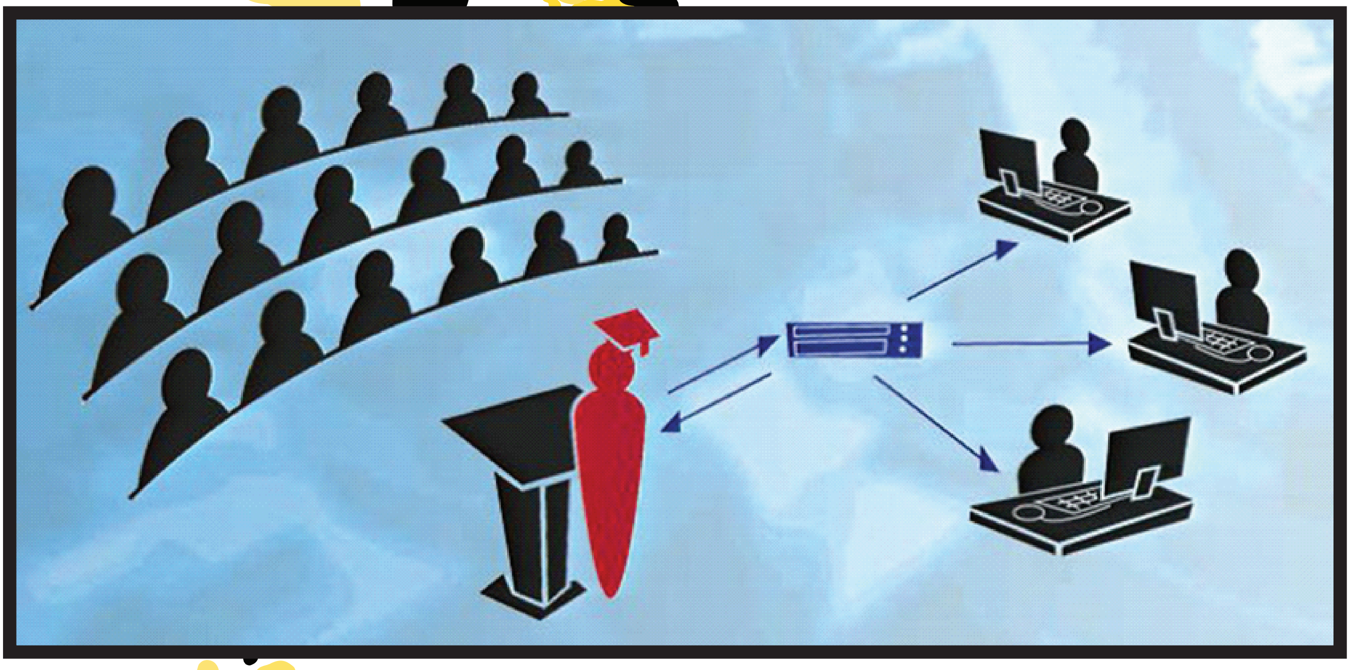


Expanding the usability of recorded lectures

A new age in teaching
and classroom instruction

E.L. de Moel

Collegerama UNIVERSITY OF TWENTE. TU Delft

Expanding the usability of recorded lectures

A new age in teaching and classroom instruction

E.L. de Moel

for the degree of:
Master of Science in Computer Science

Date of submission: 26 February 2010
Date of defense: 3 March 2010

Committee:

dr.ir. D. Hiemstra	University of Twente Faculty of Electrical Engineering, Mathematics and Computer Science
Dipl. Wirtsch.-Info R. Aly	University of Twente Faculty of Electrical Engineering, Mathematics and Computer Science
dr. T.H.S. Eysink	University of Twente Faculty of Behavioural Sciences

Chair Databases
Department Computer Science
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente

Summary

Background
At present, Delft University of Technology records around 10% of their lectures. This number is expected to increase in the following years. Having these recorded lectures opens the door to all kinds of new ideas and improvements for their educational program. At this moment they employ a video streaming system called Collegerama, which allows viewers with an active Internet connection to watch their lectures online. It combines a video stream of the lecturer with a series of screenshots of the accompanying PowerPoint slides.

Research
The main research question for this project is: "How can we efficiently and effectively present recorded lectures and course material to students at universities?" This can be divided into three sub-questions:

- How can we increase the accessibility and availability of the recorded lectures in Collegerama?
- How can we make recorded lectures easier to follow, especially for foreign speaking students?
- How can we effectively and efficiently navigate and search within recorded lectures?

The research approach for this project was to study the individual questions separately. In the second phase of the project, the individual results were combined into a set of integrated recommendations for further development in a short term implementation project.

Accessibility and availability
To increase the availability of the lectures, it is recommended to create a single video file from the Collegerama recordings. This will allow for the distribution over many other popular online multimedia platforms, such as YouTube-Edu and iTunes-U. A single video file distribution allows for offline viewing without an active broadband Internet connection (for example, while sitting in the train). This is not possible within the current Collegerama system.

In this research project, a Collegerama lecture has been converted into a single video stream, after careful review of several layout designs and technical specifications. This lecture has been published on YouTube. Several other formats have been created, so that the lecture can also be distributed on all kinds of distribution platforms. This includes a smaller sized version, created specifically for mobile devices and has been tested on Apple's latest iPhone.

Easier to follow
To make lectures easier to follow, we show that the creation and displaying of subtitles is useful. These subtitles can automatically be translated using machine translation. For this research project, Google Translate has been used which currently supports translation to 52 different languages. The quality of these is decent, depending on the target language that has been chosen. If necessary, this generated text can be enhanced by manual post-processing. The current speech recognition technology has also been evaluated for the generation of proper subtitles, using the speech recognition engine created by University of Twente called SHoUT. It is concluded that this system is not yet sufficient to generate proper subtitles and manual post-processing to improve the output is always required.

