scenes in educational videos through the use of automatic audio transcription and semantic annotation [146]; this research group [40] proposes a sys-

---

[2]https://support.google.com/youtube/answer/9884579?hl=en

tem that performs topic wise semantic segmentation and annotation of MOOC lecture videos to mitigate user effort using as input a lecture video, the speech transcript, and list of topic names. As we have already seen with the paper cited above [162], another approach to achieve video segmentation is to exploit the visual part; there is a vast literature on this, e.g. in [177] they take advantage of an HSV (Hue Saturation and Value) histogram to detect MOOC videos scene transitions. Educational videos have a different standard from generic videos, visual patterns can be detected more easily by image analysis. Within e.g. MOOCs, we can find various combinations of speakers and slides, present at the same time in the video area or at different times. Based precisely on these patterns, there is a strand of video segmentation related to slide detection [179]. A survey, at the beginning of the new millennium, already reviewed the various techniques for achieving so-called temporal video segmentation [79], in this more recent review there is a classification of video segmentation techniques identified as follows [121]:

1. Colour feature extraction

2. Text based segmentation

3. Shape based segmentation

4. Background/Foreground subtraction

5. Temporal (frame) differencing

As extensively analysed in this survey, each technique obviously has advantages and disadvantages, and assessing the type of teaching material in order to be able to carry out a quality segmentation, perhaps with a hybrid method, is an important starting point.

2.5.2   Prerequisite Concept Maps Extraction

As we mentioned earlier, prerequisite concept maps represent a specific type of concept map. They can be constructed e.g. by the teacher/instructor, but require a great deal of effort, whereas educational videos or

textbooks are presented with a large amount of minutes or pages. The process of automatically inferring prerequisites relations from learning materials can facilitate the instructor's task in constructing maps. A lot of research has been done to extract prerequisite relations from texts, while there are few attempts to extract these relations from MOOC videos. One of the most frequently used approaches is the search for lexical patterns, such as hyponyms and hypernyms [165], which in addition to being applied to texts can also be used for the transcription of an educational video.

Always using only the potential of linguistic analysis, without external knowledge support, there are several approaches for identifying prerequisite relations from the MOOC's transcript: for example [119] extracting [119] a list of contextual features such as 1) *semantic relateness* namely the cosine distance of two concepts in a vector space, 2) *video reference distance*, which represents how significant a concept is by measuring the term frequency in all the related video, 3) *sentence reference distance*, which is similar to the previous principle, but on the sentence granularity. In most cases, techniques for prerequisite relation extraction exploit external structured representation of knowledge, such as Wikipedia [151] [165]. Another approach, the RefD, again based on the external wikipedia reference, using the reference distance metric, models the relation by measuring how differently two concepts refer to each other. In [1] , the proposed method of extraction is based on Burst intervals and co-occurrence. The principle behind Burst intervals is that there are textual intervals within which certain concepts are more prominent, thus also combined with the principle of co-occurrence of two concepts within the same interval, the extraction of pairs of concepts that are prerequisite to each other is realised.

To summarize, in this chapter, an overview of the state of the art of the main themes of the doctoral thesis work is proposed. In particular, the focus has been on giving an introduction to the literature on VBL, which is at the heart of the project, with a critical approach

that has led to emphasising the advantages of using video in education, but also highlighting the limitations related to the problem of cognitive overload, dropout and the issue of visual clutter during video consumption. The theme of interactivity applied to video and in particular with so-called HVs was declined from a technical, domain and instructional point of view, then going on to give a series of practical examples of HV applications in the educational sphere, which also constitute for us the baseline for the development of an original app in the EDURELL Framework that we will present in chapter 4. We introduced the prerequisite concept maps, which in fact represent one of the main augmented methods proposed in our project, and finally the automatic extraction methods and in particular the video segmentation and extraction of the aforementioned prerequisite concept maps.

Chapter 3

Issues and Research Goals

3.1 Issues arising from the Literature

In this chapter we will focus on relevant issues in VBL emerging from the literature. In particular, we will consider issues that impact the learning experience of students but also issues that impact the provision of augmentation tools for learning. Thereafter, from the issues identified and our own experience, we will formulate the problems we aim to address with this research and the related research questions.

3.1.1 Issues regarding Video Consumption

As we have extensively discussed in the state of the art analysis, although the use of video in education is widely recognised as well as its advantages and also limitations (subsection 2.1.1), there are issues emerging from the literature with regard to the inherent limitations of videos as learning tools. In [161], the scholars focus on the crucial issue of *linear consumption*, and this passage in our opinion clearly explains the rationale for a non-linear view of educational videos: *"While videos are known to deliver an engaging and immersive experience, some of them, especially instructional and explanatory videos, are often long and require the viewer to spend time in proportion to the duration of the video. This is often sub-optimal for viewers interested in only specific parts of the video as it requires them to spend more-than-required*

time in skimming through the parts that are not of interest. The current video consumption format does not allow viewers to easily find the specific parts that they are interested in." [161].

Hence, the most limiting issues in the video consumption phase can be summarised by the following points: 1) limits in sequential consumption; 2) required time almost equal to the video length; 3) poor structure and no index of content; 4) poor navigation and retrieval of concepts.

In the case of well-designed educational videos that follow the guidelines of instructional design principles [68], this problem can be limited. However, during the last three years, due to the Covid 19 pandemic crisis, we have witnessed a proliferation of video production, very often involving hour-long recordings of lectures by academics. A system that could allow "raw" videos to be processed and given a structure could help in the exploitation of such educational materials.

On the issue of video-to-speech, almost all platforms, including commercial ones, provide subtitling services, transcription, either to be entered manually by the user or extracted automatically. Very often, however, these transcripts, navigable and interactive they may be, still have a level of granularity that does not allow for the identification of relevant concepts, especially of a specific domain, which highlighted and with the possibility of also having a mapping of synonyms detected within the same text, can provide further support to the student. On the other hand, as far as video segmentation is concerned, which acts as an index of the content and allows a not necessarily linear navigation of the video, in order to reach a specific portion of the video, we still observe limitations and difficulties in offering this particular type of service.

In general, we observe that videos can be difficult to navigate, especially when there is a need to recall specific concepts, they often lack content tables or indexing systems to explore it.

3.1.2 Issues regarding Video Augmentation Methods

The issues concerning video augmentation methods can be divided into two areas, the first concerns the production of augmented content and the issue of time and energy expenditure on the part of the teacher/instructor, the second concerns open issues concerning interactivity per se. We see these in detail in the following sections.

Production of Augmented Methods

The creation of multimedia educational content, especially video content, is certainly an onerous task on the part of the teacher/instructor. Even in the design of a MOOC [117], for example, there are a number of operations to be taken into account in order to produce effective content. On the success of this design phase may then also depend the effects on study performance and persistence [171].

If we delve into the HV ecosystem even further [26], the tasks to be performed in order to provide the video with a series of augmented methods that can facilitate the use of educational video are time-consuming. Among the various methods identified in the literature that characterise HVs (figure 2.5), the creation of an internal index of the video, possibly linked to a video segmentation, requires manual pre-processing by the teacher/instructor on the entire video or on a series of videos, with the creation of individual chapters (as is also the case on YouTube, for example).

Another highly time-consuming operation is, for instance, the identification of concepts [163] and the creation of concept maps extracted from the implicit knowledge of the video. Especially if we refer to a large amount of videos of varying length, manually constructing lists of concepts and identifying prerequisite relationships between them is an additional task for the instructor.

Other issues emerging from the literature concern the topic of extracting prerequisite concept maps. Assuming that external resources offer a

trustworthy representation of the domain knowledge, the majority of approaches for prerequisite concept maps extraction rely on external structured representation of knowledge, such as Wikipedia [58, 152, 164]. In order to identify which relations of the reference knowledge graph match to required relations, these approaches try to do so. External knowledge graph can be exploited for different purposes, for instance for filter out relevant keywords in the resource that do not match concepts from external graphs [120, 164], or to calculate the strength of the relation between pages connected by a Wikipedia link [87], but if a domain is not well covered by external resources, these approaches cannot be used.

Exploitation of Augmented Methods

The survey of video augmentation experiences in section 2.3 reveals a number of issues that are still open to research. With regard to one of the oldest projects, namely LectureScape [75], questions were already being asked in the mid-2000s about the problem of bias in data-driven projects with large amounts of data, with a reflection that links us back to the previous section and leads us to question the extent to which automatic methods of knowledge extraction are supportive in reducing the problem of time-consuming content creation, but on the other hand by the need to also have a hybrid and semi-automatic approach to allow for a speeding up of the process but also for quality control on the part of the human.

Another issue mainly noted by Cattaneo [26], is that of the effectiveness of the interactivity itself as a support along the learning process, especially in the different "learning contexts" which as we shall also see in our experiments will be the starting point for the design of the evaluations.

In the area of information visualisation, consideration must also be given in the design of a video augmentation platform, which by its very nature has a number of additional features that enrich the video player, as to how to limit the levels of distraction that interactivity and excessive elements can cause. In Hyperbutton [76] or EmotionCues [66], these

issues are also brought to light and also represent a constraint in the design of a tool.

3.2 Problem Statement and Research Question

Problem Statement

The main problem that emerges from the analysis of the state of the art is that the inherent limitations of video lessons can in some cases reduce understanding and engagement, even causing dropouts, as seen in the literature review. Video augmentation methods, such as content enhancement and browsing support, could help the learner's educational experience, but the effort to build augmentation tools is high, with the consequence that their use is limited. In addition, the effectiveness of the augmentation methods is not straightforward and depends on several factors, including in particular the learning context, i.e., the learning phase and the goal of the learner related to the activity of video consumption.

Research Question

Given the problem stated above and the specific issues described in the previous section, the main objective of this thesis is to define methods to improve the use of educational videos. More specifically, we are interested in video augmentation methods that exploit the information and knowledge included in the video. The objective is reflected in the following research question:

RQ: To what extent video augmentation based on knowledge extracted from the video can improve the learning experience?

The approach followed to address this research question, has been:
- identifying what knowledge can be extracted from videos and used for learning support;

- building a framework for knowledge extraction and video augmentation;

- testing the learning experience and video comprehension using augmentation methods in different learning contexts, in different domains.

3.3   Contribution of the Thesis

The most significant contributions of the thesis are:

- the study and design of visual interfaces for education from a HV perspective. Specifically, the study has been carried out starting from the analysis of the platforms considered as baselines and presented in the previous chapter. Through the realisation of use cases, scenarios and mockups, the various proposals of visual interfaces have been defined. Using a user-centered iterative approach, through numerous experiments, including an expert evaluation and iterative evaluations with students in two different domains, the video augmentation platform has been realised and made available online.

- the visualisation and realisation of prerequisite concept maps as semantic temporal knowledge graphs integrated within the video augmentation platform.

- the study and design of manual, semi-automatic and automatic annotation methods. The study started from the analysis of methods implemented for the automatic extraction of prerequisite concept maps and domain concepts from textbooks, the readaptation of these methods led to the implementation of methods for automatic extraction from educational videos supported by specific video annotation platform.

To summarize, in this chapter, an attempt was made to extrapolate the main issues emerging from the literature review (Chapter 2). In particular, issues relating to video consumption related mainly to the

theme of linear viewing of very long videos without content structure, and issues specific to video augmentation. On this topic, the issues for production of augmented videos as task time consuming and those of exploitation, i.e. the effectiveness of interactivity in the educational context, were underlined. Ultimately the problem statement, the research question and the specific contribution of this thesis made in response to them have been defined.

Chapter 4

The EDURELL Framework

In the previous chapters, the state of the art has been outlined and the resulting issues identified, which process brought to the definition of the research question for this thesis.

This chapter presents the design of the EDURELL Framework, that is part of a larger project[1] and that includes specific features designed to address the research question concerning the investigation of methods to enhance learning experience and video comprehension in VLB by exploiting the knowledge and content extracted from video lessons.

The EDURELL Framework - EduResource knowledge graph extraction and in-context support to learners - is focused on the study of methods for extracting concepts and relationships from video materials and the development of intelligent user interfaces based on the knowledge extracted and represented as a temporal semantic knowledge graph. Within the EDURELL framework two tools have been developed in order to test the proposed methods: 1) The *Edurell Video Augmentation tool (EVA)*; 2) The *Edurell Video Annotation tool (EVAnn)* [2].

---

[1] The design of the framework is part of a larger project funded by the Department of Computer Science, Bioengineering, Robotics and Systems Engineering of the University of Genoa, for more information: https://dibris.unige.it/Progetti_SEED.

[2] The repository of both apps can be found on github to facilitate experimental sharing by other research groups: https://github.com/Teldh/edurell.

The thesis project aims to facilitate the experience along the learning path of a student watching an educational video by taking advantage of video augmentation features and visual feedbacks.

As a matter of fact, educational videos are created by experts to give learners an accurate selection and order of educational concepts while highlighting pertinent connections between topics. As a result, learning resources can be seen as a knowledge graph, where concepts are nodes and relationships are edges. Prerequisite relationships illustrate the most efficient learning paths to acquire a thorough understanding of a subject. As we have seen in chapters 2 and 3, existing strategies for augmentation and the extraction of prerequisite maps often rely on external knowledge and structured resources available online rather than the knowledge embedded in the resource itself. Even though they are effective, those systems fall short when applied to a different domain or when the subject matter is not sufficiently covered by outside information. For this reason, within the EDURELL Framework, the type of support and video augmentation features only come from the reworking of content already present within the video itself, which may sometimes remain more implicit and which will instead be enhanced through knowledge extraction work (manual and automatic).

In this chapter, the architecture will be explored in depth.

4.1 Architecture

The two previously mentioned tools are highly interconnected, as the annotation tool is designed for the preparatory phase of the HV functionalities exploited in the video augmentation tool.

EDURELL Video Augmentation tool - EVA

The EVA tool has been developed with the intention of supporting learners in the use of educational video contents through the HV functionalities.

The aim is to improve the experience along the learning path, receiving support during the browsing and consumption of educational videos, by taking advantage especially of concepts maps, video fragments, enhanced interactive transcripts. Figure 4.1 shows the architecture of the two tools which share the same cloud database, MongoDB Atlas. In particular, EVA uses Python Flask for the back end and React technology for the front end part.
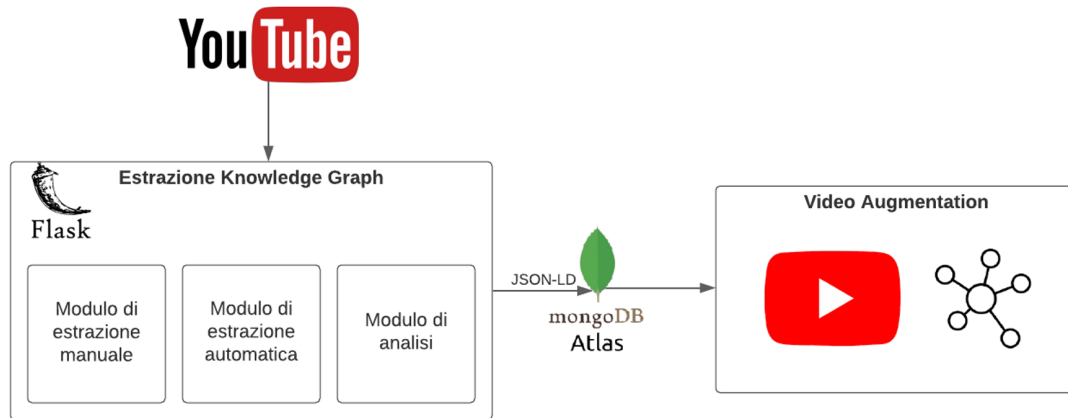
Figure 4.1: Architecture of EVAnn and EVA tools

EDURELL Video Annotation tool - EVAnn

The EVAnn has been designed with the objective of processing educational videos and annotating them through three possible methods: manual annotation, automatic and semi-automatic knowledge extraction in order to generate a knowledge graph that is then exploited to generate intelligent user interfaces in EVA.

In the current version, the educational videos are acquired from the YouTube platform, which as a matter of ease of manipulation seemed at a design stage to be the most suitable for the pilot project and chosen as a baseline in the experimental tests with EVA. The intention in future developments is to be able to extend the functionality also to the manipulation of videos from other sources.

The extraction of the prerequisite map, which can be done manually,

semi-automatically and automatically, plus the identification of the list of concepts and their descriptions, is processed using the methods described in [157], through Python Flask, which enables the organisation of the extracted contents and storage in the MongoDB database. The front end part in this case was handled with Javascript and Bootstrap.

4.2   EVA Application Design

As was already said, educational videos are among the most often used forms in online learning, but they can also have a number of limitations. For instance, when a student needs to recall specific concepts, they can be challenging to navigate.

They typically do not have explicit tables of contents or other indexing techniques for browsing the video content, unless it is the author of the video who enters the division into chapters as is the case, for example, on YouTube [3]. In this framework we studied different methods to address such limitations, tested through the EVA tool using the knowledge graph that was derived from the EVAnn.

The design of the EVA tool followed an iterative procedure, because it is not a linear operation but one that involves continuous adjustments along the way, including through evaluations that we will see in Chapter 5. Firstly, the analysis of the literature in the field of video augmentation and hypervideo provided us with the baselines from which to identify the relevant features. As we saw in section 2.3, what we were primarily interested in was to include features within the tool (see figure 4.2) that could meet the requirements and cover as much as possible the set of options we have summarised in figure 2.5.

---

[3]https://support.google.com/youtube/answer/9884579?hl=it
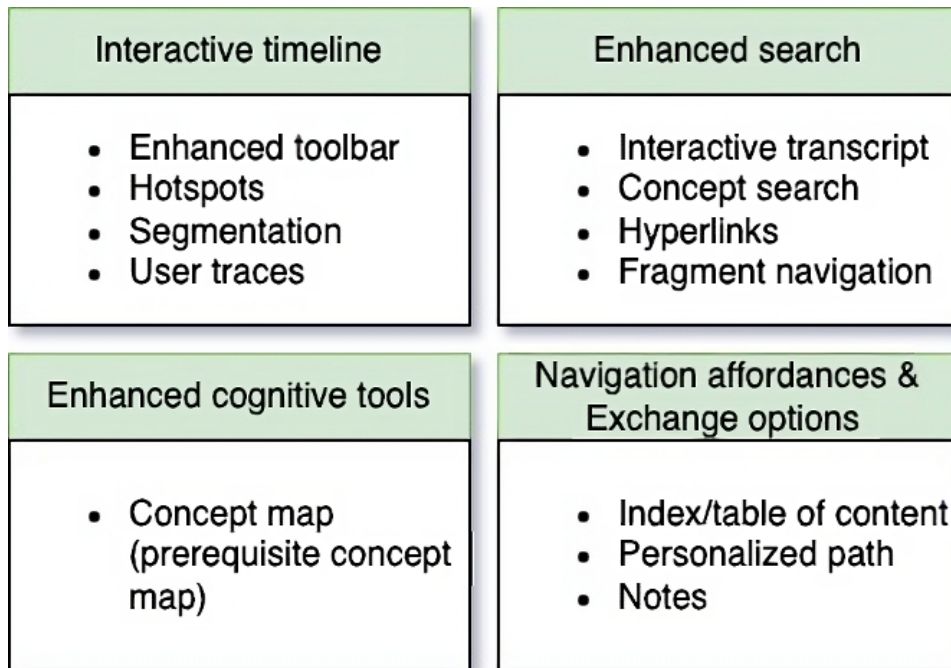
Figure 4.2: EVA affordances

From the typical affordances identified in the literature on the basis of which we classified the various platforms in the survey in section 2.3, we drew up a list of options that we would like to have within our tool, so as to be effective for the use of video augmentation methods to be tested in the experiments. In the next section we will look at the descriptions of the various interfaces, here instead we try to highlight what motivated the design choices.

Looking at figure 4.2, the first block refers to the category "Interactive Timeline". Within it, a number of features are designed within the platform:

- *enhanced toolbar*, the idea is to have an augmented toolbar, which can track concept-oriented granularity with a series of anchor points directly linked to the concept map of prerequisites. The need arises from the issues identified in subsection 3.1.1 regarding the difficulty of navigating the videos in the task of recalling specific concepts.

- *hotspots*, hotspots in EVA tools, unlike other experiences in the literature [26, 75] that use them as links to external resources, high-

light the presence of a concept at a particular point in the video. To give a further indication of the type of mention of the concept, we designed a visual feedback that allows to specify whether it is a description or an in-depth of the concept in the video.

- *segmentation*, the segmentation that is automatically produced in the pre-processing of the video is not integrated directly in the main toolbar, but in a section immediately below. This feature is of crucial importance from the perspective of non-linear video viewing [161], as it allows individual video fragments to be circumscribed and acts as a table of contents, commonly used in textbooks. The list of concepts within the fragment allows the user to have a different level of granularity compared to concept-oriented search.

- *user traces*, user traces in EVA are conceived as a fundamentally tool for keeping track, in a non-linear consumption [161], of which portions of the video have already been viewed by the user and which have yet to be viewed. They are directly linked to segmentation, in the section below the toolbar. This feature keeps track of all user logs, and when the user accesses the same video again, he or she will have track of his or her viewing rates of the fragments.

Enhanced search takes place in different modes:

- *interactive transcript*, the transcript is a useful tool to support the student while watching the video. Simultaneous scrolling allows the user to constantly keep track of the speech, but in addition to subtitling, the transcript allows the user to click on sentences of interest in order to be able to browse the entire speech in an enhanced search perspective. Moreover, with the highlighting of relevant concepts, it is also possible to increase the capability to recall concepts, which is one of the major issues identified for the limits of the video consumption (3.1.1).

- *concept search*, concept search is not only about one feature within the tool, but concept-oriented search can be performed through var-

ious strategies supported by the different features of the platform. The concept is highlighted in the transcript, present as a node in the concept map of prerequisites (which are clickable) or can be carried out through a search bar in which the user types the name of the concept of interest. The idea is precisely to support the user in finding specific concepts that the complexity and length of an educational video can be difficult to find, recall, if an internal structure is not well highlighted. This would lead the user to have to jump from one point to another in the navigation bar, resulting in wasted time.

- *hyperlinks*, hyperlinks are one of the most important features in the characterisation also provided in the literature of a video augmentation platforms [26]. In EVA, they are designed to be present in various places on the interface, we consider hyperlinks the nodes of the conceptual prerequisite map, which allows hotspots to appear on the toolbar. The concepts and sentences highlighted within the transcript are themselves clickable.

An Enhanced cognitive tool in EVA is:

- *concept map*, as we have seen in the literature [47, 61, 129], concept maps can be a valuable support for the learner throughout the learning process, it is placed in the category of enhanced cognitive tool because it involves a more complex reworking of the contents of the video. In our system the concept map is also not conceived as static, linked to prior domain knowledge and already constructed beforehand by the teacher/instructor, but rather as a map of concepts' flow, dynamic that expands and becomes more complex as the video scrolls in synchro. Since concepts evolve along with the video flow (as their explanation deepens), they are initially presented with lower complexity resulting in a context help which shows a simple knowledge graph. During the video consumption, the same concepts are deepened with other concepts that are explained. It falls into the category of the prerequisite map that we introduced in Chapter

2. The idea is to offer a contextual help for a concept that the lecturer is explaining, showing the prerequisites that the student has to know in order to get the current concept. The peculiarity of the map is also that it is directly anchored to the video to facilitate navigation as well.

The Navigation Affordances and Exchange Options (as defined in the previous chapters) that characterise the EVA system are as follows:

- *index/table of content*, the index or table of contents of a video works similarly to the book. It gives more internal structure to the video by subdividing it into "paragraphs", which in the case of videos are segments. The title of the individual segment and the list of concepts contained within it, in combination with the preview (with the keyframe of the segment) completes the feature set that serves as the index of the video.

- *fragment navigation*, closely linked to the concept of segmentation, fragment navigation allows for the non-linear navigation of portions of the video, to facilitate the user in moving from one point to another of the video while having a clearer understanding of its internal structure. In fact, as pointed out in the issues (subsection 3.1.1), this is one of the major shortcomings that videos usually have. By means of a preview of the keyframe and a title of the fragment, the user receives additional information about the content of the portion of the video, especially in the rewatch phase, he can more easily find the parts he needs to review.

- *personal path*, in EVA, is conceived in the context of segmentation, whereby thanks to the clickable segments the user can choose to have a non-linear view of the video, moreover with the presence of a percentage of completion of the segment view the user will always have an indication of which parts of the video have not yet been viewed.

- *notes*, within the classification proposed in section 2.3, notes are included as one of the possibilities for exchange options if annotations

and comments from other users are also available in the system. Although this is a goal for the future with notes from all users that can be collected and processed, in the current version of EVA the notes are personal and saved in the user account who will find a trace of them every time they enter the video.

In the feature mapping of section 2.3, there are two additional categories "Overlays" and "Extras", which were not taken into account in EDURELL. Concerning overlays our project did not originally include the use of 360° video, augmented reality, etc., this does not detract from the fact that in future developments these pluses can be incorporated. As far as the use of external materials is concerned, the choice was deliberate, because unlike numerous projects in which external support is mainly exploited (e.g. Wikipedia) [58, 152, 164], in this case the aim is precisely to augment the video by exploiting only the knowledge extracted from the video itself or with a view to inter-video anchoring, to other videos linked to the same course (in a MOOC perspective), for the reasons already explained at the beginning of this chapter and the issues identified in subsection 3.1.2, in particular with regard to the scarcity of resources in specific domains.

As a second step in the design, once the desiderata had been gathered, all design documents has been produced in order to begin modelling the system on the basis of the various use cases. In this stage the use cases that will be described in the next section (4.3) have been defined.

As a third step during the design phase, in which several mockup versions were realised on Figma platform (figure 4.2), a questionnaire has been submitted to 4 PhD students in Digital Humanities in order to receive feedback on the choices to be made in the following steps. The procedure was therefore as follows: a prototype, has been assembled, thanks to the functionality on Figma that allows the various screens realised during the mockup to be linked, producing a video recording that simulates the various operations that can be performed on the
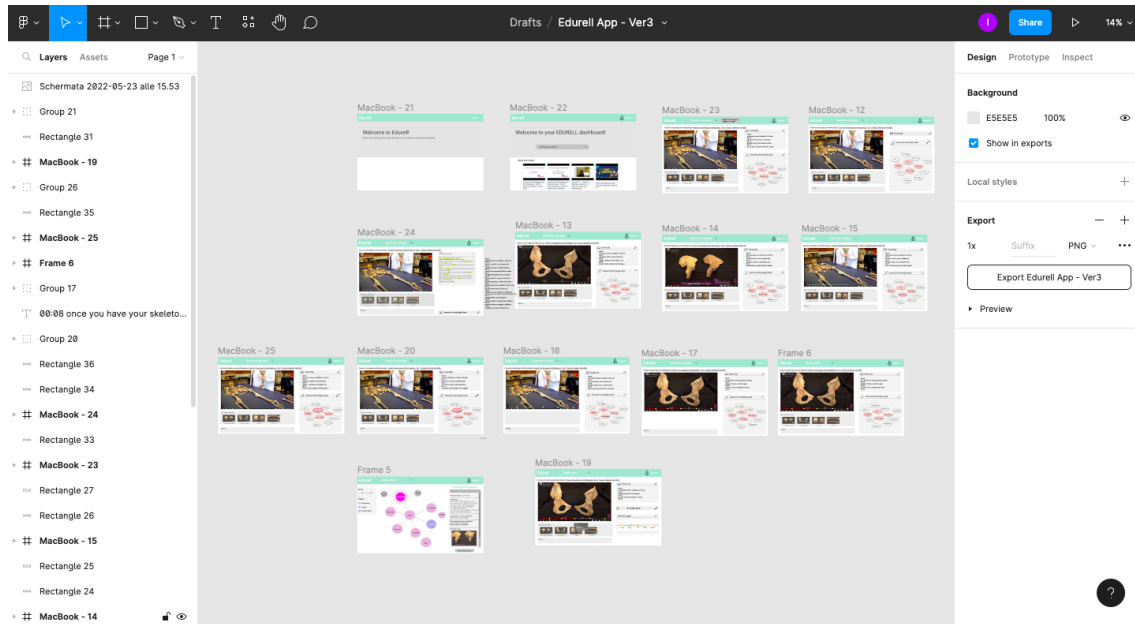
Figure 4.3: EVA mockup on Figma

interface (see figure 4.2).

At this point, after watching this video, the students were given a questionnaire, with some open answers to gather feedback or closed answers when they needed a specific suggestion in the design of the interface. The graph shows for example the result for the question *"Based on your experience what is the clearest way to define the Map?"*, to which all students responded with a preference for the option *"Map of the concepts' flow"*, which was then actually entered as the wording to identify the map within the EVA interface. With regard to open-ended answers, for example to the question *"In your opinion what is the Map useful for?"* the majority of the answers was *"To describe the relation of concepts"*. This comment confirmed to us that the main objective of the map visualisation had been grasped even at this low-fidelity prototyping stage. The feedback collected from this preliminary evaluation was used to refine the design choices and the use cases and scenarios (see section 4.3).

The next steps of the iterative design (heuristic evaluation with ex-

Figure 4.4: EVA prototype on Figma

Figure 4.5: Pie chart showing the result of the feedback question during the prototyping phase of the EVA app.

perts and three evaluations with end users) will be described in Chapter 5.

4.3   Use Cases and Scenarios

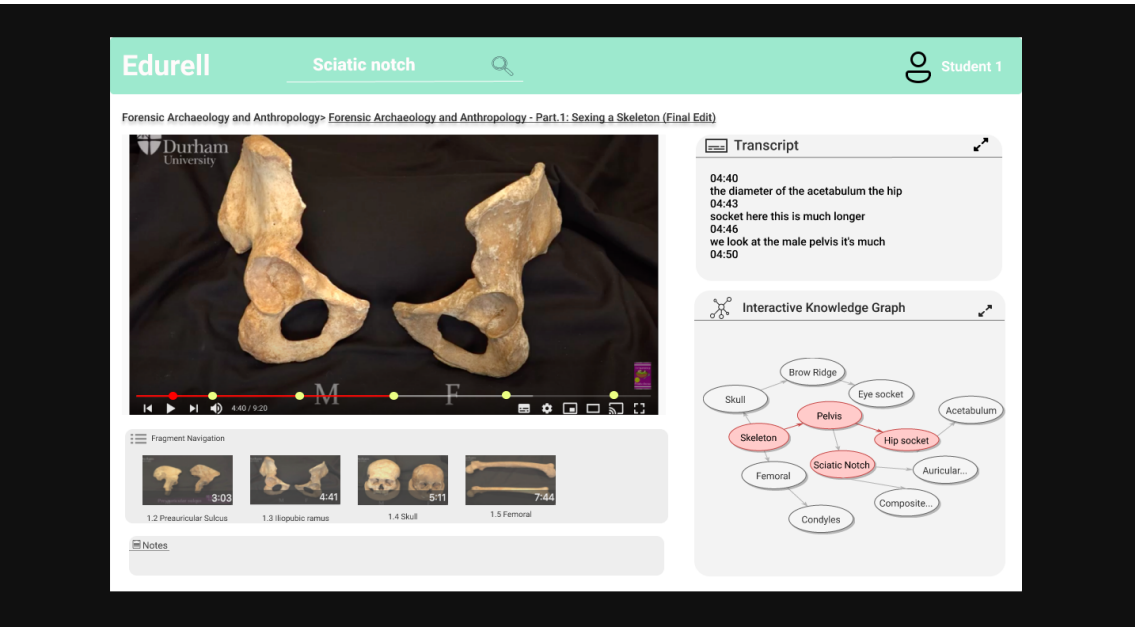We designed a set of possible scenarios to be proposed to the typical user.   Two different learning contexts (first-watch and rewatch) were defined for our study.   The first-watch occurs when the student first approaches watching an educational video, in the rewatch context the student avail himself/herself of the video again with a distance we set at 2 days. Within the two types of learning contexts, 4 tasks were identified during the video consumption event:

- exploring

- searching

- analyzing

- annotating

Within the figure 4.6, the two different learning contexts and the corresponding tasks are described in detail with reference to each individual feature of the tool.

The use cases were formulated based on a comprehensive training course offered by Durham University on the Future Learn platform and also available on Youtube. The title of the course is "Archaeology and Forensic Anthropology". The choice of domain and course type fell on archaeology because in general is strongly characterised by multidisciplinarity (biology, geology, history, art history) which makes it particularly interesting for the purpose of the project, because, despite being classified as a humanistic discipline, it also contains formalised and structured knowledge typical of the "hard sciences". Another important feature, specific to the course, is the visual part that involves a series

Visual Feedback	Learning Contexts	Tasks	Description
Augmented Transcript	First Watch	exploring	The user during the first viewing phase of the video, through the transcript receives support during the content exploration phase.
	Rewatch	exploring	The user receives support during the rewatch phase of the video by means of the transcript when exploring the content, thanks to the possibility of clicking on a single sentence to jump from one point to another in the video.
		searching	The user during the rewatch phase of the video receives support through the transcript when searching for the specific concept in order to watch the portion of the video in which it is explained.
		analyzing	During the rewatch phase of the video, the user receives support through the transcript to be able to analyse and rework the content.
Concept Map	First Watch	exploring	The user during the first viewing phase of the video, through the concept receives interactive support during the content exploration phase.
	Rewatch	exploring	During the rewatch phase, the user can navigate and interact with the map in order to carry out his research with respect to a specific concept.
		searching	During the rewatch phase, the user can search and interact with the map in order to carry out his research with respect to a specific concept.
Fragment Navigation	Rewatch	exploring	During the rewatch phase, the user can take advantage of fragment navigation to rewatch a specific portion of the video, without the need to jump from one point to another on the navigation bar.
		searching	During the rewatch phase, the user can take advantage of fragment navigation to search for a concept, thanks to the pop-up window indicating the concepts within the video portion.
Concept Search	Rewatch	searching	The user searches for the concept via the concept search bar.
		analyzing	The user uses the in-depth tab of the concept search to fix the contents of the concept and annotate them in the box.

Figure 4.6: Methods for video augmentation: learning contexts and tasks

of images shown by the lecturer to explain the anatomy of the skeleton. These features of oral explanation and direct laboratory images meant that for the first design of a use case this course could be the most suitable, among those available in the scope on the Youtube platform.

4.4 Description of the EVA tool

EVA is designed along the lines of the most popular e-learning platforms (such as Coursera [4] or edX [5]), so following a user registration phase and a customised dashboard (figure 4.7), from which the user can also trace his or her personal history that takes into account previous explorations, he or she will find the entire catalogue searchable with a bar. Once the user clicks on the video of interest, all video augmentation features are presented in the same screen (figure 4.8), in a compact manner, to

---

[4]https://www.coursera.org/
[5]https://www.edx.org/

avoid scattering and distraction while watching the video (this particular aspect will be tested in the evaluation in Chapter 5).

The user can then view the content with the video player imported directly from Youtube and the related features, such as the possibility to play/pause, activate or deactivate subtitles, set the playback speed, video quality, etc. In addition, EVA tool offers video augmentation features, which we see in detail in the following sections, namely: 1) an interactive transcript; 2) a map of concept's flow; 3) a fragment navigation; 4) a concept search.
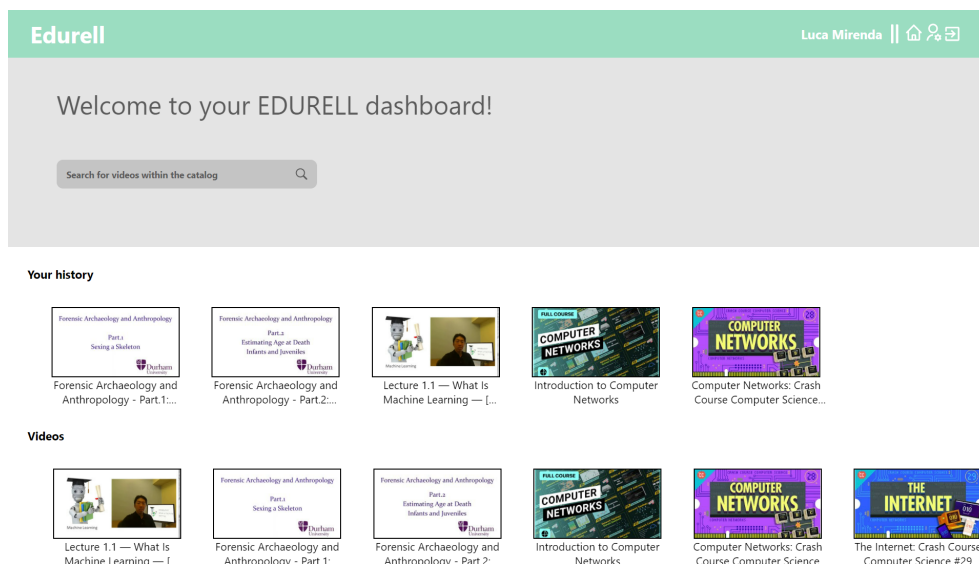
Figure 4.7: EVA dashboard
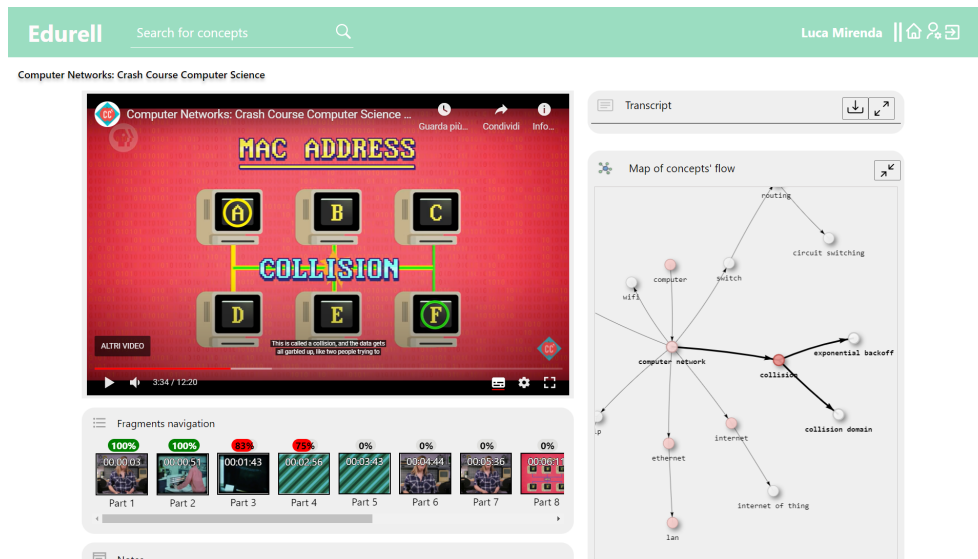
Figure 4.8: EVA interface

4.4.1 Augmented Transcript

The video transcript feature is widely spread and used in the world of video consumption, both in commercial and education-specific platforms. Transcript support is useful for a number of reasons, if the video is in a language other than the mother tongue, if the language is overly technical and the learner may find him/herself listening to domain terms not commonly used, acronyms, etc. The advantage of the video transcript is that in its interactive mode, it runs in sync with the video and is clickable, so that with direct anchoring the learner can also jump from one point to another to review unclear concepts or specific parts of the video. If the transcription of the video is not added manually by the author, subtitles can be generated automatically through the use of neural networks and language models.

In EDURELL, this process of transcript generation and/or enrichment also includes the automatic extraction of keywords to highlight the main concepts within the educational video. In EVAnn, in section 4.5 we will see how the list of concepts can in any case be edited manually, as the synonyms.

The augmented transcript (figure 4.9), which is displayed on the EVA

interface, is in a box on the right-hand side and can be zoomed in or
out if necessary so that the user can choose to view both the transcript
and the concept flow map, or only one of the two media.
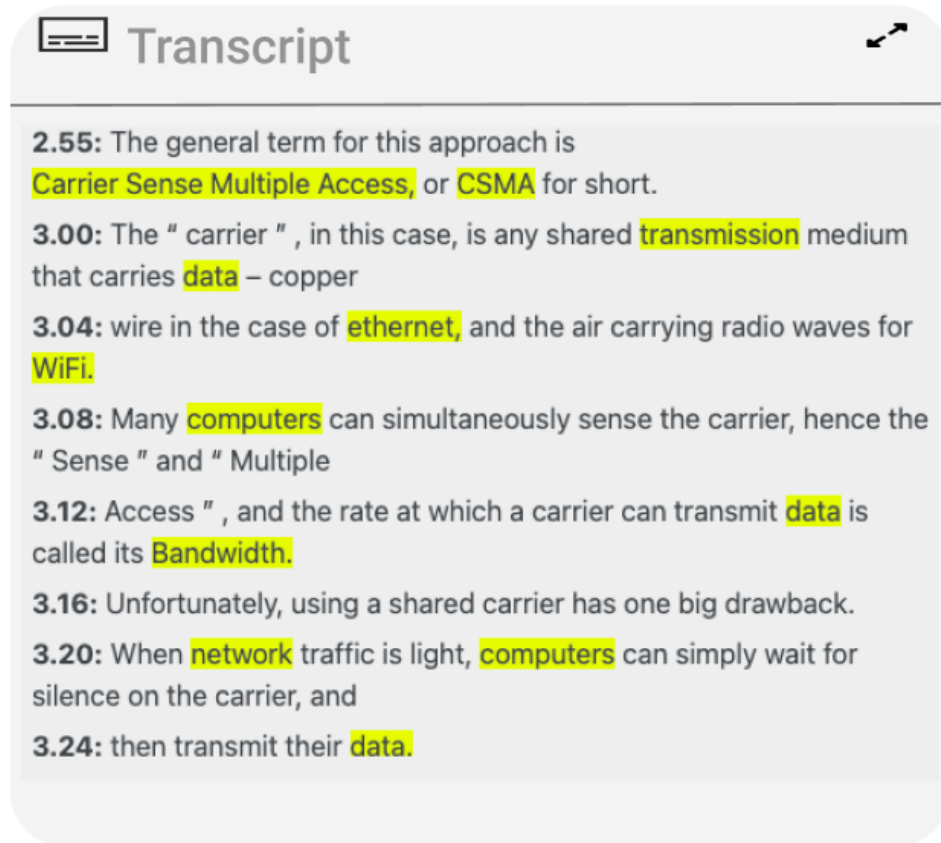
Figure 4.9: EVA Transcript

4.4.2   Map of Concepts' Flow

The Map of concepts' flow box, on the right side of the interface, shows
the concept map of the video (figure 4.10).

As we have already mentioned above, the use of concept maps in ed-
ucation is widely recognised for their support and visual feedback value.
Prerequisite maps are a particular type of map, which in the case of
our Map of Concepts' Flow is conceived as a semantic, dynamic and
interactive map. Unlike static maps that refer to eventual consolidated
and domain knowledge, the map in EVA flows synchronously with the
video. Concepts evolve together with the video flow and they are ini-

tially presented with lower complexity resulting in a map that shows a simple knowledge graph. During the consumption, the same concepts are deepened with other notions that are explained. This is reflected in the dynamic map, that is progressively enriched with new edges and nodes. The underlying graph is built as a view over a semantic RDF graph, automatically extracted from the video and annotated according to the W3C Web Annotation Vocabulary.

At the level of visualisation interface when concept is in focus, within the video, it will be highlighted and its node turns in red. In parallel when a prerequisite relations starts, the involved nodes are colored with light red, moreover, the system automatically zooms the map to the concept that is being explained at the current moment. During the video consumption the already mentioned concepts are coloured in light red. Concepts not yet introduced during the video are coloured in grey. Through this colour scale, therefore, the user is always informed about the concepts that have already been introduced in the video, those in focus, and those yet to be explained and/or mentioned. The relationships between them are also highlighted with a more intense/less intense black. In this way the user has dynamic support while watching the video, but can always decide to have an overview of the entire map instead. In any case, the map at the end of the video will be fully coloured and all relations made explicit.
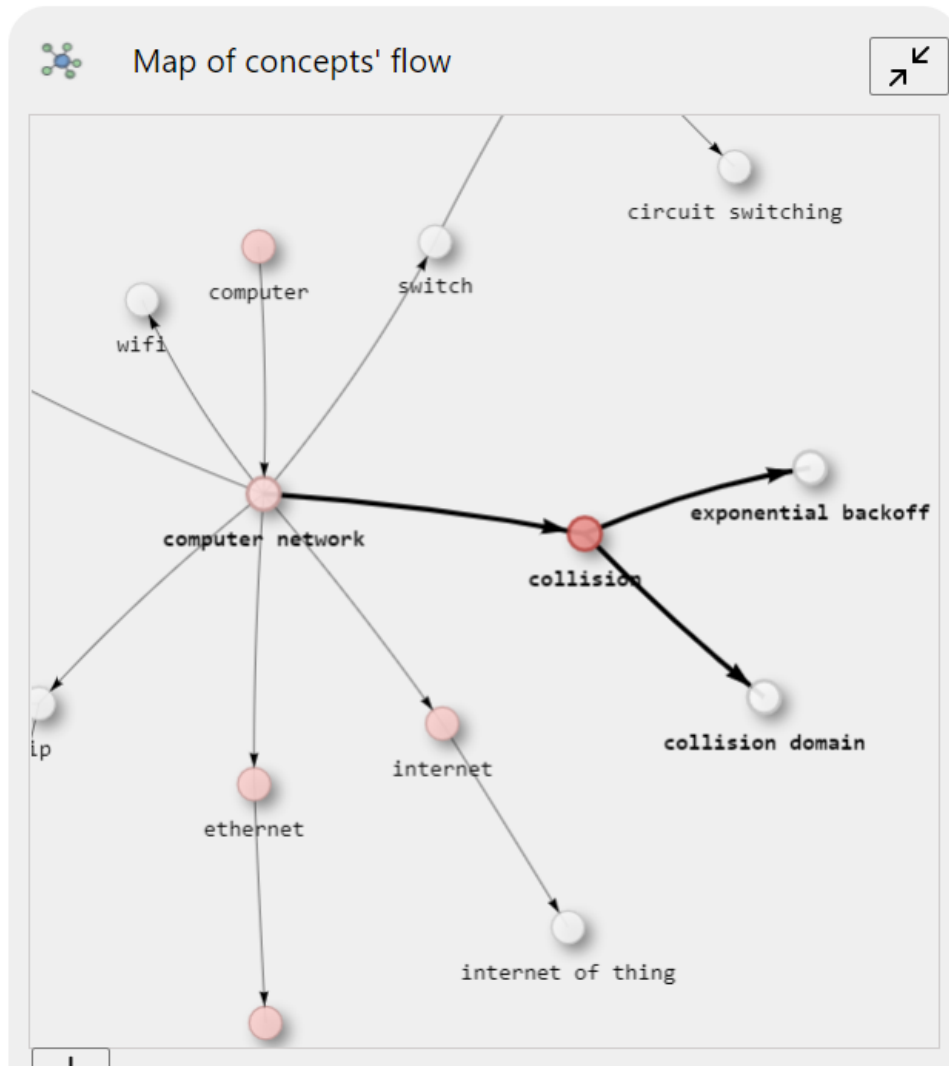
Figure 4.10: Map of concepts' flow

4.4.3   Concept Search

In the EDURELL framework the concept plays an important role, as we have already seen in the introductory chapter, it is not always easy to circumscribe the boundaries of concept identification. Depending on the domain, a concept may have value and be stand-alone or depend on other concepts in order to be understood. In EDURELL, identification even in the annotation phase follows the guidelines proposed by Wang et. al [163].

Thus, starting from this theoretical basis, the granularity of the search within EVA varies from the video fragment (which we will see in the

next section) to the single concept. While watching the video, or even in the subsequent stages of consolidating the educational materials, the student, starting by clicking on the node of the concept of interest within the Map of Concepts' Flow (figure 4.11), will see hotspots appear on the progress bar of the video (figure 4.12). As you can see in the pop-up in the figure 4.11, clicking on the node of the concept of interest to the user appears various options relating to the occurrence of the concept within the video: 1) with *Definition* we define the concept presented within the video in an extensive and probably introductory way by the lecturer/instructor, within the progress bar it is therefore signalled with the green hotspot; 2) with *In Depth* we define the concept mentioned again by the lecturer/instructor to complete its description or possibly to provide an example, within the progress bar it is signalled with a hotspot of grey colour and smaller in size than the hotspot of the *Definition*.
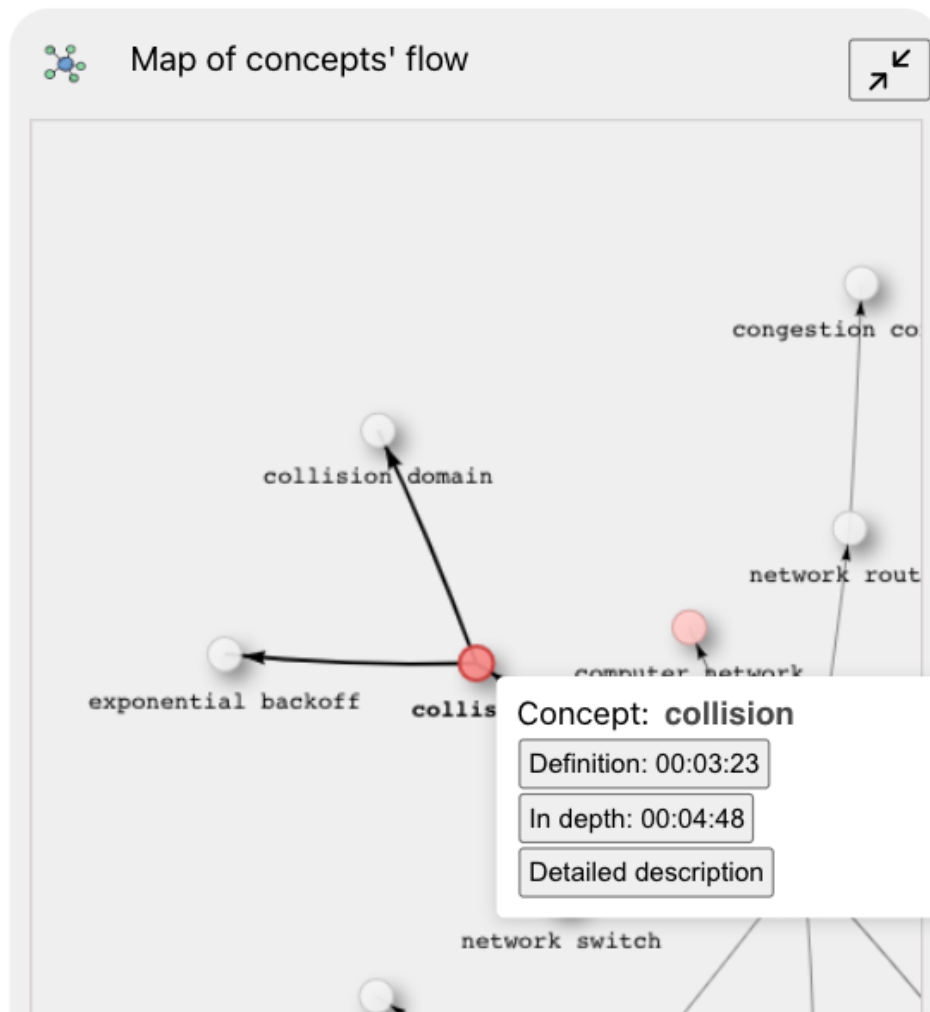
Figure 4.11: Concept Search within the Map

Figure 4.12: Concept hotspots on the progress bar

In the pop-up that appears when clicking on the node, there is also a *Detailed Description* button. By clicking on it, the user is shown a new overlay dialogue box (figure 4.14) with a detail of the sub-graph where the searched concept is in focus. This sub-graph shows the learning path of the concept. In the EVA visualization the red concept is the one that the user selected, the blue concepts are the *primary notions*, which

are the fundamental concept that does not have any prerequisite (nodes without incoming edges), the pink concepts are the prerequisites of the selected concept and the grey concepts are so-called "target" concepts which in the subgraph are the children of the concept in focus (red nodes).

In the box on the right there is instead a description, which collects the various portions of text within which the concept is described within the video, a link to the time-stamp of the occurrence of the concept within the video and when available an image of the concept extracted from the video.