Navigation and search
This research project has shown that to properly navigate through the available recorded lectures, the input from teachers is important. They need to provide the lecture title and divide their lectures into several chapters with a proper chapter title, based on separate timeframes (start time and end time). These chapters together with the slide titles and slide content form the foundation for navigation and searching. The search element can be further expanded by the available subtitles. For the purpose of this research project, all lecture titles

and chapters provided by the lecturer, slide titles and content and the generated SHoUT transcripts for all 14 lectures (28 lecture videos) have been collected. The slide metadata has been digitally and automatically extracted from the original PowerPoint files.

All this new information and metadata has been stored in a multimedia database, so that the retrieval options for the lecture content could be researched. This database serves as the source for all the additional options for navigation and searching:

- generating a static and/or interactive table of contents for each lecture (based on lecture chapters)
- generating tag clouds
- displaying subtitles in several different languages
- searching within lecture material

To demonstrate its functionality, a prototype for a Collegerama lecture search engine has been developed. This is an online web application that can be accessed from any location with an active Internet connection and searches within all the above mentioned data linked to a lecture. Every search result provides a link to Collegerama, so users can immediately see the related part of the lecture.

Future developments

It is concluded that a better system for recording slides needs to be developed. Looking at the future of education and the increasing developments in technology, it's clear that presentations are going to be supported by more animation and video. This means that an old screenshot recording system will no longer be sufficient to properly record PowerPoint slides.

To further increase the usability of the recorded lectures, a new interactive way to discuss lectures with the teacher and other students can be introduced. It promotes the asking and answering of questions, not just by the teacher but also by fellow classmates. This can be done through the use of a dynamic message board that is linked to the timeline of each lecture. Students can comment and discuss on the different topics in the lecture. To support such a system, an extension of the current multimedia database is required, so that the messages along with their optional timeframes can be stored.

With these recommendations, it is possible to use recorded lectures as a foundation for future online-given courses without the need for live lectures.

Abstract

The status of recorded lectures at Delft University of Technology has been studied in order to expand its usability in their present and future educational environment. Possibilities for the production of single file vodcasts have been tested. These videos allow for an increased accessibility of their recorded lectures through the form of other distribution platforms.
Furthermore the production of subtitles has been studied. This was done with an ASR system called SHoUT, developed at University of Twente, and machine translation of subtitles into other languages. SHoUT generated transcripts always require post-processing for subtitling. Machine translation could produce translated subtitles of sufficient quality.
Navigation of recorded lectures needs to be improved, requiring input of the lecturer. Collected metadata from lecture chapter titles, slide data (titles, content and notes) as well as ASR results have been used for the creation of a lecture search engine, which also produces interactive tables of content and tag clouds for each lecture.
Recorded lectures could further be enhanced with time-based discussion boards, for the asking and answering of questions. Further improvements have been proposed for allowing recorded lectures to be re-used in recurring online-based courses.

Preface

This report has been written as a result of my research project at University of Twente, in cooperation with Delft University of Technology. It was originally a project that started at TU Delft, in which my father was involved. He has been active at the university for the past 8 years at the chair of drinking water engineering, to improve their educational programs. When the development of a system for recording and sharing recorded lectures started (Collegerama), they were one of the first chairs at the university that started recording all their lectures.

At that time, I was active as an online poker instructor, teaching enthusiastic players ways to improve their game. I told my father the techniques we used to teach and instruct students all over the world through the use of the Internet, either one on one or via online streaming recorded lectures. We began exchanging ideas about this subject and started to see the remarkable potential that lies ahead with this new form of online multimedia education. That is how I got involved with this research project.

The goal of this project is to research possibilities for expanding the usability of recorded lectures at TU Delft and University of Twente and improve the means for distributing and sharing lecture and course material to students at universities.

I would like to thank the following people:

- Djoerd Hiemstra, my first supervisor, for assisting and guiding me during my research project
- Robin Aly, my second supervisor, for providing ideas about evaluating certain topics in my thesis
- Peter de Moel, for providing feedback and to bounce ideas back and forth about the future of online recorded lectures
- Thijs Verschoor, for generating the subtitles for the 28 lecture videos of course CT3011, using the SHoUT engine developed by University of Twente
- Willem Jansen, for helping me with the automatic conversion from pdf PowerPoint sheets to an Excel data sheet and a SQL 2005 database using a C++ script
- Koen Zomers, for his useful Visual Studio 2008 tips while programming the Collegerama lecture search engine
- my parents, sister, family and friends for their support during the past 9 months and during the course of my master

February 2010
Erwin de Moel

Table of contents

List of figures

List of tables

1. Introduction

Background

For the past 10 years, there has been little or no change in the way that lectures are given at the various Universities throughout the world. With the emerging of new technologies, there are numerous new possibilities for improving the method in which information is shared between student and teacher. Through the use of the Internet, there is an incredible amount of additional material that can be found in order to delve even deeper into the subject matter. Most universities already employ an online community and messaging system where lecture sheets, additional subject material and practice exams are shared.

Every year, a teacher of a course gives a similar lecture compared to the previous year, while a new group of students follows the course. As long as both the course and lecture material don't go through a significant change, this seems somewhat redundant. In the past year, TU Delft has also been faced with a problem. The amount of registrants for certain courses exceeds the maximum capacity of the largest available classroom.

TU Delft has been developing its own system for the production and streaming of digitally recorded lectures, called Collegerama. They stream lectures on a web server to further support their learning programs. Figure 1.1 shows an example of a lecture given at the University of Delft that has been recorded and can be viewed online.

Figure 1.1: Screenshot of a recorded lecture at TU Delft in Collegerama
(Source: http://collegerama.tudelft.nl/mediasite/Viewer/?peid=7548f752-101b-417e-a4e7-58aebc595376)

The recorded lectures on Collegerama contain the following elements:

- a video stream of the lecturer
- screenshots of the presentation sheets or an interactive screen (tablet PC) on which the presenter writes notes
- navigation tools for scrolling through the video and/or slides
- controls for play/pause, full screen mode and to modify the playback speed

Goals

TU Delft would like to investigate possibilities of expanding the usability of their recorded lectures. They have several ideas for achieving this:

- subtitling the lectures for students with hearing problems
- subtitling in other languages (English subtitles for Dutch lectures and vice versa)
- translated subtitles spoken over the original video stream (with or without subtitles)
- searching in lecture content (whether in transcripts/subtitles)
- searching in lecture content (by handmade content overview and/or computer generated keywords)
- distribution as a vodcast (for PDA, iPod, iPhone or another type of mobile phone)

Research questions

The main research question for this project is:

> *How can we efficiently and effectively present recorded lectures and course material to students at universities?*

This question can be divided into three sub questions:

- *How can we increase the accessibility and availability of the recorded lectures in Collegerama?*
- *How can we make recorded lectures easier to follow, especially for foreign speaking students?*
- *How can we effectively and efficiently navigate and search within recorded lectures?*

Project boundaries

This research project will not include investigations on user preferences (teachers and students), best educational practices, optimal teaching methods for recorded lectures etc. It will be restricted to alternatives technically feasible within the E-learning and ICT environment of TU Delft. This is not restricted to presently used ICT tools, but might include commercial available products implementable within the TU Delft environment.

Starting point for this research project are the presently produced Collegerama recordings. No other methods of recording lectures will be evaluated. This research project includes a technology-centered approach to the subject. A user-centered approach might be taken in a succeeding research project, evaluating the different proposed extending products and applications for using recorded lectures and the benefits and problems it might bring to the different user groups (teachers, local students, foreign exchange students, the University etc).

Report outline

The entire report can be separated into three parts; chapters 2 and 3 give an introduction into systems currently used for the distribution of recorded lectures, followed by chapters 4, 5 and 6 in which possibilities for expanding the usability of these recorded lectures are discussed. Chapter 7 contains the conclusions of the entire research project.

In chapter 2, a detailed history of recorded lectures is described, starting with the way that the Massachusetts Institute of Technology produces and distributes its lectures. This is followed by the current Collegerama system that is used by both University of Twente and Delft University of Technology. Chapter 3 describes several formats for producing and storing recorded lectures. It mentions several audio and video formats that can be used, in what way a timeline can be determined for a lecture and how to handle the link between the video stream of the lecturer and the slides that accompany the presentation. After this introduction into the world of online recorded lectures, chapter 4 discusses new possibilities for expanding its usability through the means of subtitling and translation. In chapter 5, a description of several different methods for navigating and searching through the various recorded lectures is given. Chapters 4 and 5 are concluded in chapter 6, by describing a list of proposed improvements. An actual prototype for a lecture search engine has been designed and a new way of browsing through a group of lectures for a single course is demonstrated. Finally, the conclusions of the entire report are presented in chapter 7.

The printed version of the report does not include the annexes. The annexes are only included in the electronic version of the report. An accompanying DVD includes the intermediate and final products of this research project.

2. Existing systems for digitally recorded lectures

Ten years ago (in 1999), the Massachusetts Institute of Technology started broadcasting several unique physics lectures over a local TV channel. This was primarily done to gain more exposure for their educational programs. It received a lot of positive results, which caused them to start recording and distributing more lectures from other sciences by use of the Internet. As time went on, the recorded lectures were also being used to improve and expand their learning programs for their own students by publishing them on the Internet.

Several years later, a trend started to emerge and several other large universities in the United States, such as Berkeley and Stanford started to do the same. In 2000, Delft University of Technology, followed by University of Twente, started its own lecture recording programs. After running a few successful pilots, they are now recording more and more lectures each year. It's not going to be surprising to see that within the next couple of years, all the Dutch universities are doing the same with their courses and lectures.

In this chapter, the history, recording process and developments with regards to recorded lectures are discussed. The differences between the techniques used at MIT, TU Delft and University of Twente will be shown and several drawbacks in the current system that both Dutch universities are using will be described.

For further background information about the research on this topic, see Annex A and B.

2.1 Massachusetts Institute of Technology

Massachusetts Institute of Technology (MIT) is a private research university located in Cambridge, Massachusetts in the United States. It has five schools and one college, containing a total of 32 academic departments, with a strong emphasis on scientific and technological research. It is one of the most prestigious technical universities in the world. Their reputation is based on their scientific output through the publishing of scientific articles and reports and the awards received by their staff. Seventy-three members of the MIT community have won the Nobel Prize, including seven current faculty members.[24]

MIT enrolled 4,232 undergraduates and 6,152 graduate students during the fall of 2009–2010.[25] It employs about 1,000 faculty members. Its endowment and annual research expenditures are among the largest of any American university. 75 Nobel Laureates, 47 National Medal of Science recipients and 31 MacArthur Fellows are currently or have previously been affiliated with the university.[24] The aggregated revenues of companies founded by MIT alumni would be the seventeenth largest economy in the world.[26][27]

OpenCourseWare

In 2000, MIT started the concept of publishing their course material on the Internet, which would be publically available for everyone. They called this project OpenCourseWare (OCW). The first proof-of-concept site was published in 2002, containing 50 courses. By November of 2007, MIT completed the initial publication of almost their entire curriculum which contained over 1,800 courses in 33 academic disciplines.[29]

MIT also publishes some of their courses in one or more translated versions and have formally partnered with four organizations that are translating OCW course material into Spanish, Portuguese, Simplified Chinese, Traditional Chinese and Thai. Their material has already been translated into at least 10 different languages, including French, German, Vietnamese, and Ukrainian.