It is also possible to reach the *Detailed Description* screen by searching directly for the concept within the *Concept Search Bar* at the top of the video's main page (figure 4.13). The figure 4.14 instead shows the display of a specific sub-graph.
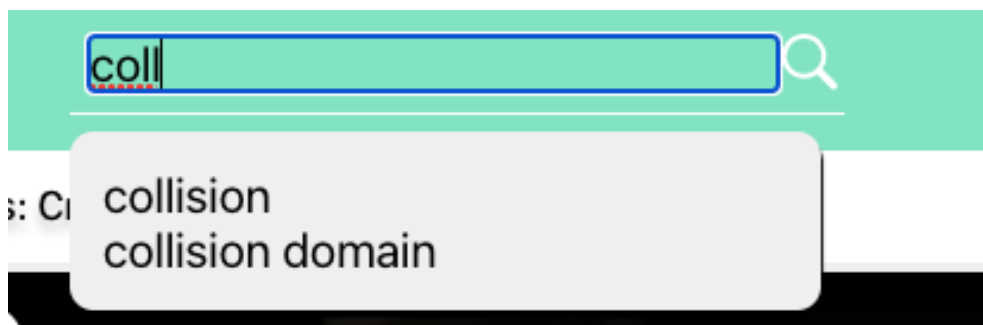
Figure 4.13: Concept Search Bar
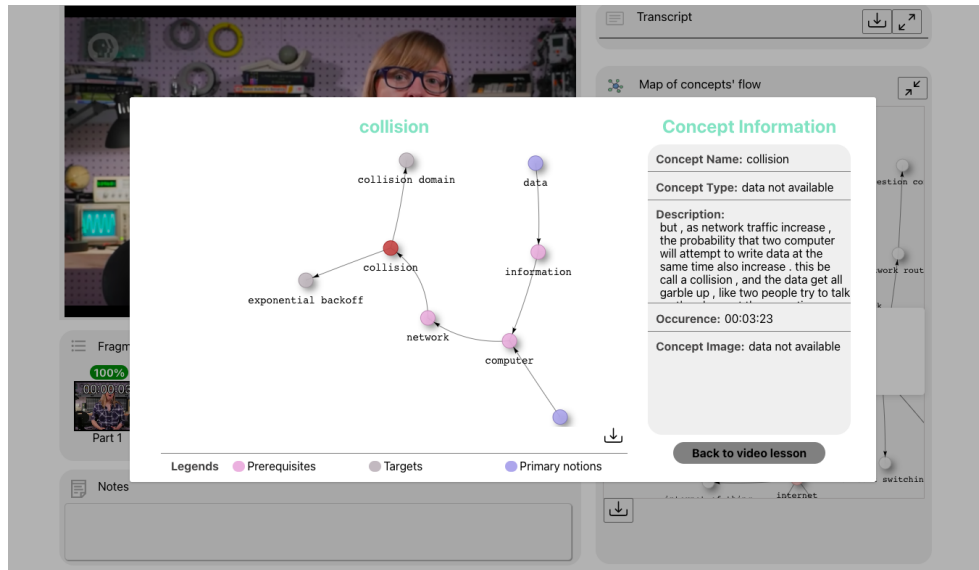
Figure 4.14: Concept sub-graph

4.4.4   Fragment Navigation

The *Fragment Navigation* (see figure 4.15) is a further navigation feature within the EVA interface. As mentioned in the previous section, if with concept search we have a concept-oriented level of granularity, in the case of *Fragment Navigation* we refer instead to portions of videos (segments) that within them deal with a broad topic. Thus within a fragment we may find a number of concepts, which are also present within the map and in the keyword highlight of the transcript.

The *Fragment Navigation* is conceived, in a non-linear consumption vision, as a table of content of the video, which, thanks to the identifying image of the segment (keyframe), the title of the fragment and the pop-up overlay with a list of the concepts within it (by hovering the mouse on the image), allows the user to be able to look at unclear parts of the video. Since the user's traces are saved within his or her account, he or she will be able to keep track of which parts of the video are viewed in full or not thanks to the mini completion bars on each fragment: 1) in green if the fragment was viewed by the user in full; 2) in red if it was viewed only in part (with an indication of the percentage) 3) in grey if the fragment was completely skipped (figure 4.16).
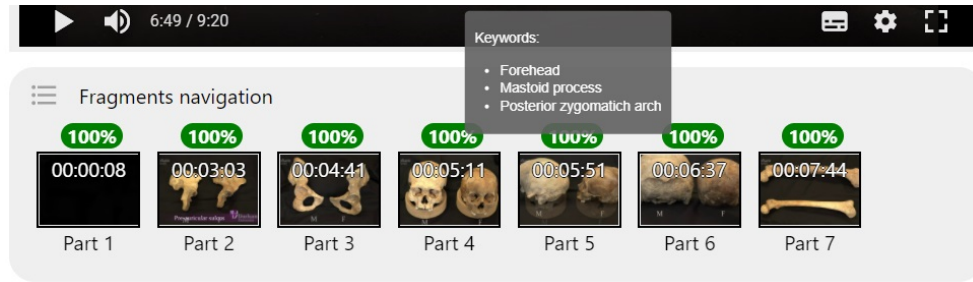
Figure 4.15: Fragment Navigation

Figure 4.16: Fragment Completion Bars

An additional implemented feature is the ability to take notes while watching the video. Figure 4.17 shows the textboard in which a student can type his notes about the topic he or she is learning.
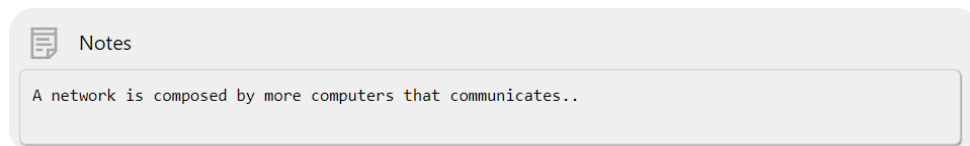
Figure 4.17: Notes

4.5 EVAnn tool

Within the EDURELL Framework EVAnn enables the video processing for video augmentation. This app is closely linked to EVA, in fact, as we have already seen in section 4.1, they share the same database. The user account is unique for both apps, and changes made in one app also automatically appear in the other.

EVAnn has been conceived as a support for the annotator, i.e. a domain expert figure who could be configured as a teacher/instructor, who thanks to the pre-processing and automatic extraction tools (which we will see in the next section) exploits the platform to annotate an educational video to be then submitted in EVA to a class of students. To the best of our knowledge, there is no app that allows the annotation of the educational video with the peculiarity of building a prerequisite map.

Within EVAnn there are numerous functionalities, which we will see in detail later on, that allow the construction of the graph that will be the main support in EVA, i.e. the Map of the Concepts' Flow and all the features connected to it. EVAnn consists of an annotation module, an analysis module and a Burst module for automatic and semi-automatic extraction.

4.5.1   Annotation Module

The annotation module, which will be described in detail in this section, is the heart of the EVAnn application. Within it, a main, compact interface contains a number of related interfaces: 1) Add Relation Interface; 2) Add Concept Description Interface; 3) Concepts and Synonyms Interface; 4) Annotation Visualisation.

Main Interface

The EVAnn main interface (figure 4.18) is designed for real time annotation with a number of specific features that we see in detail below. Following the upload of the YouTube video, which is incorporated into the catalogue of the EDURELL tool, the user can start the annotation. On the left-hand side of the screen is the video player, while on the left-hand side is the transcript, which, as in EVA, runs synchronously and with the precise timestamps of each sentence. Keyword highlighting allows the annotator to have an automatically pre-processed dataset.

Fragment Navigation is also present in this interface to facilitate the annotator's work.
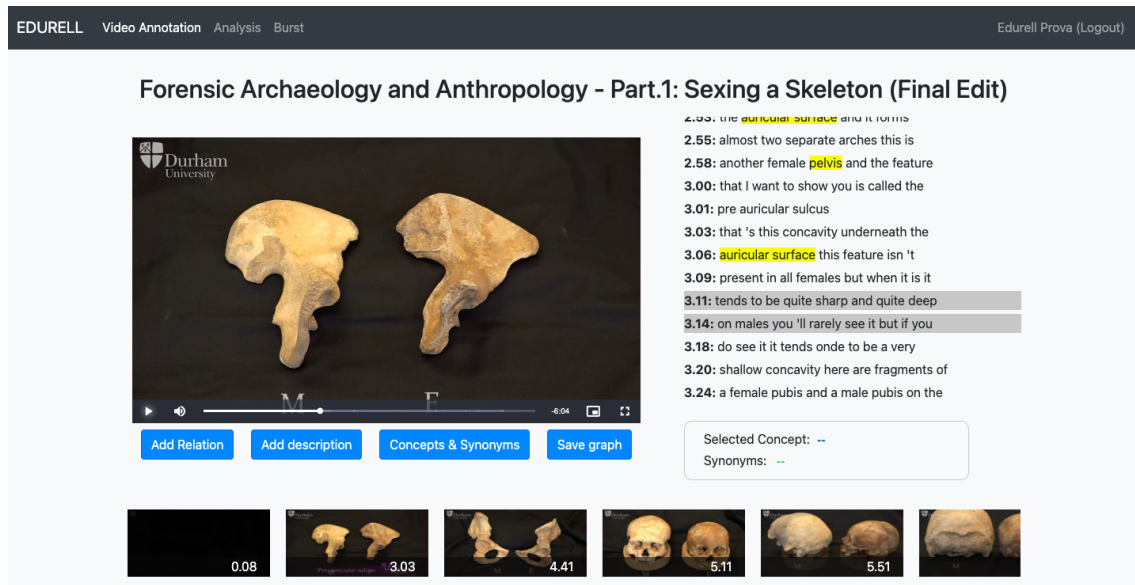
Figure 4.18: Main interface for video annotation

Below the video player there are four button:

- *Add relation*, that allows the insertion of the prerequisite relations.

- *Add description*, to handle the insertion of the concept description.

- *Concept and Synonyms*, to manage the concept list.

- *Save graph*, that saves the video on MongoDB and it downloads the JSON-LD to the user.

Add Relation Interface

The *Add relation* interface (figure 4.19), enables one of the fundamental operations for annotating the video, namely the actual construction of the prerequisite map. In this interface, after clicking on the relevant button on the main menu, the video is stopped and the last sentences of the transcript are displayed below the video frame, with the concepts highlighted. The user can then enter the "target" concept and the "pre-requisite" concept in the box on the right and assign a weight to the relationship, strong or weak.
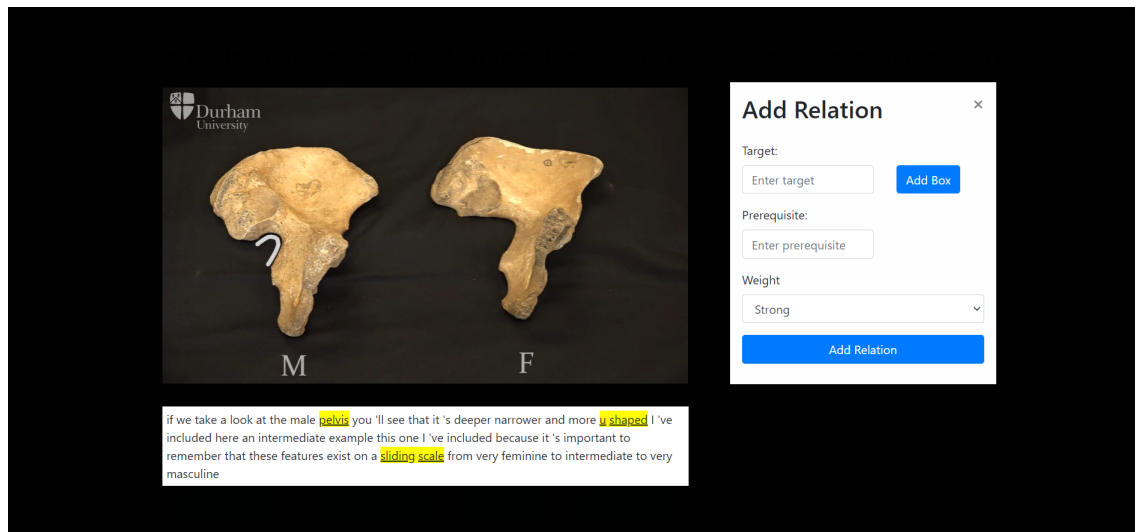
Figure 4.19: Interface for prerequisite relation insertion

It is feasible, but not mandatory, to add the area of the frame where the concept resides by clicking "Add box". The idea behind this additional feature is that it is possible for a concept to be merely displayed in the video part but not named in the audio part by the instructor or that the concept is both named and displayed within the video part; in figure 4.20 there is an example in which the target of the relation is "skeleton" and as it appears in the video, a box can be added to indicate where it is in the frame.
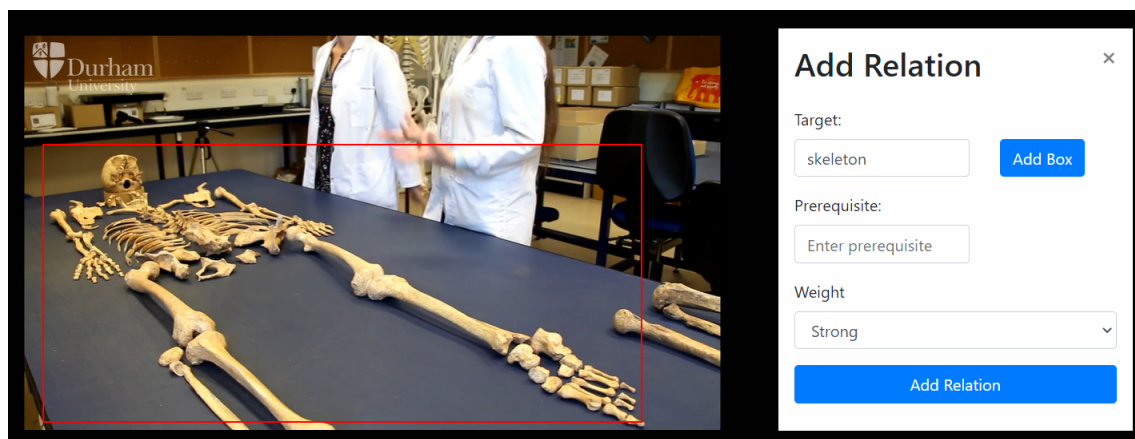
Figure 4.20: Add bounding box

Add Concept Description Interface

The *Add description* interface (figure 4.21) allows the annotator to define the portion of the video, with a start time and an end time, in which a specific concept is described. The description can have two different types: it can be either a "definition" or an "in-depth", which means that the annotator will enter for the first type the portion of the video in which the concept is presented by the lecturer/instructor in a fairly exhaustive and complete manner, while he/she can choose the "in-depth" type when in the course of the video the concept is again mentioned, explained with an in-depth, example etc. This will also result in the similar feature being displayed in the EVA user interface. In order to change the start time and end time of the description, the user can slide the bar below the video. The play button is needed in order to rewatch the inserted description.
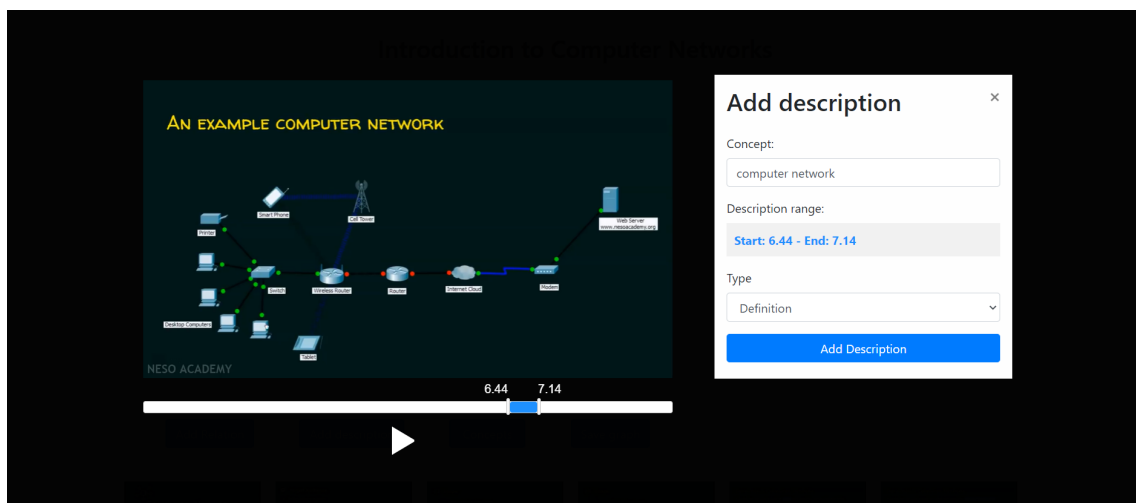
Figure 4.21: Interface for concept description insertion

Concepts and Synonyms Interface

The *Concept and Synonyms* interface allows the annotator to edit the list of concepts which, thanks to automatic extraction, already contains a set of keywords. In some cases, however, despite the refinements, it may happen that some domain concept is left out or some too general
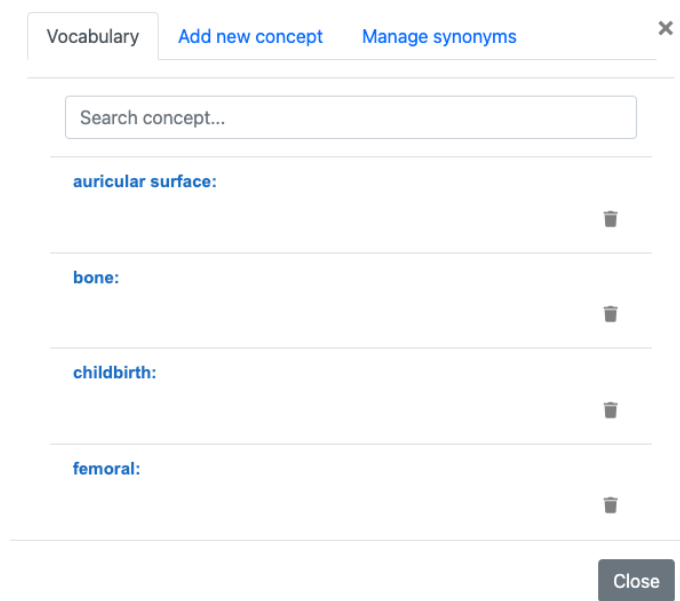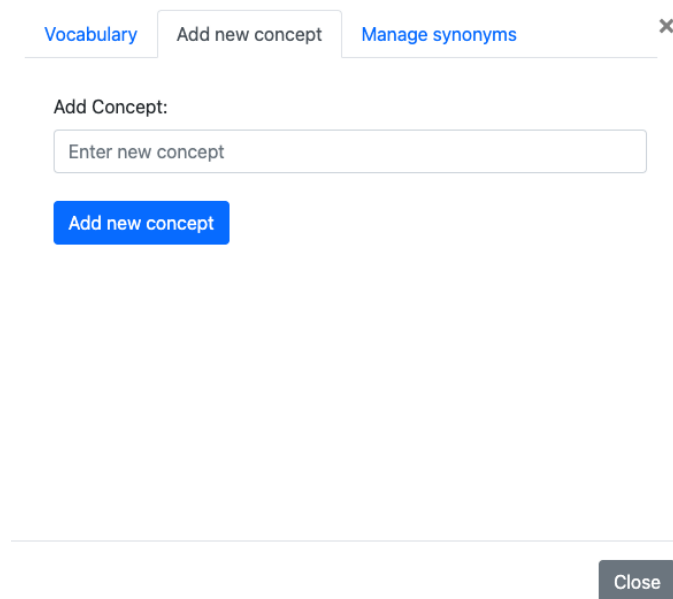
Figure 4.22: Vocabulary

concept is inserted, the annotator is therefore given the possibility to add and/or delete to clean up the list. In the case of the addition of a new concept the system checks if the inserted concept is in the transcript, whether yes, all the occurrences will be highlighted.
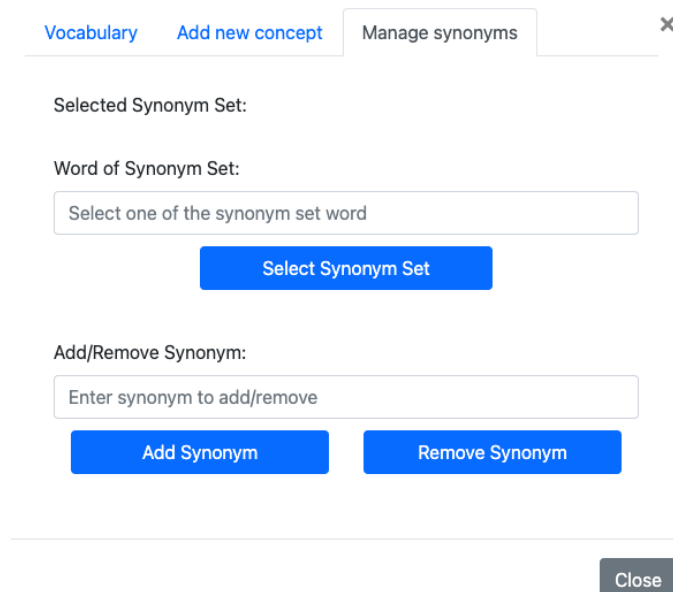
The dialogue box is tripartite, so there is a "Vocabulary" (figure 4.22) with the list of concepts, an "Add new concept" feature (figure 4.23) and finally a "Manage synonyms" feature (figure 4.24) that allows one to select a word from a *Synonym Set* and associate it with another term, considered as a synonym.

This association of synonyms takes place within the annotation of the specific topic of the video where two terms, which might not normally be synonyms in common language, but which in the domain knowledge sphere can be considered overlapping. The display of synonyms takes place on the main screen, where there is a box below the transcript: clicking on the highlighted concept will also display its synonyms (figure 4.25).

Figure 4.23: Add new concept

Figure 4.24: Manage Synonyms

Annotation Visualization

To complete the presentation of EVAnn's annotation module, we also provide visualisation screens. Below the navigation section of the fragments, a visualisation interface (figure 4.26) allows all added relations and definitions to be checked. An annotation can be deleted using the

Figure 4.25: Concept and Synonyms box

Figure 4.26: Annotation visualization

"Trash" button.

By clicking the "Graph" button, the user is able to visualize in real time the concept map of the video (figure 4.27).
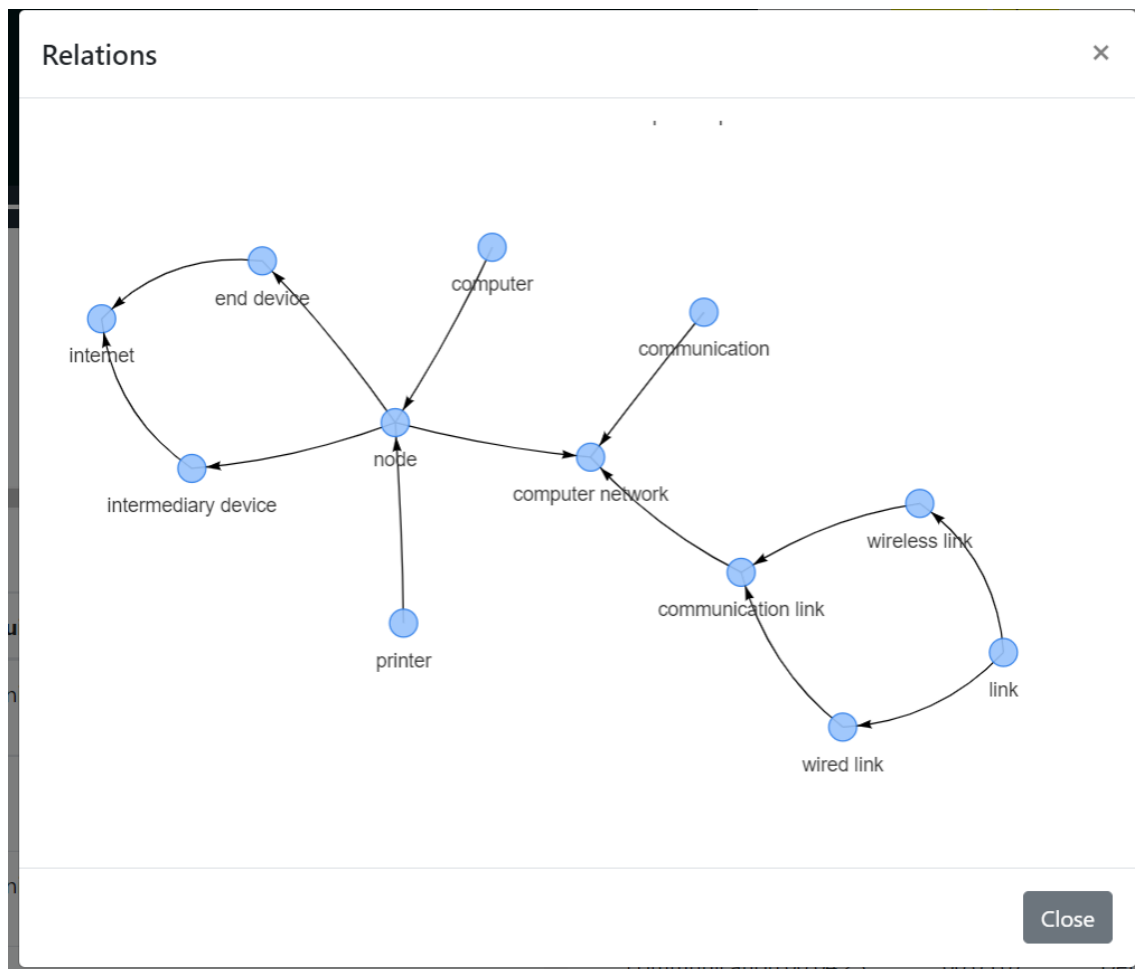
Figure 4.27: Concept map visualization

4.5.2   Analysis Module

The EVAnn tool is composed, in addition to the manual annotation module, also of the annotation analysis module and the automatic and semi-automatic extraction module. Within the analysis module (see figure 4.28) we can find several features:

- *Data Summary*

- *Linguistic Analysis*

- *Agreement calculation*

- *Fleiss kappa*
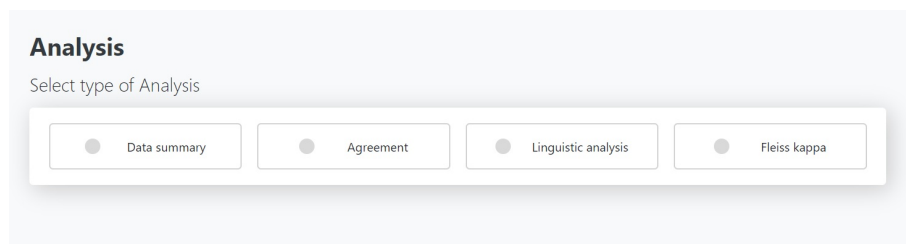
Figure 4.28: Type of analysis selection

Data Summary

The *Data Summary* section (figure 4.29) is designed to provide the user a quantitative report of the prerequisite relationships between two concepts (with respect to the graph) and also on the number of descriptions/in depth of the single concepts.

Linguistic Analysis

The *Linguistic Analysis* section (figure 4.30) offers the possibility of analysing a specific prerequisite relationship between two concepts, including the weight parameter. The relationship will be shown in its context within the transcript, POS (part of speech) tagging and visualisation on a graph complete the analysis features available.

Data Summary

Video title	**Forensic Archaeology and Anthropology - Part.1: Sexing a Skeleton (Final Edit)**
Annotator	**Maria Rossi**
Number of total descriptions	37
Number of definitions	27
Number of in depths	10
Number of relations	100
Unique relations	98
Strong relations	100
Weak relations	0
Number of concepts	60
Number of transitive relations	17

Figure 4.29: Data summary

Agreement Measures

The *Agreement* section enables the user to calculating the agreement between two annotations measuring the similarity and annotation reliability. It is based on the Cohen's kappa measure, which value ranges from 0 to 1, where 1 identifies agreement. Cohen's kappa is defined as:

$$k = \frac{P_o - P_e}{1 - P_e}$$

Where P_e denotes the agreement anticipated by chance and P_o represents the percentage of agreement between two concepts (i.e., the probability of each individual category). This metric also makes use of the transitivity of relations to determine whether two annotations are consistent with one another. Another measure of agreement can be calculated within EVAnn: Fleiss' kappa, computed between all annotations produced for the same video.
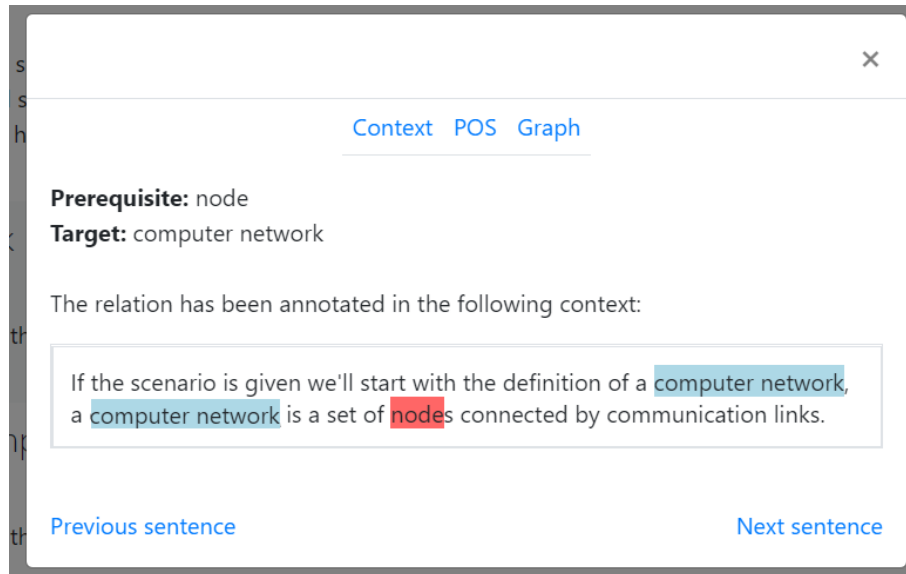
Figure 4.30: Linguistic Analysis

4.5.3   Burst Module

The Burst module is dedicated to the automatic and semi-automatic extraction of prerequisite concept maps. In subsection 4.6.3, we will see in detail how the extraction algorithm works. In this section, we show the module from the point of view of the interface.
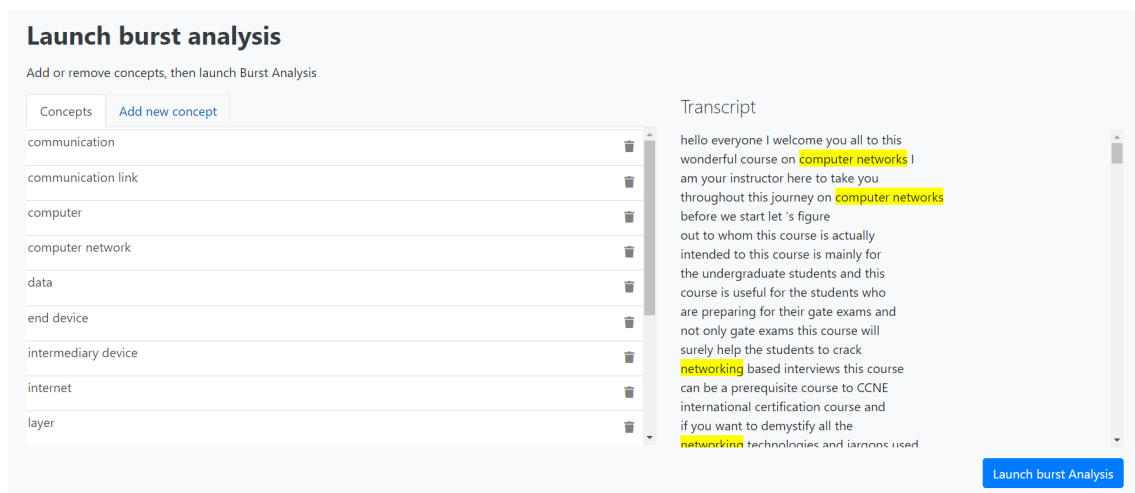
Figure 4.31: Semi automatic interface

Through the semi-automatic approach (figure 4.31), the user is asked to manually insert the concepts within the video. For this purpose, a

list of concepts are initially shown to the user that he/she can edit. The transcript of the video, with the highlighted concepts, is shown to facilitate the operation. If the video has not been previously annotated manually, the highlighted list of concepts will be extracted automatically, if not, the system shows the manually annotated list. This is useful in order to have a more precise extraction and to be able to do a comparison between the burst method and the manual extraction.

Burst analysis results

Data Summary	Definitions	Relations

Video title	**Forensic Archaeology and Anthropology - Part.1: Sexing a Skeleton (Final Edit)**
Number of concepts	**54**
Number of extracted descriptions	**31**
Number of definitions	**24**
Number of in depths	**7**
Number of extracted relations	**90**
Unique relations	**90**
Number of transitive relations	**25**
Agreement with Maria Rossi	**0.406**
VEO with Maria Rossi	**0.453**

Figure 4.32: Burst's results interface

In the automatic approach the Burst analysis is performed instantly, with no possibility of prior modification by the user.

Once the graph is extracted, all the information about it are shown to the user. The interface is composed by three tabs: Data Summary, Definition and Relations (figure 4.33).

The first tab contains the report of data summary analysis, if there is a manual annotation there is a comparison with the latter. The second tab shows all the descriptions extracted by the burst (figure 4.34). In the Relations tab there is a visualization of the extracted relations extracted, browsing the prerequisite concept map to follow the learning paths (figure 4.34).
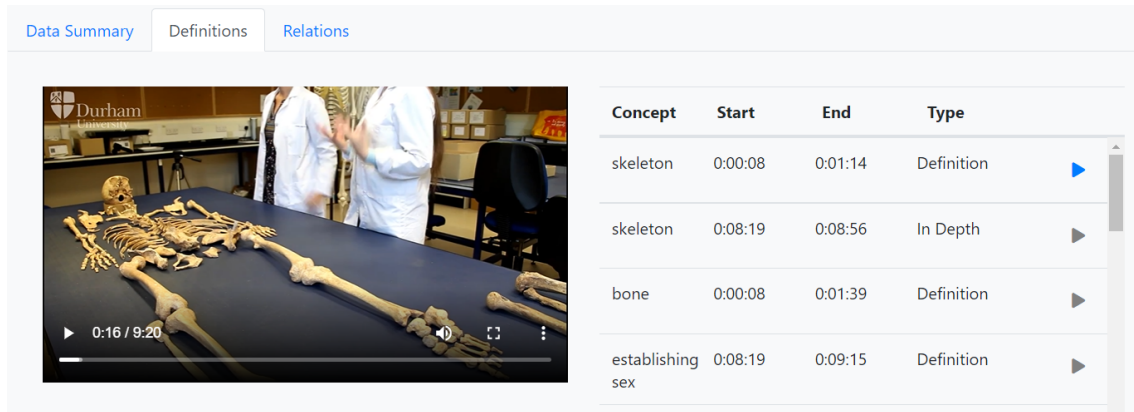
Figure 4.33: Burst's extracted definitions

4.6 Automatic Methods for Knowledge Extraction

Within the EDURELL framework, the automatic extraction of knowledge from videos focused mainly on three main strands: 1) keyword extraction; 2) fragment extraction; 3) prerequisite concept map extraction. The following subsections discuss in detail how the extraction takes place.

4.6.1 Keywords Extraction

The process of transcript generation occurs automatically within the EVAnn app, which uses the YouTube API to produce the subtitles, both auto-generated and manual-generated, with their starting and ending times. Once the subtitles are downloaded, the system will create the transcript. If the subtitles does not have punctuation, a neural net is add it using *punctuator*. From the obtained transcript, the CoNLL is generated using the UDPipe APIs [6], and parsed using the *conllu* library. This operation allows the visualization of the synchronized transcript and to browse the video through it. Concerning the process of automatically extracting keywords, which are needed both to enrich the transcript during the visualisation phase on EVA, and to help the annotator's work

---

[6]https://lindat.mff.cuni.cz/services/udpipe/api-reference.php
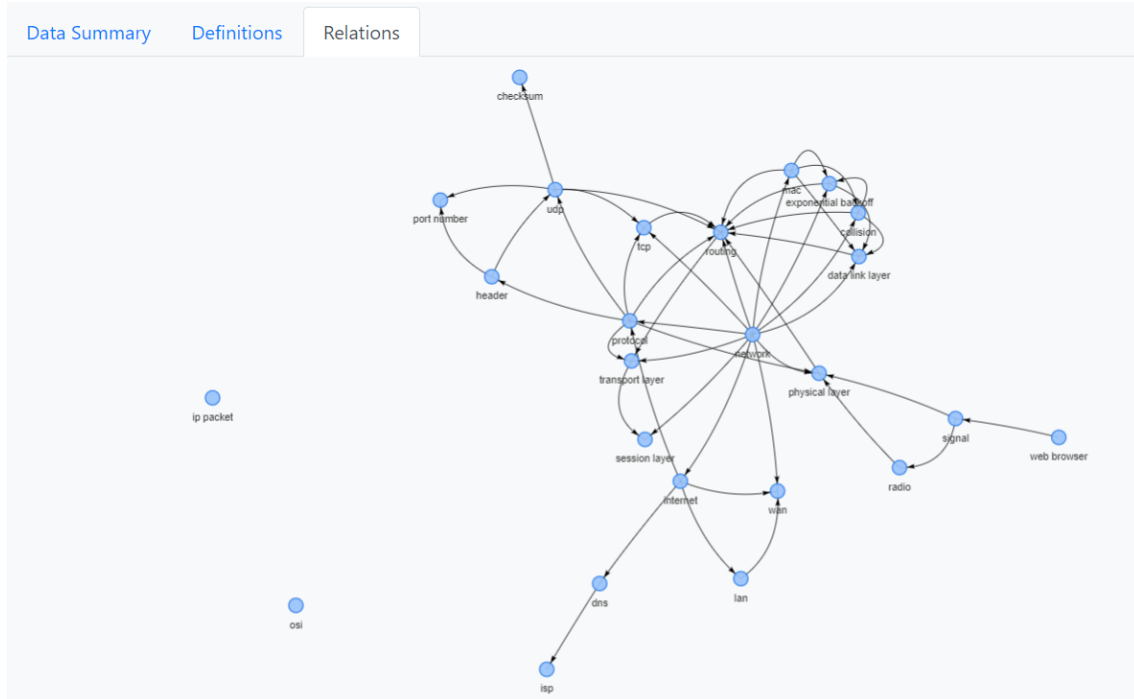
Figure 4.34: Burst's extracted relations

in the preparatory phase on the EVAnn tool, various methods were tested using different Python libraries (Gensim, Pke, Spacy, keyBERT, Python-rake, PhraseMachine). The transcript dataset of five educational videos on YouTube in the domains "archaeology" and "computer science" were manually annotated by experts. The results of the automatic extraction were compared with the manual annotations to assess the correctness of the results, calculated with the metrics of precision, recall and f1score. The analysis shows that the best method is the one that combines the results of RAKE with the 3 most common words extracted with *phrasemachine*.

4.6.2 Fragment Extraction

The fragment extraction consists of video segmentation performed by automatic methods. As we also saw in the section on the state of the art, numerous techniques are exploited to achieve this task. In our project, we carried out several experiments and combined various approaches. We explored the *NoVoExp* solution [162] using semantic similarity and

we started by their pipeline. The punctuated transcript has been splitted in sentences using the *nltk* tokenizer, each sentence has been associated to its starting and ending timestamp, obtained from the times of the Youtube subtitles. Each sentence has been mapped in an embedding and converted in a tensor, using the BERT model[7].

For every the mean cosine similarity of the current segment has been computed. In parallel the cosine similarity between the current embedding and the previous one has been calculated, too. If one of the two similarities was above a threshold, the two embeddings has been merged into a single cluster. Obtaining these "raw" segments, two refinements has been applied:

1) Aggregation of segments shorter than a given duration (40 seconds).

2) Adjusting start and end time of each segment based on detected scene changes.

The color histogram method has been exploited to detect scene changes: for each frame we compute the distribution of the colors and, if the difference between two frames is beyond a threshold, a scene change is detected.

4.6.3   Prerequisite Concept Map Extraction

The extraction of prerequisite maps is the last of the automatic extraction methods that takes place within the EVAnn app. In this case, extraction being of fundamental importance for the purposes of our project, it was also made semi-automatic with the possibility of being able to check the list of incoming concepts and to make the extraction as precise as possible. The algorithm used for the extraction is the Burst [1, 78], which is often used in the literature for the extraction of prerequisite maps from texts. Although in our case we were therefore talking about videos, by exploiting the transcript and with a few adjustments it was also possible to exploit it for our purposes. In fact, the proposed method takes into account not only the textual part, i.e. the transcript,

---

[7]https://huggingface.co/sentence-transformers/paraphrase-distilroberta-base-v1

but also the visual features. After a usual NLP pipeline of pre-processing
with the addition of punctuation, part of speech tagging etc., to which
the text of interest is subjected, the Burst algorithm, from [78], is ap-
plied to identify BICs (Burst Intervals of Concepts), namely the portions
of text in which concepts have a high density. This method [157] is use-
ful not so much for concept extraction, a task which is extensively dealt
with in the literature and for which many methods of various natures
have been tested, but for tracking the role the concept plays within the
text, i.e. to understand whether the concept in focus is described, in-
depth, recalled for an example or to introduce other concepts related to
it. Starting from the Hidden Markov model [126] to find BICs, which
show the stream's emphasis on a concept, through their lenght and in-
tensity we obtain the heuristic role identification. The video analysis
improves the identification of BICs thanks to OCR (Optical Character
Recognition) [110] techniques for slide detection and analysis. After the
BICs extraction phase, through the principle of Allen's interval algebra
for spatial-temporal reasoning [112], the prerequisite relations between
two concepts are extracted. According to Allen, in fact, two time inter-
vals can have different types of relations: X precedes Y, X meets Y, X
includes Y. Depending then on the weights assigned to the BIC pairs,
the result produces a score of the probable prerequisite pairs. This is
visualised by means of a PR graph.

To summarize, in this chapter the EDURELL project has been pre-
sented together with the description of the two main components of the
platform, the EVA and EVAnn modules. For both tools, but in partic-
ular for EVA, which represents the main contribution provided within
this thesis work, an attempt has also been made to outline the steps that
led to the design of the platform, starting from the research issues and
baselines drawn from the previous chapters. In this respect, this chapter
has also presented the use cases and scenarios used for the experimental
evaluations that will be presented in the next section.

Chapter 5

Experiments and Evaluations

5.1 EDURELL User Evaluation

The main research question has been defined in section 3.2 as follows:

RQ: To what extent video augmentation based on knowledge extracted from the video can improve the learning experience?

Then it has been detailed into sub-questions and corresponding research hypotheses in order to evaluate the tools. To achieve this objective a series of iterative evaluations has been carried out.

Sub-questions

The sub-questions underlying the evaluation are:

SQ1) What is the general user perception while using EVA tool?

SQ2) Are EVA features useful for immediate understanding of the content of an educational video within the first-watch learning context?

SQ3) How are EVA features perceived to support the first-watch learning experience?

SQ4) Are EVA features suitable to support the rewatch learning ex-

perience?

SQ5) How are EVA features perceived to support the rewatch learning experience?

SQ6) Is the method implemented within the EDURELL Framework useful for automatically extracting prerequisite relations from videos?

5.1.1   Preamble on the Choice of Domains

As already investigated in chapter 2, the use of VBL and interactive video is open to all levels of education and domains. Nonetheless, research and experiences, including commercial ones, seem to focus more on scientific disciplines. To give an idea in terms of numbers of this quantitative difference, of online course presence by domain, we quickly checked how many courses are present on two open MOOC platforms:

- **FutureLearn** is an English language MOOC platform, based in the UK. It is the UK's largest MOOC platform comprising 54 UK and international universities and four non-academic partners: British Museum, British Council, British Library and National Film and Television School.

- **EduOpen**, a project funded by the Ministry of Education, University and Research in Italy aimed at creating a platform for the delivery of courses defined as Mooc (Massive Open Online Courses) by a network of Italian universities and a group of of selected partners.

The choice of topics, which will also be the subject of subsequent evaluations, as part of the thesis project, fell on "computer science" and "archaeology". The search on the engines of both platforms gave as results:

- **FutureLearn**, 14 courses about "Archaeology", 388 courses about "Computer science".

- **EduOpen**, 0 courses about "Archaeology", 24 courses about "Computer science".

Within the EDURELL framework (in chapter 4), although the design of the tools developed are aimed at materials from any domain, through diversified experiments it was tested whether there might be differences between these two disciplines.

5.2 Evaluation Through Experts

The first evaluation carried out as part of the EVA tool experiments involved a heuristic evaluation by experts [36]. The evaluation, following Nielsen's standard metrics [114], focused mainly on the evaluation of the usability of the interface, from the identified issues, again with a view to formative evaluations aimed in an iterative process at constantly improving the EVA tool releases some modifications were then applied for the subsequent evaluation carried out with end users.

5.2.1 Method

The experiment was conducted by submitting 3 different scenarios to the experts and with the choice of standard metrics that we will detail below.

Scenarios

Three scenarios have been defined and then we asked the experts to face the three situations described below:

1. **First-time viewing (first watch):** "Imagine you are a bachelor's degree student in Archaeology, whose video lectures are included in the EVA system. Imagine that, within your course of study, face-to-face courses, online lessons and courses on MOOC platforms (as additional activities to get credits) are provided. Then you decide to attend the "Forensic Archaeology and Anthropology Course"in autonomous mode. The EVA platform, with its HV functionalities, will provide you with the needed support to follow the course on

your own. In this case, specific features of the application can improve your first viewing of the video lesson, in order to support your understanding of the video content".

Given this scenario, specific actions were explicitly required to the experts in order to use all the functionalities intended to support the first watch, including the Transcript and the Map of Concepts' Flow.

2. **Second time viewing (rewatch):** "In the same scenario as before, you have already watched the videos of the course and now you may want to browse again the contents of the first lesson because you feel you need to watch it again and in particular some of the explained concepts. In this case, you might not want to watch the whole video again, but just get an overview of the structure of the concepts, rewatch the parts you do not remember and get details on the related concepts". As above, specific actions were required to the experts to simulate this scenario, including the interaction with the Concept Graph and its related features for detailed view and concept search.

3. **Focus on contents with and without HV support:** In the third scenario, the experts were required to switch between the HV functionalities provided by the EVA platform while watching the videos and perform some actions aimed to make them focus on the content of the video. Then they are instructed to list the prerequisite concepts for a certain concept in the notes panel. The objective is to get the expert to consider which elements of an educational video help students better understand the concepts and how they relate to one another.

Metrics

In order to obtain an initial expert evaluation of the interface it was necessary to refer to a standard recognised by the scientific community, the history of ISO standards in this area is long and we will relate to the

ISO 9241-11 standard, that gives a definition on what is usability and its application in different fields of application [16]. This standard defines usability as "the extent to which a product can be used by specified users to achieve specified objectives with effectiveness, efficiency and satisfaction in a specific context of use"[16]. The ISO standard provides also a description of the terms in the definition, as follows:

- User, the person who interacts with the product;

- Objective, the expected result;

- Effectiveness, accuracy and completeness with which users obtain specified goals;

- Efficiency, resources spent in relation to accuracy;

- Satisfaction, freedom from discomfort and positive attitudes towards the use of the product;

- Context of use, users, activities, equipment, and the physical and social environments in which a product is used.

It is important to emphasise that the concept of usability [85] cannot be measured scientifically in the round, but as it is also a condition related to subjective condition, however much standards and metrics may facilitate evaluation. For example, even the mentioned standard is still under review [124]. With these considerations and premises one of the most widely used techniques is to carry out a heuristic evaluation to find any usability problems, based on the so-called "heuristic principles" or "usability heuristics".

There is general agreement about the value of adopting a preliminary heuristic evaluation to identify *a priori* any usability issues before performing further evaluations, including extensive usability tests with end-users, although each domain should have an adequate set of usability heuristics because the more generic or traditional ones will not be able to accurately evaluate the specific characteristics of the various types of software and applications.

According to the model proposed by Nielsen and Molich [114] involving usability experts who inspect the interface of the tool for possible usability issues, the evaluation has been performed by 5 experts in HCI and usability tests.

The method chosen to run the test was to provide descriptions of the scenarios and then leave the experts free to navigate the interface through the entire pipeline, from the new account registration procedure to the exploration of the various hypervideo features. They were then asked to evaluate their experience following the aforementioned Nielsen heuristics and provide comments and feedback for improvement.

5.2.2   Results and Discussion

The results of the evaluation are given below, with the reference of the heuristic the issue and its suggestion for improvement is made explicit.

1. Visibility of the system status. – *The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.*

   Issues: At the first attempt, the platform does not provide information nor a catalog neither a preview of the available videos. Users must search the videos within the system through a search bar and then, when typing, automatic suggestions appear. Only when you are registered and have already watched some videos, recent videos are shown in "your history". A short description is missing in general, and for watched videos, some more information should be shown such as, e.g., watched, watching, or even progress within the time-line.

   Suggestions: Such an issue can be fixed by adding a visual catalog on the start screen, presenting videos as cards with short synopsis and making search available as a second choice. Furthermore, users should have personalised home pages withe their favourites and most popular videos (possibly per category), as well as the recently seen

ones. This will also resolve the problem of having an empty starting page.

2. Match between system and the real world – *The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system oriented terms. Follow real world conventions, making information appear in a natural and logical order.*

   No issues were reported by none of the reviewers, comments agreed that the used language is not too technical, nor ambiguous, neither chaotic.

3. User control and freedom – *Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.*

   Issues: escape routes are missing after search results are presented.

   Suggestions: an emergency exit should be provided by adding a "This is not what you were looking for?" message on the results page. Also, a "Back to previous page" command would be helpful.

4. Consistency and standards – *Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.*

   Issues: The navbar includes a "Home" button but the logo is not a link to the home page.

   Suggestions: While the "Home" button is still useful for a certain class of users, the link to Home should be in the logo too.

5. Error prevention – *Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error prone conditions or check for them and present users with a confirmation option before they commit to the action.*

   Issues: buttons in the navbar are too much near to each other .

Suggestion: redesign the navbar and adding a hover effect on suggested videos, to clarify the current position.

6. Recognition rather than recall – *Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.*

   No issues were reported by none of the reviewers.

7. Flexibility and efficiency of use – *Accelerators unseen by the novice user may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.*

   Issues: given the small number of functionalities, such feature is completely missing, apart for the chronology in home page.

8. Aesthetic and minimalist design – *Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.*

   No issues were reported by none of the reviewers.

9. Help users recognize, diagnose, and recover from errors – *Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.*

   Issues: such feature is completely missing.

   Suggestion: it is good practice to design tools keeping in mind that errors can occur at every time, e.g., when selecting a video, it should be made possible to make changes or deleting from own history.

10. Help and documentation – *Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search,*

focused on the user's task, list concrete steps to be carried out, and not be too large.

Issues: such feature is completely missing.

The experts also noticed that exporting data from the platform is possible in the `json` format, however, depending on the browser/OS combination, it may happen that the file is not downloaded, yet visualized on the screen, which results in a blank page. All in all, the heuristic evaluation provided us with numerous insights that were then exploited for the improvement of the platform, with a view to testing the tool in a later release to an initial group of students considered to be the typical end-users of EVA.

5.3 First Evaluation with Students

As we have just mentioned in the previous section, the heuristic evaluation with experts served to give a first round of feedback in order to submit the platform to the end-users. The idea was to start with an initial small group of university students, already beginning to identify on the basis of the tasks and learning context of the figure 4.3 to submit EVA features considered as so-called Visual Feedbacks (VF). This first evaluation has been designed first of all to investigate the specific sub-questions SQ2 and SQ3 concerning the general browsing experience and the immediate understanding of notions from educational videos in the first watch learning context [156].

5.3.1 Method

The controlled experiment has been conducted, in the first-watch learning contest, to verify effectiveness and limitations of the VF methods, with the objective of increasing the understanding of the video content. The independent variable of the experiment is the Visualization Feedback (VF) type. The three VF types defined for this experiment are as follows:

1. VF1) Watching the educational video with Transcript only support.

2. VF2) Watching the educational video with the support of the Map of Concepts' Flow only.

3. VF3) Watching the educational video in a normal video player as a baseline.

The hypothesis for this experiment is is that both the augmented transcript (VF1) and the map of concepts' flow (VF2) can increase the immediate understanding, helping the learner focus on the relevant concepts and their related concepts and contents. For the augmented transcript there are already evidences in the literature for its effectiveness, whereas for the map of concepts' flow studies are scarce and the risk for learners to get distracted by the dynamic map could be relevant, with the effect of potential reduced attention and comprehension.

Setting

As we have already seen in 4.4 EVA interface merges different canvases containing: 1) the video player; 2) the video transcript; 3) the map of concepts' flow; 4) a bar for navigating the video through indexed fragments; 5) a bar for searching for concepts; 6)a suited space for students' to take notes. The frames containing the transcript and the map can be toggled on and off according to the user preferences. For this experiment, participants are required to use only the video player, the video transcript and the map of concept's flow. The video used in the experiment is a video lesson on computer networks, available on YouTube[1].