Since 2008, MIT has added audio and video-taped lectures to their OCW website. These lectures were recorded between 1999 and 2008 and have been published on YouTube, iTunes and VideoLectures.net. The OCW concept has received an enormous amount of attention from all over the world, both from students as well as from universities. In 2005, the OpenCourseWare Consortium was established to advance education and empower people through open courseware. At present, about 200 higher education institutions and associated organizations from around the world are a member of this organization, including TU Delft, the Dutch Open University and HAN University of Applied Sciences (Hogeschool van Arnhem and Nijmegen). Because of the positive response on their OCW activities, MIT employs a special OCW office where close to 20 people are working every day.[30]

Walter Lewin

In 1999, MIT started recording the lectures of their most popular courses. Professor Walter Lewin is one of the most well-known lecturers today, who has been made famous through TV and the Internet. He is an extremely enthusiastic physics teacher who received his Ph.D. degree in nuclear physics in 1965 at Delft University of Technology. He joined MIT in January of 1966 as a post-doctoral associate and became an assistant professor later that year.[28]

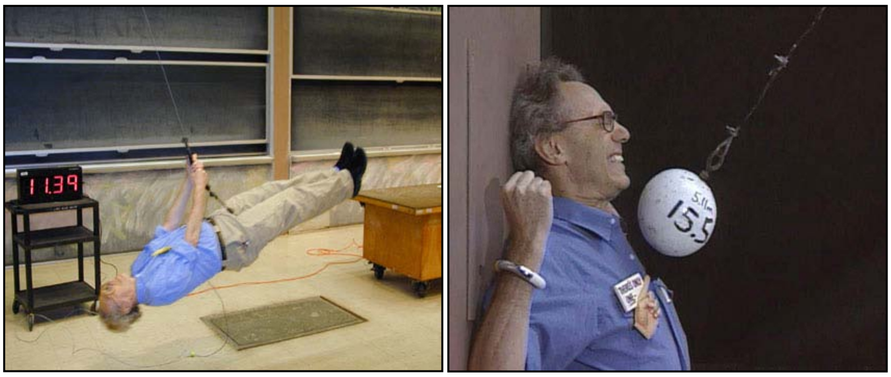


Figure 2.1: Prof. Walter H.G. Lewin, the YouTube superstar
(Source: http://bibliotematica.wordpress.com/2009/06/05/walter-lewin-quiero-morir-en-una-clase/
and http://www.pbs.org/kcet/wiredscience/blogs/2007/12/free-to-be-mit.html)

Even before the advent of MIT OpenCourseWare, Lewin's lectures could be found on UWTV in Seattle, where he reached an audience of about four million people, and on MIT Cable TV, where he helped freshmen with their weekly homework assignments. Lewin's lectures on "Newtonian Mechanics, Electricity and Magnetism" and on "Vibrations and Waves" comprise some of the most popular content on MIT OpenCourseWare. He consistently holds a spot in the most downloaded videos on Apple's iTunes-U as well as on YouTube-Edu. His unique style of teaching has captured the attention of a broad range of students, educators and self-learners.[28] Thanks to the various distribution channels that MIT OCW employs, the lectures of Walter Lewin now receive about 3,000 views a day, from people all over the world.

Online distribution

YouTube is the most popular website for online video content in the world. Nearly 20% of all global Internet users visit YouTube, an average of 16 page views per visit. In October 2009, they were ranked the 4th in the top 500 websites list, right after Google, Facebook and Yahoo.[33]

Since March 2009, YouTube has a special section for education called YouTube-Edu. In April of 2009, about 150 universities and colleges in the United States have submitted around 25,000 educational videos. 8 months later, in December of 2009, there are already 298 participating universities. The videos on YouTube-Edu are not all recorded lectures, but also short movies (6 to 12 minutes).[34]

iTunes-U is a part of the iTunes Apple Store. The service was created to manage, distribute, and control access to educational audio and video content for students within a college or university or for outside viewers. The member institutions are given their own iTunes-U site that makes use of Apple's iTunes Store infrastructure. The online service is without cost to those uploading or downloading material. Content includes course lectures, language lessons, lab demonstrations, sports highlights and campus tours provided by many top colleges and universities from the US, United Kingdom, Australia, Canada, Ireland and New Zealand.[35]

In November of 2009, iTunes-U holds over 200,000 educational audio and video files from top universities, museums and public media organizations around the world. About 200 international universities and colleges have published content on iTunes-U, including MIT, Yale, Stanford, UC Berkeley, Oxford, Cambridge, Freiburg, Lausanne, TU Aachen and Melbourne. The number of participating universities, as well as the number of audio and video files, has doubled in the previous 7 months.

Apart from iTunes-U and YouTube, which are commercial services, there are also a few websites who offer their services for other reasons. A popular example of this is VideoLectures.net. Their main purpose is "to provide free and open access of high quality video lectures presented by distinguished scholars and scientists at the most important and prominent events like conferences, summer schools, workshops and science promotional events from many fields of Science. The portal is aimed at promoting science, exchanging ideas and fostering knowledge sharing by providing high quality didactic contents not only to a scientific community but also to a general public."[32]

A recent addition to this group is Academic Earth, which launched in March of 2009. Their mission statement, as stated on their website: "Academic Earth is an organization founded with the goal of giving everyone on earth access to a world-class education".[31]

Video composition

Every MIT video has a camera angle that is fixed on the front side of the classroom. Most of the time a professor is walking in front of a whiteboard while explaining several course topics. The video camera follows the professor and zooms in and out on the blackboard whenever the professor is writing on it. Sometimes during the video, parts of the surrounding classroom are visible and you can see students sitting down and/or people walking in.

Most MIT professors only use the blackboard, while PowerPoint slides, overhead projectors or projected illustrations are rarely used. In case this does happen, the content of these slides are included in the video by zooming in on the projected screen, or the recorded video might show a text screen referring to the lecture material. These slides are published as a pdf file under the "Lecture notes".

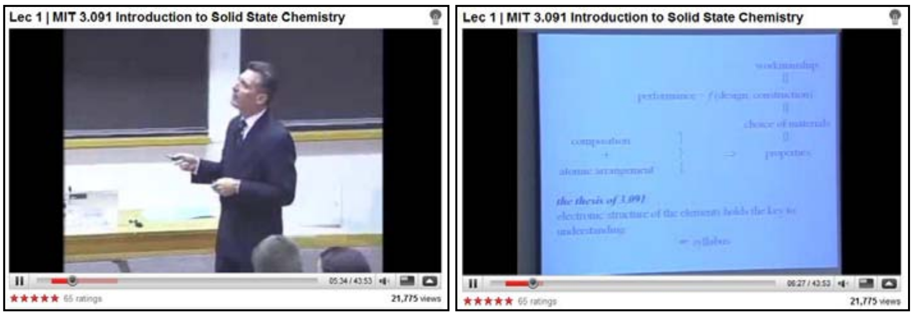


Figure 2.2: MIT lecture with a professor using slides, which are also included in the recorded video (Source: http://www.youtube.com/watch?v=R90sohp6h44)

The MIT lectures were initially recorded with two cameras; one camera was used for the overview and one camera took care of the close-ups of the blackboard. More recent recordings included two more cameras in the back of the classroom to provide a wider overview of the lecturer in front of the class. All these multi-camera lecture recordings had to undergo some form of post-production to work out the different camera angles, so that a single continuous video could be constructed combining all of the different recorded footage.