Participants

For the experiment 12 university students of foreign languages has been recruited. They 9 females and 3 males, with a range of ages between 21 and 32 years old (AVG=24.6, SD=2.7). University students are representative of the intended end-users of the system, i.e. high level students.

---

[1]https://www.youtube.com/watch?v=3QhU9jd03a0

Furthermore we assume the student's gender does not affect the results. The criterion for choosing participants was to look only for users with a very basic knowledge of computer science. In this way, their prior knowledge would not have contributed to the correct answer of the questions in our test questionnaire, but a minimum of literacy on the topics was still necessary. None of the participant was affected from learning difficulties or eyesight problems (self-reported in an anonymized questionnaire).

Procedure

An online study has been used to carry out the experiment. Before explaining the procedure, it is necessary to clarify how the groups were formed and what the VF conditions were. The presentation order of the three VF conditions was varied between participants, according to a double Latin square (3*3) with three experimental conditions and three experimental periods, i.e. sessions. All the possible combinations of the VF conditions are covered through 2 groups of 3 participants and 3 sessions. We replicated it 2 times for a total of 4 groups and 36 trials. The steps of the procedure were:

1. First of all, participants had to quickly complete a demographic questionnaire including the informed consent and the questions about learning and eyesight problems.

2. In a second phase they have been assigned to one of 4 groups of 3 people, they were each given a briefing on the experiment and encouraged to try the user interface and VFs for three minutes, playing a demonstration video.

3. Following this brief training session, participants were given instructions on the tasks they were to complete. They were each required to watch the video under one of three different visual conditions: with the *map of concepts' flow* toggled and the textit *augmented transcript* turned on (VF1), with the map of concepts' flow toggled on (VF2), with both the transcript and the map toggled off (VF3). Participants were required to respond to a few questions

(Q) designed to evaluate their understanding of the video segment they had just watched and to gather their opinions regarding the VF condition on a Likert scale. Participants were not permitted to use the interactive features on the user interface or to pause or fast-forward the video during any session. Each session had a beginning and a finish at predetermined times, and it covered a single topic. The first session was longer to lessen the impact of a "cold start" on the subsequent results.

4. Participants were then asked to rank the VF techniques in light of various learning contexts and goals. They needed about 30 minutes to complete the experiment.

5.3.2   Results

In this section, we analyse the results of the questionnaires on the comprehension of the content of the video and what we call subjective feedback, again obtained through the administration of ad hoc questionnaires.

Results for Comprehension Questions

The test has been divided into sessions. At the end of each session, participants answered a few simple, multiple-choice questions related to the topics they had just followed in the video. To evaluate the effect of vision with Augmented transcript (VF1) or the Map of concepts' flow (VF2), on the first-watch comprehension, the average score of each VF condition, considering the trials, has been computed (figure 5.1). The results are: VF1 (AVG=0.67 SD=0.27), VF2 (AVG=0.58 SD=0.33), VF3 (AVG=0.35 SD=0.36). In order to investigate statistical significance between VF1 and VF2 compared to VF3 baseline we applied the t-test and we found that VF1 results are significantly better than VF3 ($p=0.02$), but they are not significantly better than VF2.
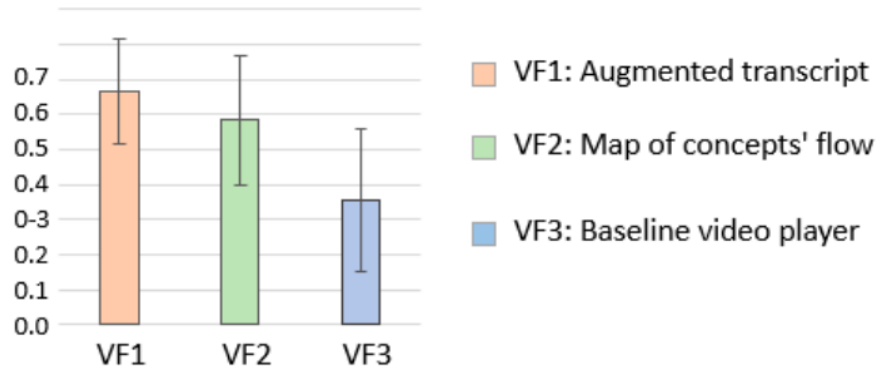
Figure 5.1: Results for video comprehension within First-watch learning context

Results for Subjective Feedback

To assess the subjective feedback regarding the use of VF1 and VF2, participants were given a questionnaire consisting of 6 items (three positive Qb1-Qb3, and three negative Qb4-Qb6)) with 5-point Likert scale (1=disagree, 5=agree). As can be seen from the figure 5.2, VF1 was more appreciated than VF2 for the Qb1-Qb3 items and less disliking for VF1 than VF2 (Qb4-Qb6). Always applying the t-test the difference is significant only for the positive questions (p=0.047). Considering the single items, only for Qb3 and Qb4 the difference is significant (p=0.006 and p=0.046 respectively).

		5-point Likert scale (1 disagree – 5 agree)	VF1		VF2	
			AVG	SD	AVG	SD
Qb1	Positive	The VF helped me to understand the content	3.33	0.78	2.83	0.94
Qb2		The VF helped me to focus on the important concepts	3.83	1.03	3.67	0.89
Qb3		The VF was in a right position	4.25	0.75	3.25	0.87
Qb4	Negative	The VF distracted me	1.75	0.75	2.42	0.79
Qb5		The VF didn't add any value to my watching experience	2.00	0.74	1.92	0.79
Qb6		I would prefer watching the video without the VF on	1.42	0.67	1.83	0.72

Figure 5.2: Results on subjective feedback after VF1 and VF2 sessions.

5.3.3   Discussion

From the results just reported, it therefore emerges that in this first evaluation with end users, VF1 and VF2 can indeed be considered as valid supports for the learner immediate comprehension, within the first-watch learning context, particularly in the aim of sending more focus to the most relevant concepts within the video. However, it must be emphasised that at the level of statistical significance, it is the augmented transcript that helps the most in the above-mentioned intent, which is in fact also confirmed by the literature [82]. This is also confirmed from the results to questions about the subjective feedback to each visual condition. As much as it can provide support even in the first watch phase the map of concepts' flow, results seem to suggest that it is expected to provide more support for other learning contexts (for example the rewatch learning context) and tasks. For this reason, the rewatch context will also be included in the subsequent evaluation (section 5.4).

It is important to emphasize that the study involved the students of a university course and it was run as part of a class activity in order to obtain more reliable results in terms of learning experience. In order to avoid, or at least reduce, complacency biases, the activity was carried out anonymously.

In general, we observe that both that both methods help in providing more structure to the video (as they were designed). By increasing the immediate understanding, an improvement in efficiency (less time to acquire the knowledge) of the learning process might occur. Another important aspect, which we have also already introduced in the state of the art analysis, is the risk of distraction (perhaps resulting from visual clutter). This risk has been confirmed by results that show a difference that is statistically significant about distraction produced by the map compared to the transcript (Qb4). We could therefore assume (considering Q3 results) that this is because more consideration must be given to the possibility, not only to activate/deactivate the panels, but also to be able to manually resize or move them. On the basis of the results and analyses, it became necessary to design a new evaluation

with students with a specific usability test and the evaluation of the further rewatch learning context.

5.4 Second Evaluation with Students

The second evaluation carried out with students aimed to test EVA in order to evaluate the usability and the experience, along the learning path, perceived by the intended users. Building on the results of previous evaluations and with a number of modifications and improvements made to the interface, the usability evaluation was designed to understand the general level perceived by users during the browsing experience. On the other hand, from the point of view of the validity of the support during the learning process, whereas only the first-watch learning contest was used in the previous evaluation, rewatch was also tested in this evaluation. The improvements made after the previous two evaluations were as follows: 1) the addition of a video catalogue; 2) the implementation of a user's video history; 3) help messages, by mouseover events, with information on the features of EVA; 4) yellow highlighting of transcript concept; 5) the addition and mapping of synonyms within the map.

5.4.1 Method

In order to be able to carry out the rewatch phase, it was necessary to design this evaluation on different days and following various steps:

1. The first part (Part 1) has been designed as a user experiment comparing between the "Augmented Transcript"-only and "Map of Concept's flow"-only features to explore the benefits and limitations of there features in terms of user experience of the learning environment.

2. The second part (Part 2) has been designed to evaluate the usability of EVA interface.

Below, we will describe the participants and the setting for the study that are common in both parts, then we provide the details of the study design of each part.

Setting

The setting of this experiment is the same as in subsection 5.3.1, with the EVA interface presenting the various canvases. For the first part of the evaluation, the setting involves using the video player, augmented transcript and map of concept's flow, wheres for the usability evaluation the whole interface will be considered. Also in this evaluation, the topic of educational videos is in the field of computer science [2] [3] [4].

Participants

For the experiment 69 university students has been recruited. These students were attending a Computer Science course at the Faculty of Foreign Languages. The composition was: 51 females and 18 males, in an age range between 20 and 31 ($AVG : 23.6$, $SD : 2.2$). All of them have at least intermediate English skills, and no one is native English speaker. University students are representative of the intended end-users of the system, i.e. high level students. We assume the student's gender does not affect the results. None of the participant declared learning difficulties or eyesight problems.

Part 1

The aim of the first part of the study was to test augmented transcript and the map of concepts' flow, to check the level of video comprehension and user experience. In particular, the purpose of this first part of the evaluation is to answer SQ2 and SQ3 in section 5.1.

---

[2] https://youtu.be/DKGZlaPlVLY
[3] https://youtu.be/xnyFYiK2rSY
[4] https://youtu.be/5o8CwafCxnU

SQ2) Are EVA features useful for encouraging immediate understanding of the content of an educational video within the first-watch learning context?

SQ3) How are EVA features perceived to support the first-watch learning experience?

Experimental Design

This part of the study was designed as a between-subject experiment with three different conditions: (C1) *YouTube video player* (C2) *augmented transcript* and (C3) *map of concept's flow*, without any visual cues as our control condition. Each participant used only one of the features.

Part 1 matches the "First-time viewing (first watch)" scenario provided to the experts (see section 4.3), with the difference that each user was required to use just one of the features, among the three conditions. Another difference is that, since we involved students from an Informatics class, the videos used in the study were not about archaeology, yet computer science. The topic of this evaluation, having recruited students from the computer science course is precisely relevant to their education. The videos, three short introductory videos, were the same for all the students in order to better compare the results, avoiding effects due to the video content difficulty. They are three YouTube videos[5], made available on the Edurell platform. All of the videos are in English.

As dependent variable for this study we used the subjective feedback in each condition, collected through questionnaire responses.

Procedure

The experiment has been conducted online, as a class activity, thus reproducing the usage condition of the EVA interface for online video learning.

---

[5]Video1: https://www.youtube.com/watch?v=MMzdKTtUIFM
Video2: https://www.youtube.com/watch?v=xnyFYiK2rSY
Video3: https://www.youtube.com/watch?v=IJv-PrInE$_A$

– The EVA platform was first explained to the participants, and then they were asked to test out the user interface and features. To achieve this, users were instructed to play one of the video already included on the platform, use each tool, and alternate between the *map of concepts' flow* and the *augmented transcript.* Participants utilized each instrument for around five minutes.

– Participants were then split in 3 groups, one for each experimental condition, and instructed about the tasks they had to perform. The tasks consisted in:

1. Create a new account and log in;

2. Set the EVA interface according to the requirements for each group: Group1 with both the transcript and the map toggled off, i.e., just YouTube video player (C1), Group2 with the *augmented transcript* toggled on and the map off (C2), Group3 with the *map of concepts' flow* toggled on and the transcript off (C3);

3. Watch Video1 trying to understand its content using if needed, the features available in the set-up (i.e., C2 and C3 hypervideo features, and the YouTube progress bar for C1);

4. Create a personal identification code to be used in the next steps, answer a few questions about learning and possible eyesight problems and past participation to other similar researches;

5. Answer a set of Comprehension Questions;

6. Repeat the same steps for Video2: watch and answer Comprehension Questions;

7. Answer a set of Subjective Questions about the user experience and the learning experience;

8. Watch Video3 without answering any questionnaire, this operation is preparatory to part two of the experiment.

Part 2

The aim of the second part of the study was to investigate the user experience in the rewatch learning context and the overall usability of the user interface. The sub-questions in this part were:

SQ3) How are EVA features perceived to support the first-watch learning experience?

SQ4) Are EVA features suitable to support the rewatch learning experience?

SQ1) What is the general user perception while using EVA tool?

Experimental Design

Part 2 of the study was structured in two sessions, aimed to address SQ3, SQ4 and SQ1 respectively.

2.1 The First Session was designed as a between-subject experiment with three conditions, like Part1: (C1) *YouTube video player* (C2) *augmented transcript* and (C3) *map of concepts' flow* without any visual cues as our control condition. Each participant used just one of the features. As with the first part, the scenario in this part can be mapped with "Second time viewing (rewatch)" provided to the experts for the Heuristic Evaluation.

2.2 The Second session involved all the participants, asked to evaluate the overall usability of the platform.

Procedure

The experiment has been conducted online, as a class activity, in subsequent later days thus reproducing the learning context of rewatch experience.

Participants were split in 3 groups, one for each experimental condition, and instructed about the tasks they had to perform.

The tasks consisted in:

– Session 2.1

    1. Log into the system using the account created in Part 1;

    2. Set the EVA according to the requirements for each group: Group1 with both the transcript and the map toggled off, i.e., just YouTube video player (C1), Group2 with the *augmented transcript* toggled on and the map off (C2), Group3 with the *map of concepts' flow* toggled on and the transcript off (C3);

    3. Rewatch Video3 trying to understand its content using if needed, the features available in the given set-up (i.e., C2 and C3 HV features, and the YouTube progress bar for C1). In the context of rewatch, students are instructed to feel free, in order to understand the content, to rewatch the video in their preferred mode, e.g. by skipping points and making use of the features or by watching the video in its entirely;

    4. Answer a set of Comprehension Questions;

    5. Answer to Subjective Questions about the user experience and the learning experience.

Participants were asked to use all the features described in the scenarios in order to complete a series of tasks. After that, they had to:

– Session 2.2

    1. Answer the questions of the System Usability Scale (SUS) Questionnaire [21];

    2. Answer the questions of the Post-Study System Usability (PSSUQ) Questionnaire [84].

5.4.2   Results

In this section, we report the results of Part 1 and Part 2 concerning the Comprehension Questions in first-watch and rewatch scenarios, the

Subjective Questions on User Experience in first-watch and rewatch scenarios, and the Usability tests (SUS and PSSUQ).

Results for Comprehension Questions

The test included, after the viewing of each educational video, a series of very simple, open-ended questions on the topics of the video to test immediate understanding. This procedure has been used in both the first watch learning context and the rewatch learning context. In this evaluation, the questionnaires were open-ended in order to allow the user full freedom in answering and to assess behaviour more closely when asked, for example, to answer questions with lists of several items, or definitions, acronyms, etc. To convert the textual answers into numerical data, each answer was assigned a score in a range from 0 to 1 based on the correctness and accuracy of the answer. As with the previous evaluation, to assess the effects of watching the video in the three different conditions, the average score for each condition was calculated.

- First-watch results

The results are as follows: C1 (AVG=0.74 SD=0.32), C2 (AVG=0.88 SD=0.21), C3 (AVG=0.86 SD=0.25), the figure 5.3 summaries them. In order to investigate statistical significance between C2 and C3 compared to C1 baseline we applied Mann-Whitney U test [180] for Two Independent Samples ith a significance criterion (alpha) of 0.05 and we found that both C2 and C3 results are significantly better than C1.

- Rewatch results

The rewatch results are: C1 (AVG=0.87 SD=0.28), C2 (AVG=0.91 SD=0.24), C3 (AVG=0.89 SD=0.24), see figure 5.4. As for the first watch learning context, in order to investigate statistical significance between C2 and C3 compared to C1 baseline we applied Mann-Whitney U test [180] for Two Independent Samples ith a significance criterion (alpha) of 0.05 and we found that both C2 and C3 results are not significantly better than C1.
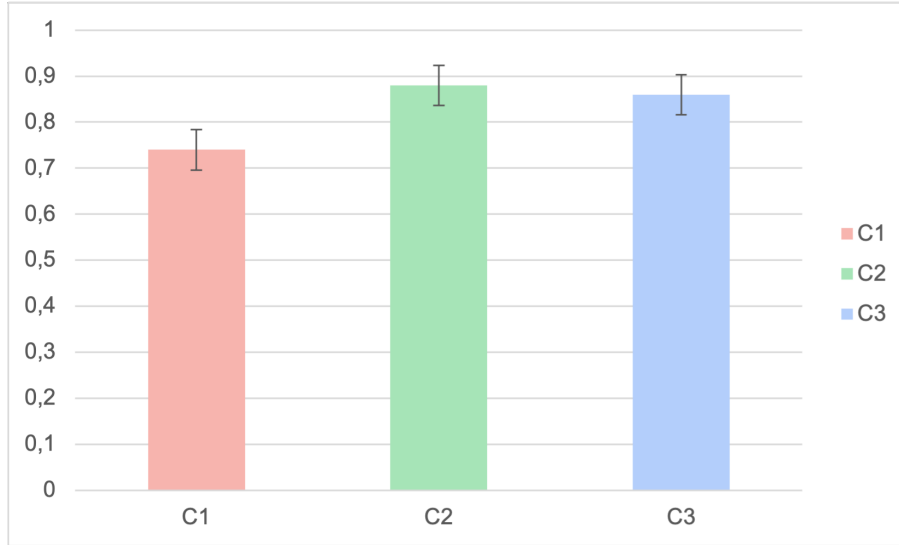
Figure 5.3: Results for video comprehension within first-watch learning context.

Results for Subjective Questions on User Experience (First-watch learning context)

After viewing the video and the comprehension questionnaire, users were given a user experience test to collect their opinions on their experience under the three conditions C1, C2 and C3. C2 and C3 questions are designed to be the same in order to compare the EVA features directly, using statistical analysis. C1 questions are used as control questions to be analyzed in relation to C2 and C3 answers for the first watch scenario. In correspondence with questions e.g. on the usefulness of having a support for the better understanding of the content of the video, in the case of C2 and C3, within the questionnaire for C1 users they are asked whether they feel the need for a support (e.g. transcript or map).

Figure 5.5 compares the results for C2 vs C3 concerning the common questions. The full description of the Subjective Questions are reported in Appendix A. We performed statistical analysis using Mann-Whitney U test [180] for Two Independent Samples. We used a two-tailed with a significance criterion (alpha) of 0.05.

The results show that both the features are perceived as useful to
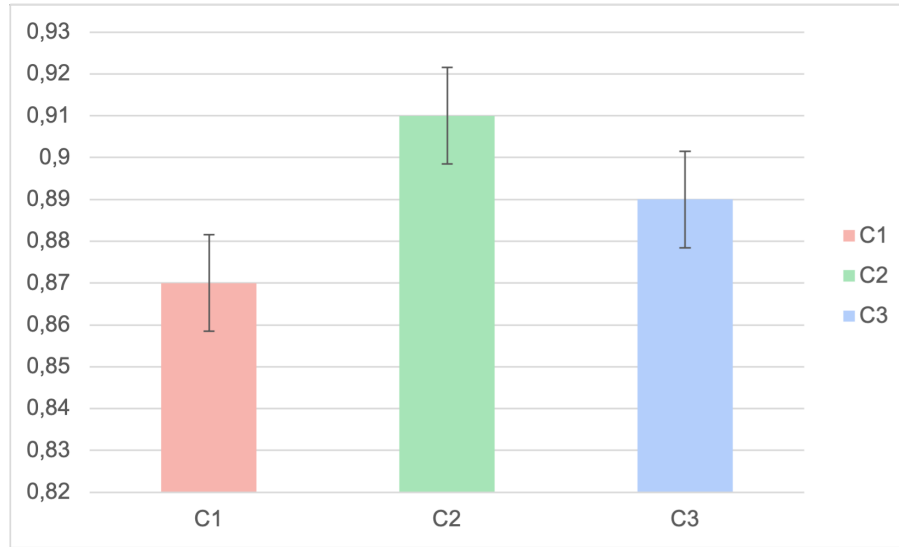
Figure 5.4: Results for video comprehension within rewatch learning context.

support the first watch learning experience, with a mean value over 3pt. The *augmented transcript* performed better, with a score that is over 4pt for all the questions aimed to investigate the support of the tool to understand and be focused on the video content (Q-C2(1).1, .2, .3 questions). In terms of usability (Q-C2(1).4, .5, .6 questions), the difference between the two features is less than 0.3, with the indication that probably the position of *the map of concepts' flow* in the bottom right-hand corner, compared to the transcript immediately to the side (right) of the video player, is less congenial for viewing in synchronisation with the video. However, the statistical analysis shows that none of the differences is statistically significant, except Q-C2(1).2 according to which the *augmented transcript* provides a greater support ($+0.77$) to understand the content in a non-native language ($U = 124.5, p < 0.05$). The answers of the students that used only the video players confirm that they would benefit from features that provide the *augmented transcript* (Q-C1(1).1 AVG:4) and outline the concepts in the first-watch context (Q-C1(1).1 AVG:4), while the perceived benefit from an interactive map synchronized with the video is a bit lower (Q-C1(1).1 AVG:3.5).
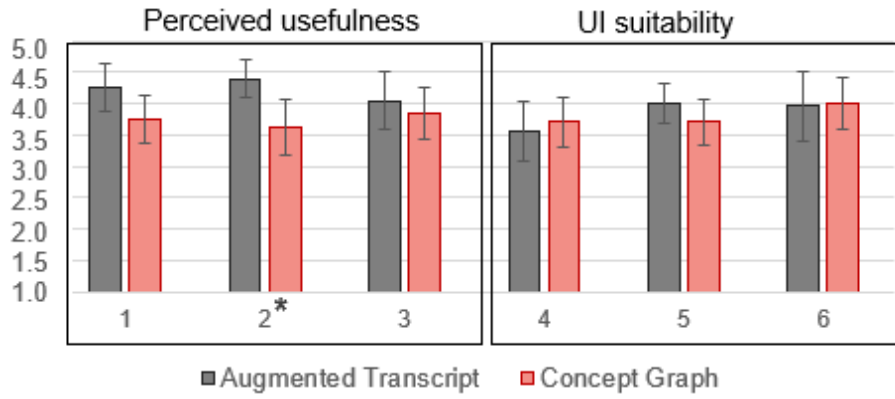
Figure 5.5: Results for Subjective Questions grouped for perceived usefulness (Q-C2/3(1).1-.3) and UI suitability (Q-C2/3(1).4-.6), in the *first watch* learning context.

Results for Subjective Questions on User Experience (Rewatch learning context)

The rewatch has learning context-specific characteristics that need to be analysed: in fact, while the first watch is mostly focused on viewing and comprehend, rewatch is focused on learning, thus involving deep understanding, exploration, navigation within the video content. The questions we designed are different in the three groups. Description of the Subjective Questions are reported in the Appendix A.
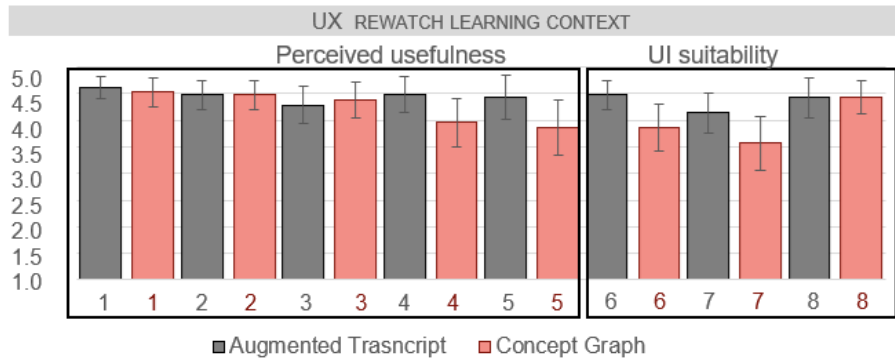
Figure 5.6: Results for Subjective Questions grouped for perceived usefulness (Q-C2/3(2).1-.5) and UI suitability (Q-C2/3(2).6-.8), in the *rewatch* learning context.

For the group under C1 condition, all the questions about the perceived benefit of EVA features in the rewatch phase obtain a score higher

than 4.2 (AVG:4,4) whereas standard deviation is lower than the case of perceived benefit of the features in the first watch scenario (average SD:0.85 vs. 1.04). The results obtained under C2 and C3 conditions seems confirmed by the results of C1.

Figure 5.6 provides an overview of all the scores, grouping them for user experience in the first watch and rewatch and usability of the user interface.

Within C2, the usefulness of the *augmented transcript* for the rewatch context obtains AVG:4.46, SD:0.7, while for C3, the usefulness of the *map of concept's flow* and related features obtains AVG:4.18, SD:0.91. Concerning the usability of both C2 and C3, results under C2 condition show an average result of AVG:3.68, SD:0.83), while results under C3 condition are AVG:4.14, SD:1.81.

Finally, for both C2 and C3 we added the common counter item to rate: "The EVA feature did not add any value to my experience of rewatching the video". It obtained high disagree (AVG: 1.57, SD:1.03) and disagree (AVG:2.14, SD:1.28) respectively.

Results for SUS and PSSUQ Questionnaire

In the case of the usability test, all the users recruited for the evaluation took the same test in order to assess the degree of satisfaction in using the platform through the standard SUS and PSSUQ metrics. The overall SUS score for EVA platform has been 82.9 (excellent, rate A).

Results for PSSUQ Questionnaire are as follows: Overall AVG: 3 on a 7pt-scale (1 agree-7 disagree). The scores for the three dimensions of the tests are: System Usefulness: 2.8, Information Quality: 2.6, Interface Quality: 3.75.

We also investigated if any effect could occur on the perceived usability, based on the group the participant belonged for conditions C1, C2, C3. However no significant difference has been found neither for SUS nor for PSSUQ results, computed using Mann-Whitney U test for each couple of conditions: C1-C2, C2-C3, C1-C3.

5.4.3   Discussion

The average results, that are above 4 points for both the transcript and the map of concepts' flow in both the first watch and rewatch learning contexts and as well as the usability test, appear to demonstrate very good results for the user experience and learning experience using EVA platform. As in the other evaluations the study involved the students of a university course, it was run as part of a class activity in order to obtain more reliable results in terms of learning experience. In order to avoid, or at least reduce, complacency biases, the activity was carried out anonymously. To address the research questions for this user study, a more in-depth analysis is provided and necessitates debate. With this evaluation, we collected data to answer 5 of the 6 sub-questions we detailed at the beginning of this chapter. Let us therefore see in detail how to answer each individual SQs.

SQ1) What is the general user perception while using EVA tool? Considering the results from the Usability tests, an overall excellent usability rate is achieved according to SUS, but a lower rate if we consider PSSUQ (that splits the results in three dimensions). This lower result is obtained for the user interface. This outcome is in line with the results of the Expert Evaluation, which offered recommendations for enhancing the user interface and highlighted in particular the need for more user control and for system feedback regarding the tool's functionalities. This is especially important for the *map of concepts' flow* because it has more features and students are not always accustomed to such kind of features. Another interesting aspect to note is that this user study did not take into account users with learning difficulties or disabilities. Although these features are often used precisely as a support for various types of disability, in other cases they may be considered an obstacle or barrier for users with e.g. motor limitations, visual impairments and screen vision problems (e.g. epilepsy).

SQ2) Are EVA features useful for encouraging immediate understanding of the content of an educational video within the first-watch learning context?

Within first-watch learning context, judging from the results obtained for the immediate understanding of the content of the video, data obtained by means of comprehension tests of the content of the video just watched by the users during the experiment, the results look promising for both EVA features, with the transcript obtaining an overall score of 0.88 and the map a score of 0.86 compared to 0.74 for unsupported viewing. The Mann-Whitney U test also confirms that there is a significant difference in the comparisons between C1 and C2 and C1 and C3. This endorses the results also obtained in the previous evaluation.

SQ3) How are EVA features perceived to support the first-watch learning experience?
Results show very good perceived usefulness of both the EVA features with higher, but not statistically significant, performance for the *augmented transcript*. The only case where it the difference is significant is for the support provided by the transcript in case of video content in a different language from the one of the user.

SQ4) Are EVA features suitable to support the rewatch learning experience?
Within the rewatch scenario, support for an immediate understanding is not to be investigated, as users are precisely in a secondary viewing phase of the video. It must be said that having set the second viewing two days later, many concepts seen only once may still be obscure and unclear to remember. In this regard, the support of the EVA features seems to be of help, judging by the results obtained from the comprehension questions answered by the users. In fact, the transcript has an average of 0.91, the map an average of 0.89 compared to the unsupported viewing which has an average of 0.87. Compared to the results of the first-watch, it must be said that there are no statistically significant differences, but the trend of the other trial and also of part 1 of this current one confirms that the transcript is a very supportive tool in both the first-watch and rewatch phases. The map still obtains very good scores, but evidently due to limitations due to the visualisation, also highlighted by the usability tests, it still obtains promising results

but a little lower than those of the transcript.

SQ5) How are EVA features perceived to support the rewatch learning experience?

While the results of the comprehension questions obtained more evenly spread scores for the three conditions C1, C2 and C3, for perceived usefulness, for users under C1 condition, it is interesting to note that the average scores of the perceived usefulness of EVA feauters in the rewatch context is higher than the usefulness perceived in the first watch context, especially for those related to the map (AVG: 4.4). This reveals a great need for this kind of support.

Considering the responses to the subjective questions about the actual usefulness of EVA features higher average score (AVG: 4.46) under C2 and a slightly lower average value (AVG:4.18) under C3 are achieved. These results confirm both the need of interactive tools for video-based learning and show that EVA features mostly satisfy this need, even though there is still room for improvement, in order to fill the gap between expected support for rewatch (4.4) and the actual perceived usefullness (4.18).

The control question to check the reliability of the other questions under both C2 and C3 conditions: "The transcript/the map did not add any value to my experience of rewatching the video" obtained high disagree (AVG: 1.57, SD:1.03) and disagree (AVG:2.14, SD:1.28) respectively in C2 and C3 conditions. This is coherent with the other results, confirming that there is room for improvement for C3.

5.5   Third Evaluation with Students

The aim of this experimentation with users is to expand the pool of students to have more data, in order to test the tool in a different domain. The experiment exploits, starting from the scenarios designed for heuristic evaluation, the possibility to test the platforms in the archaeological domain.

5.5.1 Method

This user study is divided into two parts:

1. The first part (Part 1) has been designed as a user experiment comparing between the "Augmented Transcript"-only and "Map of Concepts" flow'-only features to investigate the user experience of this learning environment, within the "first-watch" learning context.

2. The second part (Part 2) has been designed to evaluate the interface within the "rewatch" learning context.

Below, we first describe the participants and the setting for the user study that are common in the two parts, then we provide the details of the study design of each part.

Participants

For this second experiment we recruited 22 students, attending the course in "Methods of archaeological research" in Archaeology from the University of Pisa, at the faculty of Archaeology. They were 14 females and 8 males, in an age range between 18 and 27 ($AVG : 21.1, SD : 2.8$). All of them have at least intermediate English skills, and no one is native English speaker. None of the participant declared learning difficulties or eyesight problems.

Setting

For this experiment participants are required to use the developed web user interface (video player, the video transcript, the map of concept's flow). For the experiment, to make sure that the evaluation of the EVA features is not compromised by the accuracy of the knowledge extracted from the video have been post-processed after their automatic extraction.

Part 1

The aim of the first part of the study was to evaluate and compare, as in the previous study, the two EVA features tools: *augmented transcript* and *map of concepts' flow* within the archaeological domain. The purpose of this part is to answer SQ2 and SQ3 in 5.1, in the specific archaeological domain.

SQ2) Are EVA features useful for encouraging immediate understanding of the content of an educational video within the first-watch learning context?

SQ3) How are EVA features perceived to support the first-watch learning experience?

Experimental Design

This part of the study was designed as a between-subject experiment with three conditions: (1) *augmented transcript*, (2) map of concepts' flow, and (3) *YouTube video player* without any visual cues as our control condition. Each participant used only one of the tools.

The scenario is analogous to the ones described in the previous evaluations. Specifically, Part 1 of the user study matches the "First-time viewing (first watch)" scenario provided to the experts, with the difference that each user was required to use just one of the features, among the three conditions (as described below). As in the previous experiment the videos were the same for all the students in order to compare the results. They are two YouTube videos[67], made available on the EVA platform. All of the videos are in English.

As dependent variable for this study we used the subjective feedback in each condition, collected through questionnaire responses.

---

[6]Video1: https://youtu.be/eGngPn2KoPo
[7]Video2: https://youtu.be/mU5$_{e}$eH9QpM

Procedure

The experiment has been conducted in hybrid mode. The students were in the classroom in attendance, with the course lecturer supervising the activity. On the other hand, the instructor from the EDURELL team was online to make the introduction to the tool and explain all the various tasks to be performed. Each student had his or her own laptop and headphones to isolate themselves and perform the activity alone.

– The EVA platform was first explained to the participants, and then they were asked to test out the user interface and features. To achieve this, users were instructed to play one of the video already included on the platform, use each tool, and alternate between the map of concepts' flow and the augmented transcript. Participants utilized each instrument for around five minutes.

– Participants were then split in 3 groups, one for each experimental condition, and instructed about the tasks they had to perform. The tasks consisted in:

1. Create a new account and log in;

2. Set the EVA interface according to the requirements for each group: Group1 with both the transcript and the map toggled off, i.e., just YouTube video player (C1), Group2 with the *augmented transcript* toggled on and the map off (C2), with the *map of concepts' flow* toggled on and the transcript off (C3);

3. Watch Video1 trying to understand its content using if needed, the features available in the set-up (i.e., C2 and C3 hypervideo features, and the YouTube progress bar for C1);

4. Create a personal identification code to be used in the next steps, answer a few questions about learning and possible eyesight problems and past participation to other similar researches;

5. Answer a set of Comprehension Questions;

6. Answer a set of Subjective Questions about the user experience and the learning experience;

7. Watch Video2 without answering any questionnaire, this operation is preparatory to part two of the experiment.

Part 2

The aim of the second part of the study was to investigate the user experience in the rewatch learning context. The sub-questions in this part were:

SQ3) How are EVA features perceived to support the first-watch learning experience?

SQ4) Are EVA features suitable to support the rewatch learning experience?

Experimental Design

The session was designed as a between-subject experiment with three conditions, like Part1: (1) *augmented transcript*, (2) *map of concepts' flow*, and (3) *YouTube video player* without any visual cues as our control condition. The scenario for this part of the user study was analogous to the one named "Second time viewing (rewatch)" provided to the experts for the Heuristic Evaluation.

Procedure

As for Part 1 the experiment has been conducted in hybrid mode. The students were in the classroom in attendance, with the course lecturer supervising the activity. On the other hand, the instructor from the EDURELL team was online to make the introduction to the tool and explain all the various tasks to be performed. Each student had his or her own laptop and headphones to isolate themselves and perform the activity alone. Since the activity was designed as a substitute for an end-of-course quiz that the lecturer gives to his students, the assessment was carried out on the same day. Therefore for the rewatch, unlike the

previous evaluation for which we had set a distance of two days, in this case it took place approximately two hours after the first viewing of the video. This difference will be taken into account in the discussion.

Participants were split in 3 groups, one for each experimental condition, and instructed about the tasks they had to perform.

The tasks consisted in:

1. Log into the system using the account created in Part1;

2. Set the EVA according to the requirements for each group: Group1 with both the transcript and the map toggled off, i.e., just YouTube video player (C1), Group2 with the *augmented transcript* toggled on and the map off (C2), with the *map of concepts' flow* toggled on and the transcript off (C3);

3. Rewatch Video 2 trying to understand its content using if needed, the features available in the given set-up (i.e., C2 and C3 hypervideo features, and the YouTube progress bar for C1). In the context of rewatch, students are instructed to feel free, in order to understand the content, to rewatch the video in their preferred mode, e.g. by skipping points and making use of the features or by watching the video in its entirely;

4. Answer a set of Comprehension Questions;

5. Answer to Subjective Questions about the user experience and the learning experience.

5.5.2 Results

In this section, we report the results of Part 1 and Part 2 concerning the Comprehension Questions in first-watch and rewatch scenarios and the Subjective Questions on User Experience in first-watch and rewatch scenarios.

Results for Comprehension Questions

The test comprised a questionnaire of a few simple, multiple-choice questions to assess the user's immediate understanding of the content of the video they had just watched. This procedure has been used in both the first-watch learning context and the rewatch learning context. Unlike the previous test for which they opted for open-ended questions, for this test the choice fell on multiple-choice questions in order to have a clearer indication of the results. Without the need for conversion into numerical values, each incorrect answer was automatically assigned a zero and each correct answer a one.

This led to some encouraging results (figure 5.7) in terms of overall averages in the first watch scenario: C1 (AVG=0.62 SD=0.27), C2 (AVG=0.89 SD=0.13), C3 (AVG=0.75 SD=0.13). Nevertheless we found that both C2 and C3 results are not significantly better than C1. In order to investigate statistical significance between C2 and C3 compared to C1 baseline we applied Mann-Whitney U test [180] for Two Independent Samples with a significance criterion (alpha) of 0.05. With regard to the rewatch the answers were all correct in all three conditions. For this reason, no significance test was carried out; we will elaborate on this result in more detail in the discussion.
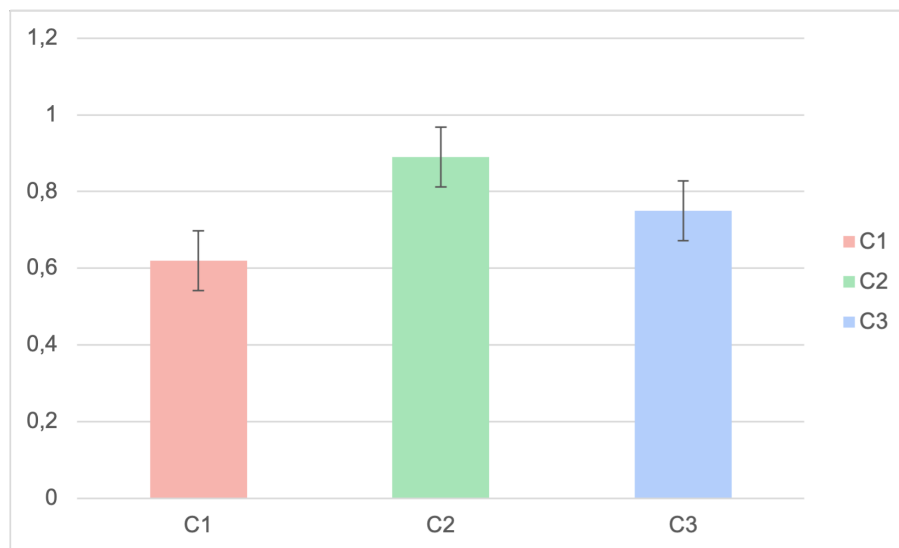
Figure 5.7: Results for video comprehension within first-watch learning context.

Results for Subjective Questions on User Experience (First-watch learning context)

As in the previous evaluation, users answered a user experience test with the aim to collect their opinions on the experience under the three conditions C1, C2 and C3. C2 and C3 questions were designed to be the same in order to compare the EVA features directly, using statistical analysis. With regard to the group questionnaire in C3, additional questions, specific to the type of support, were included. Full details of the questionnaires can be found in the Appendix B. C1 questions are used as control questions to be analyzed in relation to C2 and C3 answers for the first-watch scenario. We performed statistical analysis using Mann-Whitney U test [180] for Two Independent Samples. We used a two-tailed with a significance criterion (alpha) of 0.05. A comparative analysis of the Qs common to C2 and C3 (figure 5.8) shows that the transcript remains highly rated by users, with all questions averaging over 4 points. As for the map, the averages are equally positive but with a somewhat low result (AVG 2.63) for the Q relating to distraction. This result is also confirmed by the significance test, as it is on this Q that there is a statistically significant difference between C2 and C3.
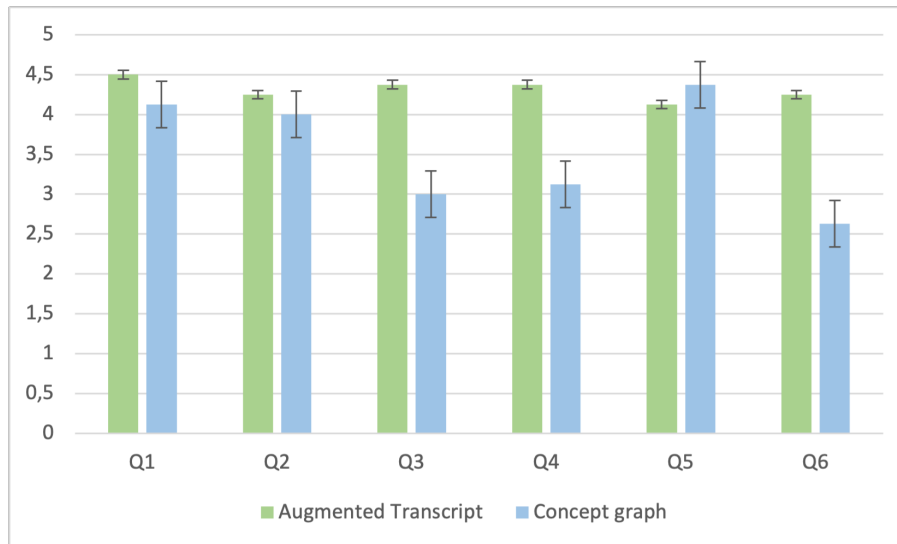
Figure 5.8: Results of subjective feedback questions in common between C2 and C3 within first-watch scenario

With regard to the user questionnaires under C1, they confirmed the results of the previous evaluation, expressing interest in receiving support from tools that provide the augmented transcript (AVG:4), concepts outlined in a map (AVG:4.5), and from an interactive map synchronized with the video (AVG:4.5).

Results for Subjective Questions on User Experience (Rewatch learning context)

As with the first-watch scenario, the questions for the rewatch scenario are differentiated according to the group. C2 and C3 are for an overlapping number of Qs to allow comparisons with significance tests, in the case of C3 specific questions were also included to explore the user experience of map peculiarities. The questions in C1 are aimed at understanding the need for support by those who tested only with the simple videoplayer. Concerning C1, questions exploring the possibility of increased support while watching the video obtained average scores from 3.8 to 4.6. The latter was obtained by Q9 *"I would benefit from a system that makes it easy for me to search for concepts within the video"*.
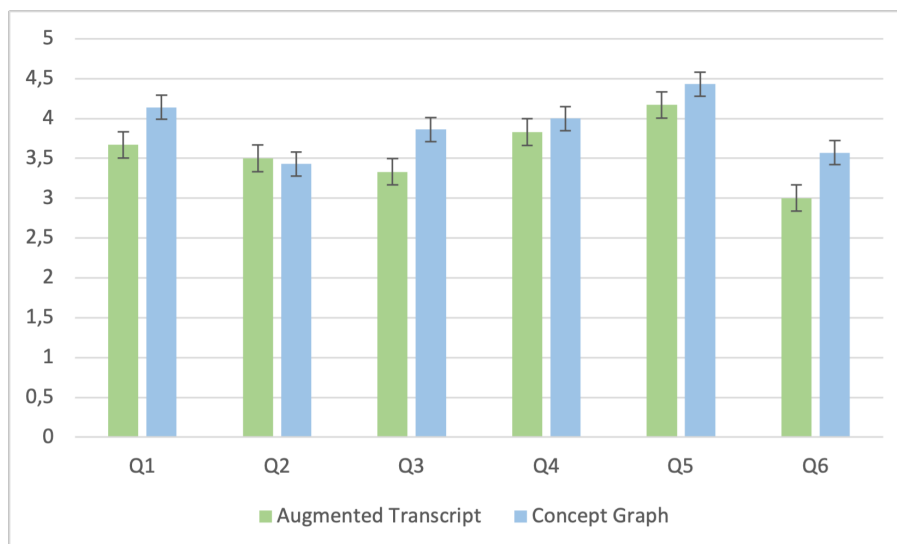
Figure 5.9: Results of subjective feedback questions in common between C2 and C3 within rewatch scenario.

In the group below C2, we notice lower generic scores compared to the first-watch scenario, ranging from AVG 3 to AVG 4.17, while C3 scores better (AVG 3.43 to AVG 4.43). However, the statistical analysis shows that none of the differences is statistically significant. In fact, we applied for these comparisons the Mann-Whitney U test [180] for Two Independent Samples and all attempts were non-significant. In figure 5.9 we can see in detail a comparison of common Qs in C2 and C3.

5.5.3 Discussion

This third evaluation with users involved for the first time a group of students from a domain other than "computer science", namely "archaeology". The intention was to test the system on different domains to see if there could be any substantial differences in performance. As in previous evaluations the study involved the students of a university course and it was run as part of a class activity in order to obtain more reliable results in terms of learning experience. In order to avoid, or at least reduce, complacency biases, the activity was carried out anonymously. The limitation of this experiment compared to the previous one is certainly related to the smaller numbers, which therefore allows less room for generalisation. Even from a qualitative approach, however, we can confirm the results of the previous experiments, with a general appreciation for the EVA features as support during video consumption. Let us now look in detail at how to answer the SQs detailed at the beginning of the chapter.

SQ2) Are EVA features useful for encouraging immediate understanding of the content of an educational video within the first-watch learning context?

As the previous experiments, in the first-watch scenario the support of EVA features always proves useful for immediate understanding, which is tested through very simple quizzes given to the students at the end of the viewing. Also in this experimentation, the *augmented transcript* is confirmed as the best support, with the possibility of viewing the transcript of the speech in real time and with the visualisation of the salient

concepts highlighted. This enables a better understanding of the content, especially in a language other than the mother tongue. The map performed equally well but probably suffers from those usability limitations in the area of visualisation that we also found in the feedback questionnaires. In general, however, compared to viewing without any support, the correct response rates confirm to us that EVA is supportive for the students.

SQ3) How are EVA features perceived to support the first-watch learning experience?
User perceptions of the use of EVA features during the first-watch scenario are generally positive. Confirming the results of previous experiments. Also in this case the transcript scores better and in the case of the Q concerning distraction we observe a statistically significant difference between the *augmented transcript* and the *map of concepts' flow*. This result suggests us that the map is a fairly complex and visually new tool for the learner, which in the case of synchro visualisation can lead to distraction. Future developments will have to take this into account and possibly design a simplified type of map to be used as a support during navigation and an expanded and detailed map for the post-viewing phase.

SQ4) Are EVA features suitable to support the rewatch learning experience?
The results on video comprehension in the rewatch scenario were a surprise and suffer from some limitations. In fact, the answers to the questionnaires were all positive in all 3 viewing conditions. The reasons for this result can be several: firstly the short length of the video and the simplicity of the content influenced the results. In fact, the students, with prior knowledge of the topics, being at the end of the academic semester course, were able to answer the questions independently with repeated viewing of the video. Another factor to be taken into account is the limited number of participants in the trial (22 people) compared to the previous one, and perhaps the choice of the multiple-choice questionnaire, which in case of doubt about an answer can help trigger memory,

unlike the open-ended answer.

SQ5) How are EVA features perceived to support the rewatch learning experience?

Although there were no differences in terms of performance on video comprehension, we note a general appreciation of users for the EVA features on rewatch. In this case, in contrast to the data of all previous experiments, it is the map that receives the highest average scores. The video used for Part 2 of the experiment, was about 3 minutes long and the resulting extracted map presented a few simple concepts well schematised, evidently as also hypothesised previously a more compact map support is more congenial to the student. This cue may support future improvements of the platform.

5.6 EVAnn Evaluation

In this last section of the chapter, we introduce a further experiment carried out within the EDURELL Framework, namely the evaluation of automatic extraction methods, which were implemented within the EVAnn tool [157]. In particular, it was necessary to evaluate the extraction of prerequisite maps compared with manually annotated maps. This evaluation answers the last sub-question defined at the beginning of this chapter, namely:

SQ6) Is the method implemented within the EDURELL Framework useful for automatically extracting prerequisite relations from videos?

5.6.1 Method

In order to make this comparison with manual annotations, a "gold" annotation [170] of five manual annotations, made by different expert users, was first calculated. As for the baseline methods to be compared with the burst results, three methods were considered, using different metrics. Let us see the methods in detail:

- *Method 1 - Hyponyms, hypernyms and meronyms (HH)*

In section 4.6.3 we introduced several methods used in the literature to extract prerequisite maps, mainly from textbooks. Among these, one of the most exploited methods is the one that exploits the linguistic properties of a text, with hypernymy-hyponymy and meronymy relations [164]. Therefore, when considering these types of linguistic relations, it is possible to trace prerequisite relations within them.

- *Method 2 - Reference Distance (RefD)*
  Another method introduced in section 4.6.3 is the RefD (reference distance) [87] method, i.e. how two concepts refer to each other. This method calculates, considering a threshold, the distance between two concepts within the text.

- *Method 3 - Wikipedia pages (WP)*
  The last baseline method exploits Wikipedia's external resource [164], using three sets of features extracted with Wikipedia in order to obtain the prerequisite relations. The features are:

  1. *Usage*, which captures whether a concept is used in another concept's definition.

  2. *Content Similarity*, which exploits the lexical similarity between Wikipedia concepts using cosine similarity between the concept vector and article vector.

  3. *Learning Level*, which measure whether a concept has a lower learning level and should be learned first.

5.6.2   Results

The metrics used to evaluate the performance of automatic extraction from videos are precision, recall and F1 score [123] and Vertex Edge Overlap (VEO) for graph similarity. In the figure below (5.10) the comparison between baseline methods and the Burst is illustrated. The results show that the Burst method performs better than RefD in both precision and recall, it has an higher precision compared to the Wikipedia

pages, and finally, it has a lower precision than the hyponyms method but a bigger recall. However, considering the F1-score and VEO the Burst method outperforms the others.
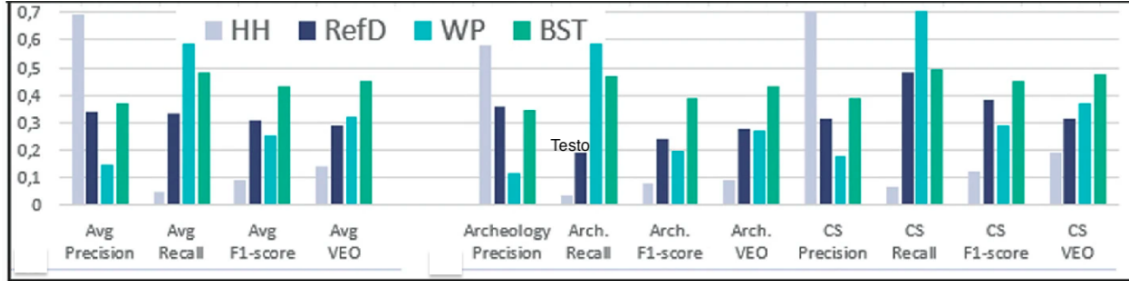
Figure 5.10: Results of the evaluation between the various methods.

Finally, to further explore whether there are differences between the algorithms used to textbook vs. transcript, we examined the methods' performance on a manually annotated textbook chapter. Results shows an average non-significant decrease of all the metrics (AVG -12%) using transcript.

5.6.3   Discussion

The evaluation of automatic methods showed promising results, with the method developed within the EDURELL Framework performing better than the baseline methods according to the F1 score and VEO metrics. Actually, we expected HH to perform better in terms of accuracy and WP in terms of recall. Another consideration to be made is that for both domains, "computer science" and "archaeology", there are no significant differences from the comparison. In general, even the metrics that perform better than the baselines, i.e. F1 score and VEO, do not reach the 0.5 threshold. This limitation is therefore to be emphasised, with the results certainly improving with refinements (such as video analysis and synonym handling, which was not carried out at this stage), and considering that these are unsupervised automatic methods, it is still an acceptable score [157].

Chapter 6

Conclusions, Limits and Future Work

The thesis work that has been presented leads us to a varied set of considerations, also in light of the experiments and evaluations gathered in Chapter 5. Many of the discussions related to each experiment that we have already enunciated in the previous chapter lead us to draw conclusions with an overview of the EDURELL Framework.