Transcripts, captions and annotations

About 60% of the recorded lectures at MIT are provided with a transcript. The transcripts are presented on the MIT-OCW website, on the page of the related lecture under the embedded YouTube movie. Most of the time these transcripts are also available as a pdf file.

In YouTube, these transcripts are used for the YouTube Caption option that shows subtitles in the bottom part of the movie. Captions or subtitles are available in YouTube since August of 2008.

2.2 Delft University of Technology

Development of Collegerama

In the year 2000, the section Multimedia Services (MMS) of Delft University of Technology started with the development of Collegerama in a pilot project on streaming media.[38][39] The main goal of this pilot was the recording of lectures which could be viewed by students within Blackboard, their digital learning environment. These “web lectures” were regarded as instruments to improve study results and to increase the efficiency of the education at the university.

MMS selected the commercially available Mediasite system, created by Sonic Foundry, as a basis for Collegerama. The term Collegerama is a private brand created by TU Delft, so that they could be independent from the technical infrastructure for their web lectures. Selecting a standard product avoids the high development cost for creating a new system. By using an existing solution, the university also has the added benefit of getting new updates and features within the Mediasite platform.

The early years

In April and May of 2004, Professor Barend Thijsse was teaching the BSc course TN2012 Quantum mechanics. He was giving the course for the last time, because he was leaving the university. Since he was recognized as an outstanding teacher, TU Delft wanted to record his lectures now that they still had the chance. He gave the course and lectures together with his successor, Professor Leo Kouwenhoven.

Figure 2.3: Older Collegerama lectures (TN2012) recorded in 2004 had a smaller video size (Source: http://collegerama.tudelft.nl/mediasite/Viewer/?peid=735a8c5902864988b01157c16f8e632e)

Mediasite was used for recording the 25 lectures (40-45 minutes) of the BSc course. A Tablet PC functioned as a blackboard to write notes on and both lecturers had a speaker microphone attached to their jackets. The recorded courses were used during the succeeding years as a reference until a drastic curriculum change in September of 2008.

After this successful project, there were 3 additional presentations recorded using Mediasite from September until December of 2004, as part of tests for the technical infrastructure of Collegerama. These web lectures were filmed with poor audio recording equipment (no special microphone for the speaker) and a small sized video recording (256x192 resp. 240x180). By that time, 240x180 was the standard video size for Mediasite recordings.

In January of 2006, Collegerama was used for the recording of the closing speech by the Rector Magnificus, Prof. Dr. Ir. J.T. Fokkema, at the 164th Dies Natalis of TU Delft. This was the start of a yearly tradition where all the Dies Natalis speeches were recorded. The video was recorded at a higher resolution of 320x240, which is still the standard Collegerama video resolution in 2009.

Between September and December of 2006, the 30 lectures (40-45 minutes) of the BSc course TN2545 Signals and Systems by Professor Lucas van Vliet were recorded. This course was normally given in Dutch, but for the sake of the recordings they decided to give them in English to allow non-Dutch speaking students to follow the course. The recorded lectures consist of videos showing the lecturer and synchronized screenshots of a Tablet PC, used as an interactive blackboard. These recorded lectures were actually used for several years, until in September 2009 a new lecturer took over the course. They are currently available on Blackboard as reference material.

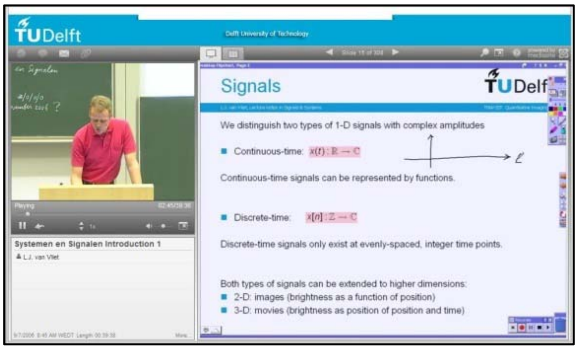


Figure 2.4: Collegerama recording using a Tablet PC as an interactive blackboard
(Source: http://collegerama.tudelft.nl/mediasite/Viewer/?peid=b7d4c81eed134ff68781e84ba05002e9)

Collegerama recording

Collegerama has two possibilities for recording lectures. They can either use a stationary setup that has been placed at a few classrooms at TU Delft, or they can use the mobile station which can be used at any given location. Both of the systems consist of a stationary webcam which can be operated remotely by use of a joystick. The operator, usually a student aid, makes sure that the camera is always pointed at the lecturer while he or she is moving around the classroom. The laptop that comes with the presenter unit is connected to a beamer, so that the PowerPoint slides can be viewed in the classroom and recorded by the system. The recording system takes screenshots of the beamer screen, based on computer activity. Every 1 to 4 seconds, the system checks for a change on the screen. If a different slide has been loaded or the position of the mouse has been changed, a new screenshot will be saved as a jpeg image file.

Figure 2.5: Stationary and mobile recording unit of Collegerama

The current system used to record the slides of the lectures relies on the fact that changes on the screen always correspond to a change in the presentation. This is clearly not the case and several scenarios can cause a faulty screenshot to be taken:

- a video is played within a PowerPoint slide
- the lecturer inadvertently moves the mouse
- the lecturer leaves PowerPoint to demonstrate an application on his PC

This recording flaw creates a problem, because a lot of Collegerama lectures contain a lot of abundant images that were accidentally saved. Some of these lectures contain 400 screenshots, when in fact the original PowerPoint presentation only had about 50 slides. While playing the lectures, the interface relies on the screenshots that are created during the recording for navigation. The problem with this navigation system is that once the lecture contains an overflow of useless slides, there is no other way of browsing through the lecture except for the video timeline. An example of the navigation element in such a lecture is shown in Figure 2.6.

Figure 2.6: Screenshot of a Collegerama lecture with too many screenshots because of mouse movement

After the lecture has been given, the data is sent to the presentation server. It will process the different data sources and create three different outputs:

- an audio/video stream (wmv file)
- pictures of the different PowerPoint slides or computer screenshots (jpeg files)
- different settings and additional information about the lecture (xml file)

The presentation server will synchronize all the different elements and will store the required information in the xml file. This information will later be used to correctly display the video in combination with the screenshots. When the presentation has been processed, it is written to the Collegerama web server and is now available for students with Internet access all over the world.[42]

Presentation options

During the presentation, the lecturer is provided with three different presenting options:

- *blackboard*
  The lecturer uses the blackboard or an overhead projector to give his lecture, while the video camera records the content.
- *PowerPoint*
  This works in combination with a prepared set of PowerPoint slides that will be displayed while the presentation is being given.
- *screen capturing*
  The contents of the computer screen will be displayed during the presentation, which allows for the lecturer to use external software such as computer simulations or written text on a Tablet PC and record the results as separate screenshots.

Figure 2.7: Examples of the three presentation options

Each of these presentation options uses the same storage system, which is based on screen activity. Especially while using the blackboard or desktop methods, there will be an abundant amount of images stored, since every mouse movement and change on the screen, when writing down notes, will cause a new screenshot to be saved. Collegerama uses a uniform view for all three presenting options, as is shown in the examples given in Figure 2.8.

Figure 2.8: Collegerama screenshots of the three different presentation options
(Source: http://collegerama.tudelft.nl/mediasite/Viewer/?peid=ca42dce5-bb51-4c39-93de-50528dd6b880
and http://collegerama.tudelft.nl/mediasite/Viewer/?peid=724886f7-cfd0-441d-ae85-1fae0cbb28a1
and http://collegerama.tudelft.nl/mediasite/Viewer/?peid=b7d4c81eed134ff68781e84ba05002e9)

Figure 2.8 illustrates that Collegerama is suitable for the showing of lectures in which a PowerPoint presentation or a Tablet PC is used (middle and right screenshot). In these cases the most detailed information is presented on the presentation block. For a lecture with blackboard only (left screenshot), the Collegerama system is a little superfluous. The three presentation options differ significantly in the number of slides (or screenshots). This difference is illustrated in Table 2.1.

Table 2.1: Number of slides/screenshots for the three presentation options

Presenting option	**Number of slides**	**Navigation pages (list – small – large)**
Blackboard	0 (no slides picture)	0 - 0 - 0
PowerPoint	30	2 - 3 - 5
Screen capturing	308	12 - 15 - 29

With respect to navigation, only lectures with PowerPoint slides seem to be suitable for a Collegerama recording. Blackboard lectures lack the navigation by slides/screen shots, while Tablet PC lectures have too many screenshots for a proper navigation. For the latter, the screenshots can be clustered in chapters as part of the post-processing process of a Collegerama recording.

Collegerama as a service

Starting in September of 2007, Collegerama became part of the regular facilities for education at TU Delft, under the responsibility of the University Corporate Office for Education and Student Affairs (O&S). This office is also responsible for the electronic learning system Blackboard. As a consequence, recording of lectures was financed by the Corporate Office and became free for the lecturers at the different faculties. Before that time recordings were made at a rate of € 500,- per recorded session of 45 minutes. The scheduling of recording units and operators is now organized by O&S and lecturers can apply there to have their lectures recorded. This service has resulted in a huge increase of recorded lectures. In September and October of 2009 alone, around 60 to 75 lectures were recorded each week (30 to 40 sessions of 2 lectures of 45 minutes each). This amounts to 5% of all lecture hours given each week at TU Delft.

In September of 2009, the faculty Mechanical Engineering was faced with a huge student overflow. The 500 first-year students did not fit in their largest lecture room available, which had a capacity of 300 students. To overcome this problem, they used two lecture rooms. In one lecture room, the lecturer gives the live lecture which was recorded using a high quality camera. This recording was then streamed to the other lecture room via a larger data stream to accommodate the higher quality. The recorded lectures were afterwards also available at a lower quality via Blackboard and Collegerama. The faculty called this service: “lectures in a movie theater”.

Figure 2.9: The Collegerama lecture at Mechanical Engineering is streamed to the “movie theater” next door (Source: Delta 27, 17 September 2009)

Collegerama live

The mobile recording units of Collegerama have a personal storage unit. After the lecture recording has been completed, the stored data is uploaded to the central Collegerama server. It is also possible to stream this recording to the server immediately while recording, thus generating a live stream to the outside world. This live streaming process has a 5 to 10 second delay between recording and broadcasting. In the Collegerama setup, a URL for a Collegerama lecture is automatically created 4 hours before the recording. This URL is published before the lecture starts, so that every student can watch it from their own room or any other location that has live Internet access.

This live streaming system was used for the course CT2011 Watermanagement in September-October of 2009. The course was moved within the curriculum from the third year to the second year, which caused the student attendance to double to about 500 students. This again largely exceeded the maximum seating capacity of the largest classroom available at the faculty of Civil Engineering (it holds only 350 students). To reduce the number of students attending, the lectures were scheduled on Monday and Friday during the first two lecture hours. The lectures were also announced to be broadcasted live and received a wide media attention under the title “lectures in bed”. The system was a huge success. After the initial lecture, the number of attending students reduced to around 100 attendees, with a large number of online viewers during lecture hours or several hours after the lecture. The movie theater lecture room stayed empty after the first lecture.