We presented the EDURELL Framework, which includes several interactive visual tools designed with the goal of improving video-based learning by providing a structure to video content and enhanced HV capabilities. This goal of building interactive HV tools for learning support is achieved through a process of extraction of concepts and relations from video lectures and represent them as a knowledge graph. From the work done on the literature review, it emerged the need for tools aimed at improving the path along the learning process by strengthening navigation capabilities through interactivity. One of the major issues proved to be the linear and sequential viewing of the video and the lack of a structure that could improve the organisation of the content. In many cases, this augmentation activity can also be provided in popular platforms such as YouTube, but with additional work on the part of the designer of a video lesson: such as adding subtitles (which can actually also be automatically added on YouTube but often with deficiencies in punctuation

and linguistic correctness in the case of readers speaking in a language other than their mother tongue), structuring in chapters, etc. The work behind this project therefore moved along two main strands, on the one hand the need to highlight and bring out the needs to enable the design of tools that could help in video consumption, and on the other hand trying to provide these tools also by means of a manual and automatic and semi-automatic extraction system. During this long project, which has involved numerous actors so far, it has been necessary to test the platforms and methods developed through different types of evaluations. First and foremost, the heuristic evaluation carried out by the experts gave us first and fundamental insights into the usability of the EVA tool. In fact, if the usability of a tool is poor, it already represents the first obstacle in the correct use of a functionality. These hints led to the so-called iterative process of continuously improving even very small features of the tool, whenever it was to be evaluated. Hence, three user evaluations were carried out to investigate not only the usability aspect and the related satisfaction in using the tool, but also ad hoc customised questionnaires to investigate the perceived usefulness of these features by the users and and finally the detection of video comprehension levels with or without support. All this took place within two different learning contexts, which had also been previously identified by the literature survey, namely the first watch and rewatch scenario.

The decision to differentiate the domains and act on the double track of computer science and archaeology was due to the fact that the framework, as we have often pointed out, by exploiting the knowledge inherent in video, wanted to overcome the limitations of lack of support for some specific domains. Thus, testing the platform on two quite different domains, one coming from the STEM world and the other from the non-STEM world, could give us an idea of its applicability in different areas of knowledge, without, of course, claiming to generalise. In fact, one of the possible future developments will be to broaden the user base on a large scale to have more data to support our hypotheses.

The first experimentation, carried out only in the context of the first

watch, had already given promising results in terms of average scores on video comprehension and in general on the perceived usefulness of EVA's features. What emerged already as a first limitation during the first-watch phase was that the map was less welcome than the augmented transcript, which proved to be a very useful support, right from the start, during the synchronised viewing of the video. The map, which probably presented a complex structure at that stage, with an expanded graph, was more difficult to consult in interactive mode during navigation. The perception from the questionnaires was that it is seen more as a tool to be used at a later stage during the learning process in order to better grasp concepts and the relationships between them. However, in order to improve its usability and visualisation after these two evaluations, some modifications were made.

One of the most important changes made to improve the map was therefore to try to deal with synonyms, not included in the first version. The management of synonyms is quite complex and is generally an open issue in linguistics, because no matter how overlapping two terms may be, given the complexity of language and its various facets, the slightest ambiguity may always arise. In our case, we have decided to be guided strictly by the domain, considering as synonyms terms that might possibly not be so in common language and vice-versa, supported by domain-specific thesauri. This also practical handling has resulted in the creation of synsets of terms, including, for example, acronyms, which in the map display have been associated with the same node. In this way, the map of concepts' flow was lightened with fewer nodes and a hoped-for improvement in terms of information visualisation. The handling of synonyms also led to changes for the transcript with different colouring in the case of the concept in focus (blue) and its synonym (green). On the strenght of these changes the second and third evaluations were carried out with a larger number of users. In particular, during the second evaluation, it was decided to test the entire system according to usability standards by submitting the SUS and PSSUQ questionnaires, which are a recognised measure in the literature. The two evaluations

were also important because we tested the rewatch learning context for the first time. The findings show that the transcript and the map tools are perceived as useful tools to address such needs and the overall user experience is good. The results of the second and third evaluations were in line with the previously collected data, showing a consistently higher acceptance of the transcripts than the map, not only in terms of the perceived usefulness for users but also for the results obtained by them with quizzes for video comprehension. Overall, the two features consistently scored better than watching videos without support.

However, user experience questionnaires from the third evaluation as part of the rewatch showed that users liked the map more than the transcript. This entices us, despite the limitations and the work that will still need to be done to improve the system, to think that this tool is perceived to be supportive during the in-depth stage of specific concepts and at a more advanced stage of consulting educational materials, during the process along the learning path. Probably, although we had conceived the map as an additional tool to improve navigation, what will have to be exploited as a suggestion for future work is to hive off its function. Already with the additional tab of the "detailed description" offering the display of a sub-graph we were trying to go in that direction. Probably a more compact and interactive "path graph" will be able to help more in the immediate visualization of concepts, and the complete prerequisite map can perhaps be consulted as an in-depth tool at a later stage. This is to limit the issue of distraction that has arisen in the various evaluations.

Another complementary part of the work was the processing of the videos by manual annotation, automatic and semi-automatic extraction of concepts and relationships between them. The work as we also anticipated in the introduction of this thesis, comes from years of research in the field of textbooks. The application of such methods on video content required further research in order to adapt it to the peculiar characteristics of video and exploit them. The results of automatic extraction (see section 5.6) were encouraging in comparison with methods considered

as baselines. Even in this strand, limitations have emerged because we still stay at average results below 0.5, which leads us to want to address new challenges for the improvement of extraction, possibly exploiting video analysis. To the best of our knowledge an attempt at automatic extraction of prerequisite maps from videos is still a novelty factor in research, so despite the inevitable limitations it certainly represents an interesting challenge for future project developments. By automating the development of this kind of interactive HV tools, the goal is to make them available in educational learning environments and MOOCs.

Appendix A

Subjective Questionnaires (1)

The questions relating to the subjective questionnaires submitted to the users during the second evaluation are shown below.
First watch (1) and rewatch (2) experience.
5pt-Likert scale (1 disagree-5agree).

Condition C1: Augmented Transcript and Concept Graph toggled off, i.e., YouTube-like video player

1. First watch scenario	AVG	Dev ST
Q-C1(1).1 I would benefit from a system that would let me view an interactive transcript of the video while watching	4	1.21
Q-C1(1).2 I would benefit from a system that allows me to see the concepts outlined in a map	4	0.99
Q-C1(1).3 I would benefit from a system that allows me to view an interactive map synchronized with the video	3.5	0.93
2. Rewatch scenario	AVG	Dev ST
Q-C1(2).1 I would benefit from a system that would allow me to easily search for concepts within the video	4.25	1.07
Q-C1(2).2 I would benefit from a system that would allow me to see the concepts in a map.	4.5	0.83
Q-C1(2).3 I would benefit from a system that would allow me to view pop-ups so that I could find concepts that I do not remember	4.35	0.81
Q-C1(2).4 I would benefit from a system that summarizes in a tab the descriptions of a concept found within the video	4.6	0.68
Q-C1(2).5 I would benefit from a system that would break the video automatically into smaller, clickable portions of the video like a sort of index to a book	4.3	0.86

Figure A.1: Second Evaluation with students: C1 condition

Condition C2: Augmented Transcript-only toggled on

1. First watch scenario	AVG	Dev ST
Q-C2(1).1 The augmented transcript helped me understand the content of the video	4.26	0.92
Q-C2(1).2 The augmented transcript supported me in understanding content in a language other than my native language	4.39	0.72
Q-C2(1).3 The augmented transcript helped me keep my attention on the content while watching the video	4.04	1.11
Q-C2(1).4 The augmented transcript is in the right position within the interface	4.00	0.80
Q-C2(1).5 The augmented transcript runs too fast	2.43	1.16
Q-C2(1).6 The augmented transcript distracted me	2.04	1.33
2. Rewatch scenario	AVG	Dev ST
Q-C2(2).1 The augmented transcript helped me deeply understand the content of the video	4.62	0.50
Q-C2(2).2 The augmented transcript supported me in understanding content in a language other than my native language	4.48	0.68
Q-C2(2).3 The augmented transcript helped me recall the concepts while rewatching the video	4.29	0.85
Q-C2(2).4 The augmented transcript made it easier for me to navigate the video in the "rewatch" phase	4.48	0.81
Q-C2(2).5 The augmented transcript did not add any value to my experience of "rewatching" the video	1.57	1.03
Q-C2(2).6 The augmented transcript is in the right position within the interface	4.48	0.68
Q-C2(2).7 The augmented transcript runs too fast	1.86	0.91
Q-C2(2).8 The augmented transcript distracted me	1.57	0.93

Figure A.2: Second Evaluation with students: C2 condition

Condition C3: Map of concepts' flow-only toggled on

1. First watch scenario	AVG	Dev ST
Q-C3(1).1 The map of concepts' flow helped me understand the content of the video	3.75	0.94
Q-C3(1).2 The map of concepts' flow supported me in understanding content in a language other than my native language	3.63	1.10
Q-C3(1).3 The map of concepts' flow helped me keep my attention on the content while watching the video	3.83	1.01
Q-C3(1).4 The map of concepts' flow is in the right position within the interface	3.71	0.91
Q-C3(1).5 The map of concepts' flow flows too fast	2.29	0.95
Q-C3(1).6 The map of concepts' flow was distracting	2.00	1.02

2. Rewatch scenario	AVG	Dev ST
Q-C3(2).1 The map of concepts' flow helped me navigate the video concepts	4.52	0.68
Q-C3(2).2 The map of concepts' flow made it easier for me to find the "definition" of a concept	4.48	0.68
Q-C3(2).3 The map of concepts' flow made it easier for me to find the "in-depth" description of a concept	4.38	0.80
Q-C3(2).4 The "Search for Concepts" supported me in learning more about a specific concept	3.95	1.12
Q-C3(2).5 The map of concepts' flow added no value to my experience of "rewatching" (i.e., second viewing) the video	2.14	1.28
Q-C3(2).6 The "Search for Concepts" bar is placed in the right position within the interface	3.86	1.06
Q-C3(2).7 The tab that appears with the "Search for Concepts" was suitable for the task	3.57	1.25
Q-C3(2).8 The nodes on the map of concepts' flow are easily navigable and clickable	4.43	0.75

Figure A.3: Second Evaluation with students: C3 condition

Appendix B

Subjective Questionnaires (2)

The questions relating to the subjective questionnaires submitted to the users during the third evaluation are shown below.

First watch (1) and rewatch (2) experience.

5pt-Likert scale (1 disagree-5agree).

Condition C1: Augmented Transcript and Concept Graph toggled off, i.e., YouTube-like video player

	AVG	Dev ST
***1.** First watch scenario*	*AVG*	*Dev ST*
Q-C1(1).1 I would benefit from a system that would let me view an interactive transcript of the video while watching	4.00	0.89
Q-C1(1).2 I would benefit from a system that allows me to see the concepts outlined in a map	4.50	0.84
Q-C1(1).3 I would benefit from a system that allows me to view an interactive map synchronized with the video	4.50	0.84
2. Rewatch scenario	AVG	Dev ST
Q-C1(2).1 I would benefit from a system that would allow me to easily search for concepts within the video	4.60	0.55
Q-C1(2).2 I would benefit from a system that would allow me to see the concepts in a map.	4.20	0.84
Q-C1(2).3 I would benefit from a system that would allow me to view pop-ups so that I could find concepts that I do not remember	4.40	0.85
Q-C1(2).4 I would benefit from a system that summarizes in a tab the descriptions of a concept found within the video	4.40	0.89
Q-C1(2).5 I would benefit from a system that would break the video automatically into smaller, clickable portions of the video like a sort of index to a book	3.80	0.84

Figure B.1: Third Evaluation with students: C1 condition

]

Condition C2: Augmented Transcript-only toggled on

1. First watch scenario	AVG	Dev ST
Q-C2(1).1 The augmented transcript helped me understand the content of the video	4.50	0.76
Q-C2(1).2 The augmented transcript supported me in understanding content in a language other than my native language	4.25	1.04
Q-C2(1).3 The augmented transcript helped me keep my attention on the content while watching the video	4.38	0.92
Q-C2(1).4 The augmented transcript is in the right position within the interface	4.12	0.83
Q-C2(1).5 The augmented transcript runs too fast	1.62	0.74
Q-C2(1).6 The augmented transcript distracted me	1.75	0.89
2. Rewatch scenario	AVG	Dev ST
Q-C2(2).1 The augmented transcript helped me deeply understand the content of the video	3.67	0.82
Q-C2(2).2 The augmented transcript supported me in understanding content in a language other than my native language	3.50	1.05
Q-C2(2).3 The augmented transcript helped me to keep my attention on the content while watching the video.	3.33	1.6
Q-C2(2).4 The augmented transcript runs too fast	2.17	1.6
Q-C2(2).5 The augmented transcript is in the right position within the interface	4.17	0.98
Q-C2(2).6 The augmented transcript added no value to experience of "rewatching" the video	3.00	1.1

Figure B.2: Third Evaluation with students: C2 condition

Condition C3: Map of concepts' flow-only toggled on

1. First watch scenario	AVG	Dev ST
Q-C3(1).1 The map of concepts' flow helped me understand the content of the video	4.13	0.83
Q-C3(1).2 The map of concepts' flow supported me in understanding content in a language other than my native language	4.00	1.07
Q-C3(1).3 The map of concepts' flow helped me keep my attention on the content while watching the video	3.00	1.41
Q-C3(1).4 The map of concepts' flow is in the right position within the interface	4.38	0.74
Q-C3(1).5 The map of concepts' flow flows too fast	2.88	1.46
Q-C3(1).6 The map of concepts' flow was distracting	3.38	1.41
Q-C3(1).7 The map of concepts' flow made it easier for me to find the 'definition' of a concept.	4.00	1.20
Q-C3(1).8 The map of concepts' flow made it easier for me to find the 'in-depth of a concept.	4.00	1.07
Q-C3(1).9 The nodes of the map of concepts' flow are easily navigable and clickable	3.75	1.39

2. Rewatch scenario	AVG	Dev ST
Q-C3(2).1 The map of concepts' flow helped me deeply understand the content of the video	4.14	1.46
Q-C3(2).2 The map of concepts' flow supported me in understanding content in a language other than my native language	3.43	1.'40
Q-C3(2).3 The map of concepts' flow helped me to keep my attention on the content while watching the video.	3.86	1.46
Q-C3(2).4 The map of concepts' flow runs too fast	2.00	0.82
Q-C3(2).5 The map of concepts' flow is in the right position within the interface	2.14	1.28
Q-C3(2).6 The map of concepts' flow added no value to experience of "rewatching" the video	2.43	1.27
Q-C3(2).7 The map of concepts' flow made it easier for me to find the 'definition' of a concept.	3.71	1.38
Q-C3(2).8 The map of concepts' flow made it easier for me to find the 'in-depth of a concept.	4.00	1.07
Q-C3(2).9 The nodes of the map of concepts' flow are easily navigable and clickable	4.43	0.79
Q-C3(2).10 The 'Search for Concepts' supported me in deepening a specific concept.	3.86	0.69
Q-C3(2).11 The 'Search for Concepts' bar is placed in the right position within the interface	4.43	0.79
Q-C3(2).12 The 'Search for Concepts' tab was useful for me to answer the test questions.	3.86	0.90

Figure B.3: Third Evaluation with students: C3 condition

Bibliography

[1] G. ADORNI, C. ALZETTA, F. KOCEVA, S. PASSALACQUA, AND
    I. TORRE, *Towards the identification of propaedeutic relations
    in textbooks*, in International Conference on Artificial Intelli-
    gence in Education, Springer, 2019, pp. 1–13.

[2] G. ADORNI AND F. KOCEVA, *Educational concept maps for per-
    sonalized learning path generation*, in Conference of the Italian
    Association for Artificial Intelligence, Springer, 2016, pp. 135–
    148.

[3] R. AGRAWAL, B. GOLSHAN, AND E. PAPALEXAKIS, *Toward
    data-driven design of educational courses: A feasibility study.*,
    Journal of Educational Data Mining, 8 (2016), pp. 1–21.

[4] L. ALBÓ, D. HERNÁNDEZ-LEO, AND V. MORENO OLIVER,
    *Smartphones or laptops in the collaborative classroom? a study
    of video-based learning in higher education*, Behaviour & Infor-
    mation Technology, 38 (2019), pp. 637–649.

[5] M. ALMOUSA, R. BENLAMRI, AND R. KHOURY, *Nlp-enriched
    automatic video segmentation*, in 2018 6th International Confer-
    ence on Multimedia Computing and Systems (ICMCS), IEEE,
    2018, pp. 1–6.

[6] C. ALZETTA, I. GALLUCCIO, F. KOCEVA, S. PASSALACQUA,
    AND I. TORRE, *Digging into prerequisite annotation.*, in iText-
    books@ AIED, 2020, pp. 29–34.

[7] C. ALZETTA, F. KOCEVA, S. PASSALACQUA, I. TORRE, AND G. ADORNI, *Pret: Prerequisite-enriched terminology. a case study on educational texts*, in Proceedings of the Fifth Italian Conference on Computational Linguistics CLiC-it, vol. 10, 2018, p. 12.

[8] C. ALZETTA, A. MIASCHI, G. ADORNI, F. DELL'ORLETTA, F. KOCEVA, S. PASSALACQUA, AND I. TORRE, *Prerequisite or not prerequisite? that's the problem! an nlp-based approach for concept prerequisite learning.*, in CLiC-it, 2019.

[9] T. ANDERSON, *Towards a theory of online learning*, Theory and practice of online learning, 2 (2004), pp. 109–119.

[10] D. P. AUSUBEL, *The psychology of meaningful verbal learning.*, (1963).

[11] ——, *In defense of advance organizers: A reply to the critics*, Review of Educational research, 48 (1978), pp. 251–257.

[12] I. E. AYDIN AND M. YAZICI, *Drop-out in moocs.*, Turkish Online Journal of Educational Technology-TOJET, 19 (2020), pp. 9–17.

[13] M. H. BATURAY, *An overview of the world of moocs*, Procedia-Social and Behavioral Sciences, 174 (2015), pp. 427–433.

[14] A. BERNS, J. M. MOTA, I. RUIZ-RUBE, AND J. M. DODERO, *Exploring the potential of a 360 video application for foreign language learning*, in Proceedings of the sixth international conference on technological ecosystems for enhancing multiculturality, 2018, pp. 776–780.

[15] A. BES-PIÁ, B. T. ENCARNA, AND M. J. MUÑOZ-PORTERO, *Different applications of concept maps in higher education*, Journal of Industrial Engineering and Management (JIEM), 4 (2011), pp. 81–102.

[16] N. BEVAN, J. CARTER, AND S. HARKER, *Iso 9241-11 revised: What have we learnt about usability since 1998?*, in International conference on human-computer interaction, Springer, 2015, pp. 143–151.

[17] M. BIRBILI, *Mapping knowledge: Concept maps in early childhood education.*, Early Childhood Research & Practice, 8 (2006), p. n2.

[18] H. BLISS, J. ABEL, AND B. GICK, *Computer-assisted visual articulation feedback in l2 pronunciation instruction: A review*, Journal of Second Language Pronunciation, 4 (2018), pp. 129–153.

[19] C. J. BRAME, *Effective educational videos: Principles and guidelines for maximizing student learning from video content*, CBE—Life Sciences Education, 15 (2016), p. es6.

[20] C. J. BRAME ET AL., *Effective educational videos*, 2015.

[21] J. BROOKE ET AL., *Sus-a quick and dirty usability scale*, Usability evaluation in industry, 189 (1996), pp. 4–7.

[22] P. BRUSILOVSKY, G. CHAVAN, AND R. FARZAN, *Social adaptive navigation support for open corpus electronic textbooks*, in Adaptive Hypermedia and Adaptive Web-Based Systems: Third International Conference, AH 2004, Eindhoven, The Netherlands, August 23-26, 2004. Proceedings 3, Springer, 2004, pp. 24–33.

[23] P. BRUSILOVSKY, S. SOSNOVSKY, AND K. THAKER, *The return of intelligent textbooks*, AI Magazine, 43 (2022), pp. 337–340.

[24] A. S. BURLEA AND D. D. BURDESCU, *An integrative approach of e-learning: from consumer to prosumer*, in Smart Education and e-Learning 2016, Springer, 2016, pp. 269–279.

[25] A. A. CATTANEO, A. T. NGUYEN, AND C. APREA, *Teaching and learning with hypervideo in vocational education and train-*

ing, Journal of Educational Multimedia and Hypermedia, 25 (2016), pp. 5–35.

[26] A. A. CATTANEO, H. VAN DER MEIJ, C. APREA, F. SAULI, AND C. ZAHN, *A model for designing hypervideo-based instructional scenarios*, Interactive learning environments, 27 (2019), pp. 508–529.

[27] A. A. CATTANEO, H. VAN DER MEIJ, AND F. SAULI, *An empirical test of three instructional scenarios for hypervideo use in a vocational education lesson*, Computers in the Schools, 35 (2018), pp. 249–267.

[28] T. CHAMBEL AND N. GUIMARÃES, *Communicating and learning mathematics with hypervideo*, in Multimedia tools for communicating mathematics, Springer, 2002, pp. 79–91.

[29] T. CHAMBEL, L. SANTOS, S. NÁPOLES, J. F. RODRIGUES, AND T. APOSTOL, *Visualization in learning mathematics with hypervideo*, Encontro Nacional de Visualização CientÍfica, Centro Multimeios, Espinho, Setembro, (2005).

[30] H. CHAMPEAUX, L. MANGIAVACCHI, F. MARCHETTA, AND L. PICCOLI, *Child development and distance learning in the age of covid-19*, Review of Economics of the Household, 20 (2022), pp. 659–685.

[31] A. H. CHAN, E. Y. KOK, M. A. M. RAZALI, G. A. LAWRIE, AND J. T. WANG, *Student perceptions and engagement in video-based learning for microbiology education*, International Journal of Innovation in Science and Mathematics Education, 30 (2022).

[32] D. CHAND AND H. OĞUL, *A framework for lecture video segmentation from extracted speech content*, in 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI), IEEE, 2021, pp. 000299–000304.

[33] M. A. CHATTI, M. MARINOV, O. SABOV, R. LAKSONO, Z. SO- FYAN, A. M. F. YOUSEF, AND U. SCHROEDER, *Video anno- tation and analytics in coursemapper*, Smart Learning Environ- ments, 3 (2016), pp. 1–21.

[34] X. CHE, H. YANG, AND C. MEINEL, *Lecture video segmentation by automatically analyzing the synchronized slides*, in Proceed- ings of the 21st ACM international conference on Multimedia, 2013, pp. 345–348.

[35] S. CIULLO, T. S. FALCOMATA, K. PFANNENSTIEL, AND G. BILLINGSLEY, *Improving learning with science and social studies text using computer-based concept maps for students with disabilities*, Behavior Modification, 39 (2015), pp. 117–135.

[36] M. COCCOLI, I. GALLUCCIO, I. TORRE, F. AMENDUNI, A. CATTANEO, AND C. CLARKE, *Advanced visual interfaces for augmented video*, in Proceedings of the 2022 International Conference on Advanced Visual Interfaces, 2022, pp. 1–3.

[37] A. CONDE, M. LARRAÑAGA, A. ARRUARTE, AND J. A. ELOR- RIAGA, *A combined approach for eliciting relationships for edu- cational ontologies using general-purpose knowledge bases*, IEEE Access, 7 (2019), pp. 48339–48355.

[38] K. CUKA AND E. BUFASI, *Hyper video for improving students' math performance*, European Journal of Education and Peda- gogy, 3 (2022), pp. 13–15.

[39] B. J. DALEY AND D. M. TORRE, *Concept maps in medical ed- ucation: an analytical literature review*, Medical education, 44 (2010), pp. 440–448.

[40] A. DAS AND P. P. DAS, *Automatic semantic segmentation and annotation of mooc lecture videos*, in International Conference on Asian Digital Libraries, Springer, 2019, pp. 181–188.

[41] ——, *Semantic segmentation of mooc lecture videos by analyzing concept change in domain knowledge graph*, in International Conference on Asian Digital Libraries, Springer, 2020, pp. 55–70.

[42] T. DESAI AND D. KULKARNI, *Assessment of interactive video to enhance learning experience: A case study*, Journal of Engineering Education Transformations, 35 (2022).

[43] N. DOGRUER, R. EYYAM, AND I. MENEVIS, *The use of the internet for educational purposes*, Procedia-Social and Behavioral Sciences, 28 (2011), pp. 606–611.

[44] M. DOUSE AND P. UYS, *Educational planning in the age of digitisation.*, Educational Planning, 25 (2018), pp. 7–23.

[45] J. DOWELL, *Interactive concept maps can augment learning with video media*, Seventh Int. Conference on Concept Mapping, Tallinn, Estonia, 2016.

[46] C. DZIUBAN, C. R. GRAHAM, P. D. MOSKAL, A. NORBERG, AND N. SICILIA, *Blended learning: the new normal and emerging technologies*, International journal of educational technology in Higher education, 15 (2018), pp. 1–16.

[47] G. W. ELLIS, A. RUDNITSKY, AND B. SILVERSTEIN, *Using concept maps to enhance understanding in engineering education*, International Journal of Engineering Education, 20 (2004), pp. 1012–1021.

[48] P. ERUVARAM, K. RAMANI, AND C. S. BINDU, *An experimental comparative study on slide change detection in lecture videos*, International Journal of Information Technology, 12 (2020), pp. 429–436.

[49] A. ESGUERRA, *Roles of Visual Clutter, Perceptual Load, and Multitasking on Monitoring Display Systems*, PhD thesis, California State University, Long Beach, 2021.

[50] A. R. Fabbri, I. Li, P. Trairatvorakul, Y. He, W. T. Ting, R. Tung, C. Westerfield, and D. R. Radev, *Tutorialbank: A manually-collected corpus for prerequisite chains, survey extraction and resource recommendation*, arXiv preprint arXiv:1805.04617, (2018).

[51] M. Fesmire, M. C. Lisner, P. R. Forrest, and W. H. Evans, *Concept maps: A practical solution for completing functional behavior assessments*, Education and Treatment of Children, (2003), pp. 89–103.

[52] A. L. Franzoni, C. P. Ceballos, and E. Rubio, *Interactive video enhanced learning-teaching process for digital native students*, in 2013 IEEE 13th International Conference on Advanced Learning Technologies, IEEE, 2013, pp. 270–271.

[53] R. Gagné, *Learning hierarchies and learning conditions*, Journal of Curriculum Studies, 7 (1975), pp. 133–134.

[54] R. M. Gagne, *The conditions of learning. holt, rinehart and winston*, Inc., New York, (1965).

[55] P. Gang, J. Hui, S. Stirenko, Y. Gordienko, T. Shemsedinov, O. Alienin, Y. Kochura, N. Gordienko, A. Rojbi, J. López Benito, et al., *User-driven intelligent interface on the basis of multimodal augmented reality and brain-computer interaction for people with functional disabilities*, in Future of Information and Communication Conference, Springer, 2018, pp. 612–631.