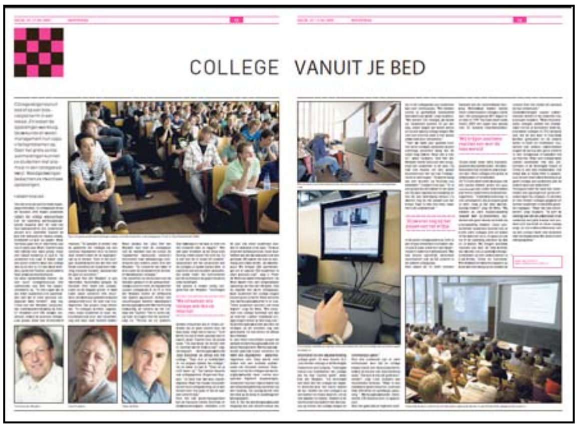

COLLEGE VANUIT JE BED

Figure 2.10: The Collegerama live streamed online recordings (CT2011) were announced as “lectures in bed” (Source: Delta 27, 17 September 2009)

OpenCourseWare

In March 2007, TU Delft started its own OpenCourseWare pilot project.[40] In this pilot project the course material of about 20 MSc courses from 6 different disciplines were published. Collegerama lecture recordings were part of this material. This initiative was very well received and students found the Collegerama recordings to be of extraordinary quality. Because of the national and international response, TU Delft decided in 2008 to continue its OpenCourseWare program at a more extensive scale.

In October 2009, TU Delft hosted the yearly conference of the OpenCourseWare consortium, in which more than 200 universities worldwide are active.[30] The Director of Education and Student affairs of TU Delft is a member of the board of the OCW Consortium.

In January 2010, TU Delft has renewed its OpenCourseWare website. One of the goals for update was to give the recorded lectures a more pronounced exposure and to give it the look and feel of the original Blackboard courses. A month later, they have created an iTunes-U account and have started to publish recorded lectures, partly as a result of the work that has been described in this research report. Annex C gives further information on this subject.

Future developments

In 2010 a new viewer for Collegerama will be implemented at TU Delft (Mediasite version 5.2). This Silverlight player has the look and feel of viewing YouTube movies, with a small slide viewer for navigation and of viewing iPhone or Windows 7 screens (dynamic screen changes). This new viewer will still encounter the major drawback of having an overload of useless slides, because this problem is related to the recording process, not to the viewer.

2.3 University of Twente

In 2007, University of Twente started a pilot project on recorded lectures. This pilot project used the experience of TU Delft with its Collegerama system. The same technical infrastructure of Collegerama was also used at University of Twente. Within the pilot project, the lectures of 10 BSc courses have been recorded. One of these was the course Algorithms, Data structures and Complexity (214020), which is also a pre-master course for the master Computer Science. Between November 2007 and January 2008, 8 of their lecture sessions have been recorded. Afterwards the 7th lecture session was not available due to technical difficulties. The recorded sessions include two lecture hours (40 minutes each) and the intermediate coffee break (a 15 minute recording of a clock).

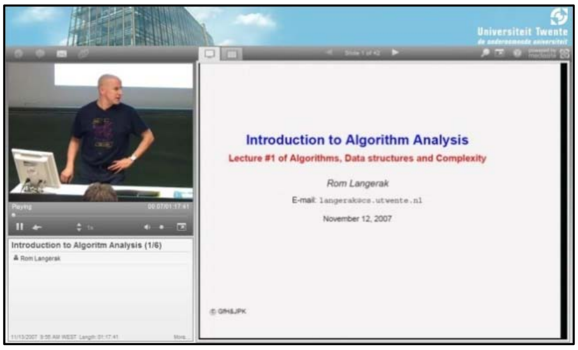


Figure 2.11: Collegerama recording (214020) at University of Twente in November 2007
(Source: http://videolecture.utwente.nl/mediasite/Viewer/?peid=bcb88779-b54c-4d38-a028-34b7f1d0dfdb)

After each recorded course, an evaluation form was used to register the opinion of the students. Based on the positive results of the pilot project it was decided to continue the project. Since September 2008, lectures at University of Twente can be recorded with Collegerama i.e. Mediasite.[41] At University of Twente, two lecture rooms are available with recording facilities for Collegerama/Mediasite (Horst C101 and Cubicus B209). There is also one mobile recording unit available (Spiegel, for room 1, 2, 4 and 5). This unit can also be used in other buildings and lecture rooms, if requested. The service for recording lectures is free of charge and is provided by the ICT Service Centre of University of Twente.

In September 2009, University of Twente started using Blackboard as its digital learning system as a replacement for Teletop. At present, TU Delft and University of Twente use the same technical infrastructure for their digital learning environment as well as their lecture recording and streaming system.

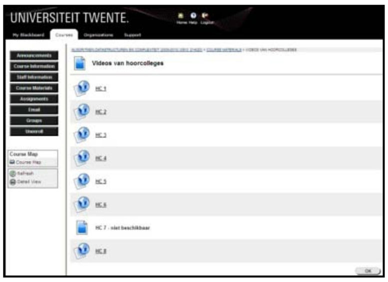


Figure 2.12: Overview of video lectures (214020) in Blackboard (2nd quarter of study year 2009-2010)
(Source: http://blackboard.utwente.nl/webapps/blackboard/content/listContent.jsp?course_id=_758_1&content_id=_92264_1)

2.4 Summary

Massachusetts Institute of Technology (MIT) was the first university to start recording their lectures back in 1999, by taking a video camera into the classroom and video-taping the lecturer as he was teaching. Several years later, TU Delft started doing the same, using a more sophisticated system that simultaneously records the slides along with a video stream of the lecturer. This made the recorded lectures easier to follow, but also added a problem. The lectures were no longer contained within a single video file, which severely limits the possibilities for different online distribution channels.

In 2007, University of Twente decided to use the same system for the recording and distribution of their lectures as TU Delft. After several pilot projects, they purchased 2 stationary recording units and 1 mobile recording unit of Collegerama.