[56] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, *A survey on deep learning techniques for image and video semantic segmentation*, Applied Soft Computing, 70 (2018), pp. 41–65.

[57] M. J. GARCÍA-ORTA, V. GARCÍA-PRIETO, AND M. SUÁREZ-ROMERO, *New audiovisual consumption habits among minors: approximation through the analysis of survey data.*, Doxa Comunicación, (2019).

[58] F. GASPARETTI, C. DE MEDIO, C. LIMONGELLI, F. SCIARRONE, AND M. TEMPERINI, *Prerequisites between learning objects: Automatic extraction based on a machine learning approach*, Telematics and Informatics, 35 (2018), pp. 595–610.

[59] F. GASPARETTI, C. LIMONGELLI, AND F. SCIARRONE, *Exploiting wikipedia for discovering prerequisite relationships among learning objects*, in 2015 International Conference on Information Technology Based Higher Education and Training (ITHET), IEEE, 2015, pp. 1–6.

[60] M. N. GIANNAKOS, *Exploring the video-based learning research: A review of the literature*, British Journal of Educational Technology, 44 (2013), pp. E191–E195.

[61] N. GUIMARÃES, T. CHAMBEL, J. BIDARRA, ET AL., *From cognitive maps to hypervideo: Supporting flexible and rich learner-centred environments*, Interactive Multimedia Electronic Journal of Computer-Enhanced Learning, 2 (2000), pp. 1–7.

[62] R. HAMMOUD, *Interactive video*, Springer, 2006.

[63] L. F. HAMOND, G. WELCH, AND E. HIMONIDES, *The pedagogical use of visual feedback for enhancing dynamics in higher education piano learning and performance*, OPUS, 25 (2019), pp. 581–601.

[64] J. HATTIE AND H. TIMPERLEY, *The power of feedback*, Review of educational research, 77 (2007), pp. 81–112.

[65] T. HAYAMA AND S. SATO, *Supporting online video e-learning with semi-automatic concept-map generation*, in International

Conference on Human-Computer Interaction, Springer, 2020, pp. 64–76.

[66] L. HOLY, K. JEZEK, J. SNAJBERK, AND P. BRADA, *Lowering visual clutter in large component diagrams*, in 2012 16th International Conference on Information Visualisation, IEEE, 2012, pp. 36–41.

[67] S. HRASTINSKI, *What do we mean by blended learning?*, TechTrends, 63 (2019), pp. 564–569.

[68] Y. ICHIMURA, H. NAKANO, AND K. SUZUKI, *Ten dimensions model of moocs for quality design: Implications from the instructional design principles*, International Journal for Educational Media and Technology, 16 (2022), pp. 39–47.

[69] T. ILARIA AND S. FERRARI, *Il video come dispositivo di anticipazione per approfondire e generare domande. un caso di studio nella didattica universitaria*, (2021).

[70] A. ITANI, L. BRISSON, AND S. GARLATTI, *Understanding learner's drop-out in moocs*, in international conference on intelligent data engineering and automated learning, Springer, 2018, pp. 233–244.

[71] A. JACOBS, Y.-C. PAN, AND Y.-C. HO, *More than just engaging? tiktok as an effective learning tool.*, (2022).

[72] H. J. JEONG, T.-E. KIM, H. G. KIM, AND M. H. KIM, *Automatic detection of slide transitions in lecture videos*, Multimedia Tools and Applications, 74 (2015), pp. 7537–7554.

[73] S. JOKO PRAYUDHA, *Video based learning as a media for teaching english during pandemic covid-19*, Journal of Language Intelligence and Culture, 2 (2021), pp. 1–11.

[74] I. KAZANIDIS, G. PALAIGEORGIOU, A. PAPADOPOULOU, AND A. TSINAKOS, *Augmented interactive video: Enhancing video*

interactivity for the school classroom., Journal of Engineering Science & Technology Review, 11 (2018).

[75] J. KIM, P. J. GUO, C. J. CAI, S.-W. D. LI, K. Z. GAJOS, AND R. C. MILLER, *Data-driven interaction techniques for improving navigation of educational videos*, in Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14, New York, NY, USA, 2014, Association for Computing Machinery, p. 563–572.

[76] J. KIM, J. PARK, AND I. H. LU, *Hyperbutton: In-video question answering via interactive buttons and hyperlinks*, in Asian CHI Symposium 2021, 2021, pp. 48–52.

[77] A. KIRILLOVICH, M. FALILEEVA, O. NEVZOROVA, E. LIPACHEV, A. DYUPINA, AND L. SHAKIROVA, *Prerequisite relationships of the ontomath edu educational mathematical ontology*, in Workshop on Engineering Applications, Springer, 2021, pp. 517–524.

[78] J. KLEINBERG, *Bursty and hierarchical structure in streams*, in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 91–101.

[79] I. KOPRINSKA AND S. CARRATO, *Temporal video segmentation: A survey*, Signal processing: Image communication, 16 (2001), pp. 477–500.

[80] E. E. KOTOVA AND I. A. PISAREV, *Expansion of the students educational indicators activity interface in the moodle environment by means of intelligent agents*, in 2021 IV International Conference on Control in Technical Systems (CTS), IEEE, 2021, pp. 158–161.

[81] G. KOVACS, *Effects of in-video quizzes on mooc lecture viewing*, in Proceedings of the third (2016) ACM conference on Learning@ Scale, 2016, pp. 31–40.

[82] C. Kramer, J. König, S. Strauss, and K. Kaspar, *Classroom videos or transcripts? a quasi-experimental study to assess the effects of media-based learning on pre-service teachers' situation-specific skills of classroom management*, International Journal of Educational Research, 103 (2020), p. 101624.

[83] T. C. Leonard, *Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness*, 2008.

[84] J. R. Lewis, *Ibm computer usability satisfaction questionnaires: psychometric evaluation and instructions for use*, International Journal of Human-Computer Interaction, 7 (1995), pp. 57–78.

[85] ——, *Usability: lessons learned... and yet to be learned*, International Journal of Human-Computer Interaction, 30 (2014), pp. 663–684.

[86] I. Li, A. R. Fabbri, R. R. Tung, and D. R. Radev, *What should i learn first: Introducing lecturebank for nlp education and prerequisite chain learning*, in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 6674–6681.

[87] C. Liang, Z. Wu, W. Huang, and C. L. Giles, *Measuring prerequisite relations among concepts*, in Proceedings of the 2015 conference on empirical methods in natural language processing, 2015, pp. 1668–1674.

[88] J. Liao, J. Tang, and X. Zhao, *Course drop-out prediction on mooc platform via clustering and tensor completion*, Tsinghua Science and Technology, 24 (2019), pp. 412–422.

[89] C. Limongelli, F. Gasparetti, and F. Sciarrone, *Wiki course builder: a system for retrieving and sequencing didactic materials from wikipedia*, in 2015 International Conference on Information Technology Based Higher Education and Training (ITHET), IEEE, 2015, pp. 1–6.

[90] C. LIU, H.-C. WANG, AND J. KIM, *Videoscape: Augmenting video learning experience with concept map.*

[91] J. LIU, Y. ZHENG, AND X. JIANG, *Design of augmented interactive video based on script engine*, in Proceedings of the International Workshop on Artificial Intelligence and Education, 2019, pp. 1–5.

[92] L. LIU, L.-T. CHEN, AND K. PUGH, *Online teaching and learning under covid-19: Challenges and opportunities*, Computers in the Schools, 38 (2022), pp. 249–255.

[93] Z. LIU, K. LI, L. SHEN, R. MA, AND P. AN, *Spatio-temporal residual networks for slide transition detection in lecture videos*, KSII Transactions on Internet and Information Systems (TIIS), 13 (2019), pp. 4026–4040.

[94] C. LOCATIS, J. CHARUHAS, AND R. BANVARD, *Hypervideo*, Educational Technology Research and Development, 38 (1990), pp. 41–49.

[95] W. LU, P. MA, J. YU, Y. ZHOU, AND B. WEI, *Metro maps for efficient knowledge learning by summarizing massive electronic textbooks*, International Journal on Document Analysis and Recognition (IJDAR), 22 (2019), pp. 99–111.

[96] R. MACLEAN, *Television in education.*, (1968).

[97] N. MARASCHIO, *Le nuove fonti della lingua: radio e televisione*, V. Coletti (a cura di), L'italiano dalla nazione allo Stato, Le Lettere, Firenze, (2011), pp. 161–171.

[98] F. MARTINS BATISTA LÃ, J. SUNDBERG, AND S. GRANQVIST, *Augmented visual-feedback of airflow: Immediate effects on voice-source characteristics of students of singing*, Psychology of Music, (2021).

[99] I. A. MASTAN, D. I. SENSUSE, R. R. SURYONO, AND K. KAUT-SARINA, *Evaluation of distance learning system (e-learning): a systematic literature review*, Jurnal Teknoinfo, 16 (2022), pp. 132–137.

[100] M. T. MAYBURY, *Intelligent user interfaces for all*, Lawrence Erlbaum Associates, Publishers, Mahwah, NJ, 2001.

[101] R. E. MAYER, *Introduction to multimedia learning*, The Cambridge handbook of multimedia learning, 2 (2005), p. 24.

[102] R. E. MAYER AND R. MORENO, *Nine ways to reduce cognitive load in multimedia learning*, Educational psychologist, 38 (2003), pp. 43–52.

[103] D. F. MCLEAN, *Restoring baird's image*, IEE Review, 46 (2000), pp. 9–14.

[104] C. MEIGH-ANDREWS, *A history of video art*, A&C Black, 2013.

[105] B. MEIXNER AND M. GOLD, *Second-layer navigation in mobile hypervideo for medical training*, in International Conference on Multimedia Modeling, Springer, 2016, pp. 382–394.

[106] B. MEIXNER, C. HANDSCHIGL, S. JOHN, AND M. GRANITZER, *A dual screen concept for user-controlled hypervideo-based physiotherapy training*, in Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care, 2016, pp. 9–13.

[107] A. MITROVIC, V. DIMITROVA, L. LAU, A. WEERASINGHE, AND M. MATHEWS, *Supporting constructive video-based learning: requirements elicitation from exploratory studies*, in International Conference on Artificial Intelligence in Education, Springer, 2017, pp. 224–237.

[108] M. G. MOORE, *Three types of interaction*, (1989).

[109] B. NADEAK AND L. NAIBAHO, *Video-based learning on improving students'learning output*, PalArch's Journal of Archaeology of Egypt/Egyptology, 17 (2020), pp. 44–54.

[110] G. NAGY, *At the frontiers of ocr*, Proceedings of the IEEE, 80 (1992), pp. 1093–1100.

[111] J. NAVIO-MARCO, L. M. RUIZ-GÓMEZ, R. ARGUEDAS-SANZ, AND C. LÓPEZ-MARTÍN, *The student as a prosumer of educational audio–visual resources: a higher education hybrid learning experience*, Interactive Learning Environments, (2022), pp. 1–18.

[112] B. NEBEL AND H.-J. BÜRCKERT, *Reasoning about temporal relations: A maximal tractable subclass of allen's interval algebra*, Journal of the ACM (JACM), 42 (1995), pp. 43–66.

[113] C.-W. NGO, T.-C. PONG, AND T. S. HUANG, *Detection of slide transition for topic indexing*, in Proceedings. IEEE International Conference on Multimedia and Expo, vol. 2, IEEE, 2002, pp. 533–536.

[114] J. NIELSEN AND R. MOLICH, *Heuristic evaluation of user interfaces*, in Proceedings of the SIGCHI conference on Human factors in computing systems, 1990, pp. 249–256.

[115] J. D. NOVAK AND A. J. CAÑAS, *The theory underlying concept maps and how to construct and use them*, (2008).

[116] J. D. NOVAK AND D. B. GOWIN, *Learning how to learn*, cambridge University press, 1984.

[117] E. G. OH, Y. CHANG, AND S. W. PARK, *Design review of moocs: Application of e-learning design principles*, Journal of Computing in Higher Education, 32 (2020), pp. 455–475.

[118] D. PAL AND S. PATRA, *University students' perception of video-based learning in times of covid-19: A tam/ttf perspective*, In-

ternational Journal of Human–Computer Interaction, 37 (2021), pp. 903–921.

[119] L. PAN, C. LI, J. LI, AND J. TANG, *Prerequisite relation learning for concepts in MOOCs*, in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, July 2017, Association for Computational Linguistics, pp. 1447–1456.

[120] L. PAN, X. WANG, C. LI, J. LI, AND J. TANG, *Course concept extraction in moocs via embedding-based graph propagation*, in Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2017, pp. 875–884.

[121] J. PARSOLA, D. GANGODKAR, AND A. MITTAL, *Video segmentation techniques for instructional videos–survey*, Journal of Graphic Era University, (2019), pp. 90–107.

[122] M. PERINI, A. A. CATTANEO, AND G. TACCONI, *Using hypervideo to support undergraduate students' reflection on work practices: a qualitative study*, International Journal of Educational Technology in Higher Education, 16 (2019), pp. 1–16.

[123] D. M. POWERS, *Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation*, arXiv preprint arXiv:2010.16061, (2020).

[124] D. QUIÑONES AND C. RUSU, *How to develop usability heuristics: A systematic literature review*, Computer standards & interfaces, 53 (2017), pp. 89–122.

[125] S. RABIGER, T. DALKILIÇ, A. DOĞAN, B. KARAKAŞ, B. TÜRETKEN, AND Y. SAYGIN, *Exploration of video e-learning content with smartphones*, (2020).

[126] L. RABINER AND B. JUANG, *An introduction to hidden markov models*, ieee assp magazine, 3 (1986), pp. 4–16.

[127] A. G. RADIN AND C. J. LIGHT, *Tiktok: An emergent opportunity for teaching and learning science communication online*, Journal of microbiology & biology education, 23 (2022), pp. e00236–21.

[128] S. RAWAT, D. KUMAR, P. KUMAR, AND C. KHATTRI, *A systematic analysis using classification machine learning algorithms to understand why learners drop out of moocs*, Neural Computing and Applications, 33 (2021), pp. 14823–14835.

[129] A. REGIS, P. G. ALBERTAZZI, AND E. ROLETTO, *Concept maps in chemistry education*, Journal of Chemical Education, 73 (1996), p. 1084.

[130] J. REIS, *What makes tiktok so addicting?*, (2022).

[131] M. C. REYES AND G. TRENTIN, *Un modello di interactive mooc per potenziare l'interattività e infrangere la sequenzialità*, Adorni G. e Koceva F.(2019)(a cura di), Atti di Didamatica, (2019), pp. 125–134.

[132] L. P. RIEBER, *Multimedia learning in games, simulations, and microworlds*, The Cambridge handbook of multimedia learning, (2005), pp. 549–567.

[133] E. F. RISKO, N. ANDERSON, A. SARWAL, M. ENGELHARDT, AND A. KINGSTONE, *Everyday attention: Variation in mind wandering and memory in a lecture*, Applied Cognitive Psychology, 26 (2012), pp. 234–242.

[134] R. ROSENHOLTZ, Y. LI, AND L. NAKANO, *Measuring visual clutter*, Journal of vision, 7 (2007), pp. 17–17.

[135] M. SABLIĆ, A. MIROSAVLJEVIĆ, AND A. ŠKUGOR, *Video-based learning (vbl)—past, present and future: An overview of the research published from 2008 to 2019*, Technology, Knowledge and Learning, 26 (2021), pp. 1061–1077.

[136] F. SAULI, A. CATTANEO, AND H. VAN DER MEIJ, *Hypervideo for educational purposes: a literature review on a multifaceted technological tool*, Technology, pedagogy and education, 27 (2018), pp. 115–134.

[137] N. SAWHNEY, D. BALCOM, AND I. SMITH, *Hypercafe: narrative and aesthetic properties of hypervideo*, in Proceedings of the the seventh ACM conference on Hypertext, 1996, pp. 1–10.

[138] S. L. SCHNEIDER AND M. L. COUNCIL, *Distance learning in the era of covid-19*, Archives of dermatological research, 313 (2021), pp. 389–390.

[139] K. SCHOEFFMANN, M. A. HUDELIST, AND J. HUBER, *Video interaction tools: a survey of recent work*, ACM Computing Surveys (CSUR), 48 (2015), pp. 1–34.

[140] N. L. SCHROEDER, J. C. NESBIT, C. J. ANGUIANO, AND O. O. ADESOPE, *Studying and constructing concept maps: A meta-analysis*, Educational Psychology Review, 30 (2018), pp. 431–455.

[141] B. SETTLES, *Active learning literature survey*, (2009).

[142] R. R. SHAH, Y. YU, A. D. SHAIKH, S. TANG, AND R. ZIMMER-MANN, *Atlas: automatic temporal segmentation and annotation of lecture videos based on modelling transition time*, in Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 209–212.

[143] B. SHAN, M. R. ABID, AND E. AMJADIAN, *Hybrid unsupervised scale-invariant slide detection (hussd) for video presentation*, in Proceedings of the 2020 the 4th International Conference on Information System and Data Mining, 2020, pp. 100–106.

[144] K. SHARMA, H. S. ALAVI, P. JERMANN, AND P. DILLEN-BOURG, *A gaze-based learning analytics model: in-video visual*

feedback to improve learner's attention in moocs, in Proceedings of the sixth international conference on learning analytics & knowledge, 2016, pp. 417–421.

[145] S. SINGGIH ET AL., *Video-based learning for "learning from home" solution in pandemic*, in Journal of Physics: Conference Series, vol. 1760, IOP Publishing, 2021, p. 012011.

[146] E. R. SOARES AND E. BARRÉRE, *An approach for automatic segmentation of scenes in educational videos through the use of audio transcription and semantic annotation*, in Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, 2017, pp. 229–235.

[147] J. STIGLER, E. GELLER, AND K. GIVVIN, *Zaption: A platform to support teaching, and learning about teaching, with video*, Journal of e-Learning and Knowledge Society, 11 (2015).

[148] S. SUFIRMANSYAH, L. N. PRAMESWATI, D. T. WATI, AND E. SULISTYOWATI, *Student's preferences in using video-based learning applications and its efficiency in higher education*, Nazhruna: Jurnal Pendidikan Islam, 4 (2021), pp. 272–283.

[149] J. SWELLER, *Cognitive load during problem solving: Effects on learning*, Cognitive science, 12 (1988), pp. 257–285.

[150] J. SWELLER AND P. CHANDLER, *Evidence for cognitive load theory*, Cognition and instruction, 8 (1991), pp. 351–362.

[151] P. TALUKDAR AND W. COHEN, *Crowdsourced comprehension: Predicting prerequisite structure in wikipedia*, in Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, Montréal, Canada, June 2012, Association for Computational Linguistics, pp. 307–315.

[152] ——, *Crowdsourced comprehension: predicting prerequisite structure in wikipedia*, in Proceedings of the Seventh Workshop on

Building Educational Applications Using NLP, 2012, pp. 307–315.

[153] A. J. Tan, J. Davies, R. I. Nicolson, and T. Karaminis, *A technology-enhanced learning intervention for statistics in higher education using bite-sized video-based learning and precision teaching*, Research and Practice in Technology Enhanced Learning, 18 (2023).

[154] C.-L. Tang, J. Liao, H.-C. Wang, C.-Y. Sung, and W.-C. Lin, *Conceptguide: Supporting online video learning with concept map-based recommendation of learning path*, in Proceedings of the Web Conference 2021, 2021, pp. 2757–2768.

[155] Y. Taskin, T. Hecking, H. U. Hoppe, V. Dimitrova, and A. Mitrovic, *Characterizing comment types and levels of engagement in video-based learning as a basis for adaptive nudging*, in European Conference on Technology Enhanced Learning, Springer, 2019, pp. 362–376.

[156] I. Torre, I. Galuccio, and M. Coccoli, *Video augmentation to support video-based learning*, in Proceedings of the 2022 International Conference on Advanced Visual Interfaces, 2022, pp. 1–5.

[157] I. Torre, L. Mirenda, G. Vercelli, and F. Mastrogiovanni, *Prerequisite graph extraction from lectures*, in International Conference on Artificial Intelligence in Education, Springer, 2022, pp. 616–619.

[158] F. Ulrich, N. H. Helms, U. P. Frandsen, and A. V. Rafn, *Learning effectiveness of 360 video: experiences from a controlled experiment in healthcare education*, Interactive Learning Environments, 29 (2021), pp. 98–111.

[159] E. Van Zele, J. Lenaerts, and W. Wieme, *Improving the usefulness of concept maps as a research tool for science edu-*

cation, International Journal of Science Education, 26 (2004), pp. 1043–1064.

[160] J. VANIDES, Y. YIN, M. TOMITA, AND M. A. RUIZ-PRIMO, *Concept maps*, Science Scope, 28 (2005), pp. 27–31.

[161] G. VERMA, T. NALAMADA, K. HARPAVAT, P. GOEL, A. MISHRA, AND B. V. SRINIVASAN, *Non-linear consumption of videos using a sequence of personalized multimodal fragments*, in 26th International Conference on Intelligent User Interfaces, 2021, pp. 249–259.

[162] ——, *Non-linear consumption of videos using a sequence of personalized multimodal fragments*, in 26th International Conference on Intelligent User Interfaces, IUI '21, New York, NY, USA, 2021, Association for Computing Machinery, p. 249–259.

[163] M. WANG, H. CHAU, K. THAKER, P. BRUSILOVSKY, AND D. HE, *Knowledge annotation for intelligent textbooks*, Technology, Knowledge and Learning, (2021), pp. 1–22.

[164] S. WANG AND L. LIU, *Prerequisite concept maps extraction for automaticassessment*, in Proceedings of the 25th International Conference Companion on World Wide Web, 2016, pp. 519–521.

[165] S. WANG, A. OROBIA, Z. WU, K. WILLIAMS, C. LIANG, B. PURSEL, AND C. L. GILES, *Using prerequisites to extract concept maps fromtextbooks*, in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM '16, New York, NY, USA, 2016, Association for Computing Machinery, p. 317–326.

[166] Z. WANG, K. XIAO, Z. QIN, AND S. WANG, *Extracting prerequisite relations among concepts from the introduction of online courses*, in Proc. 33rd Int. Conf. Software Engineering and Knowledge Engineering, 2021, pp. 375–380.

[167] M. M. WAQAR, M. ASLAM, AND M. FARHAN, *An intelligent and interactive interface to support symmetrical collaborative educational writing among visually impaired and sighted users*, Symmetry, 11 (2019), p. 238.

[168] T. WINKLER, M. IDE, AND M. HERCZEG, *The use of hypervideo in teacher education*, in Conference Paper, Conference: Association of Ubiquitous and Collaborative Educators international, At Cocoa Beach, Floyda, USA, OI, vol. 10, 2013, pp. 3728–9765.

[169] B. WISNIEWSKI, K. ZIERER, AND J. HATTIE, *The power of feedback revisited: A meta-analysis of educational feedback research*, Frontiers in Psychology, 10 (2020), p. 3087.

[170] L. WISSLER, M. ALMASHRAEE, D. M. DÍAZ, AND A. PASCHKE, *The gold standard in corpus annotation.*, IEEE GSC, 21 (2014).

[171] W. XING, *Exploring the influences of mooc design features on student performance and persistence*, Distance Education, 40 (2019), pp. 98–113.

[172] A. YEE, W. M. PADOVANO, I. K. FOX, E. J. HILL, A. G. ROWE, L. M. BRUNT, A. M. MOORE, A. K. SNYDER-WARWICK, L. C. KAHN, M. D. WOOD, ET AL., *Video-based learning in surgery: establishing surgeon engagement and utilization of variable-duration videos*, Annals of surgery, 272 (2020), pp. 1012–1019.

[173] A. M. F. YOUSEF, M. A. CHATTI, AND U. SCHROEDER, *The state of video-based learning: A review and future perspectives*, International Journal on Advances in Life Sciences, 6 (2014), pp. 122–135.

[174] ——, *Video-based learning: A critical analysis of the research published in 2003-2013 and future visions*, in eLmL 2014, The Sixth International Conference on Mobile, Hybrid, and On-line Learning, 2014, pp. 112–119.

[175] C. ZAHN, B. BARQUERO, AND S. SCHWAN, *Learning with hyperlinked videos—design criteria and efficient strategies for using audiovisual hypermedia*, Learning and Instruction, 14 (2004), pp. 275–291.

[176] H. ZENG, X. SHU, Y. WANG, Y. WANG, L. ZHANG, T.-C. PONG, AND H. QU, *Emotioncues: Emotion-oriented visual summarization of classroom videos*, IEEE transactions on visualization and computer graphics, 27 (2020), pp. 3168–3181.

[177] X. ZHANG, C. LI, S.-W. LI, AND V. ZUE, *Automated segmentation of mooc lectures towards customized learning*, in 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), 2016, pp. 20–22.

[178] B. ZHAO, S. LIN, X. LUO, S. XU, AND R. WANG, *A novel system for visual navigation of educational videos using multimodal cues*, in Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 1680–1688.

[179] B. ZHAO, S. LIN, X. QI, R. WANG, AND X. LUO, *A novel approach to automatic detection of presentation slides in educational videos*, Neural Comput. Appl., 29 (2018), p. 1369–1382.

[180] D. W. ZIMMERMAN, *Comparative power of student t test and mann-whitney u test for unequal sample sizes and variances*, The Journal of Experimental Education, 55 (1987), pp. 171–174.

Acknowledgements

at the beginning of my PhD programme, were crucial in setting me on my research path: Prof. Giovanni Adorni, Frosina Koceva, Samuele Passalacqua, Morteza Alinam, Selenia Anastasi and especially Chiara Alzetta, who was a formidable, brilliant and enthusiastic PhD colleague. I would like to thank my colleagues from cycle XXXV who shared this somewhat complicated but definitely challenging PhD experience with me, especially the fantastic and very nice Giulia Chiantia and Angela Zinno for her wisdom and support.

I would like to thank Prof. Gabriele Gattiglia of the University of Pisa and his Archaeology students for their enthusiastic participation in the experimentation designed specifically for the archaeological domain.

Finally, I would like to thank my parents Patrizia and Salvatore, my two sisters Gabriella and Martina, my cousin Alessia and the entire family for the support and encouragement they have given me in the most difficult moments. A big thank you to the whole host of my friends scattered around the world, in particular my best friends Anna, Nina and Rachele, who have supported me in times of discouragement with their good humour.

To all of you goes my gratitude and profound joy for this important milestone that is about to be reached.

Ilenia