The current Collegerama system has several problems:

- navigation, since it is based on inconsistent screenshots of slides
- not distributable through a single (video) file
- no easy way of browsing/searching through a lecture

3. Distribution platforms

Collegerama / Mediasite player

At present, lectures recorded in Collegerama can only be viewed as streaming video with an Internet connection to the Collegerama server. The movies are played within the custom Java player developed by Mediasite. This setup has several advantages:

- no distribution channels required, avoiding its institutional and technical requirements
- single point of entry, with its benefits on updating (its content as well as the player)
- no storage required at the point of viewing/listening

Aside from these advantages, there are also a number of severe drawbacks to the current Collegerama distribution platform:

- limited distribution options
- no offline viewing
- compatibility
- limited expansion options

Limited distribution options

The current Collegerama system can be divided into two parts:

- video stream of the lecturer (wmv)
- screenshots of PowerPoint slides (jpg)

During playback, the web player will update the screenshots based on a time index that is stored in the configuration file of the lecture. Basically, a video stream is played and the corresponding pictures are reloaded on the right side of the viewer during playback. Virtually all online distribution platforms operational today require a video file to be uploaded. This file will usually be re-encoded using a specific codec compatible with that player. YouTube for instance uses mp4 as the way of storing its online video files. Unfortunately this poses a problem when distributing lectures stored within the Collegerama server over any of these other multimedia platforms. It is possible to upload the video stream, since that component is stored in a video file format, but without the lecture slides to accompany it the lecture will miss most of its important content.

No offline viewing

Since all the lectures are streamed over the Internet, it is not possible to view the Collegerama lectures without an active Internet connection. This means that it's not possible to store the lectures and view them later on your laptop, iPhone/iPod or other mobile multimedia device.

Compatibility

The current player that is being used within Mediasite is based on Microsoft Silverlight, which has a bad compatibility with other operating systems such as Linux. There is a custom made version available created by Novell, but this solution won't always work when Mediasite releases a new version of their player. Users are dependent on the developments by Novell to keep their system up to date.[43]

Limited expansion options

At present, the Mediasite player cannot be easily integrated with (multi-language) subtitles. This might be improved in future versions, but Collegerama is dependent on the Mediasite developments in order to add custom functionality. Other channels such as YouTube do provide these options as a default and are ahead of Mediasite in this area.

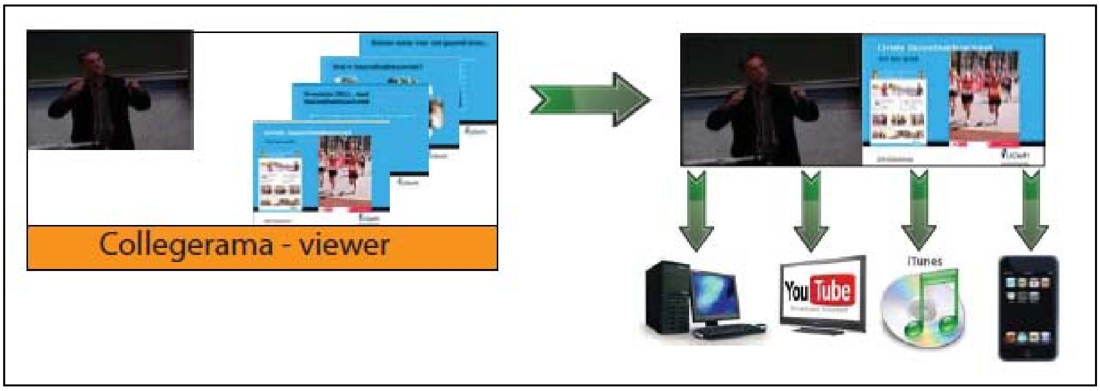


Figure 3.1: Combining the Collegerama components into a single video file enables broader distribution of recorded lectures

Other distribution platforms

In this chapter there are two important platforms for which the options and capabilities have been researched, YouTube and iTunes. These two have been selected for the following reasons:

- their worldwide exposure
- the acceptance of their technical specifications by other external platforms
- the experiences of MIT (see Annex A)
- the compatibility of these technical specifications on TU Delft's own Blackboard learning environment, the OpenCourseWare website and other web platforms

The distribution of recorded lectures through these platforms requires the creation of a single video file, which can be uploaded to their server. For the creating of such a Collegerama vodcast/podcast, the following aspects should be examined:

- content (slides, audio, video, subtitles and any combination of these)
- presentation of the content (lay-out, introduction tune/movie, branding)
- video quality (resolution, frame rate)
- format of video file (mov, wmv, flv, mp4, codec etc)
- audio quality (stereo/mono, frequency range)
- format of audio file (mov, mp3, mp4, codec etc)

Above mentioned technical specifications (quality, codec) primarily determine the file size. The technical specification should balance between quality (usability) and quantity (download time and storage requirements).

Outline

This chapter will focus on the distribution of Collegerama over various different platforms. The popular audio/video sharing mediums YouTube and iTunes will first be covered. After that, a new lecture creation tool called Adobe Presenter will be demonstrated. Each of these platforms will be thoroughly examined and a conclusion will be made about the quality of each of these systems.

For further background information about the research on this topic, see Annex C.

3.1 YouTube

YouTube is a video sharing website where users can upload and share their videos. Three former PayPal employees created YouTube in February of 2005. In November 2006, YouTube was bought by Google Inc. for $1.65 billion and is now operated as a subsidiary of Google. It uses Adobe Flash Video technology to display a wide variety of user-generated video content and is currently the biggest distributor of streaming online video content.

Unregistered users can watch the videos, while registered users are permitted to upload an unlimited number of videos. Videos that are considered to contain potentially offensive content are available only to registered users over the age of 18. The uploading of videos containing copyright violations is prohibited by YouTube's terms of service. Accounts of registered users are called "channels".[44]

In the last few years YouTube became a medium for several Universities to publish their recorded lectures on. One of the first was MIT (Massachusetts Institute of Technology), who joined in October of 2005. Later, other Universities like Purdue (2006), Stanford (2006), UC Berkeley (2007) and Harvard Business (2007) started publishing recorded lectures and course material via the popular Internet medium.

Video formats for YouTube

YouTube's video playback technology, based on the Adobe Flash Player, allows the site to display videos with quality comparable to more established video playback technologies such as Windows Media Player, QuickTime, and RealPlayer. These formats generally require the user to download and install a web browser plug-in to view video content. Viewing Flash video also requires a plug-in, but market research from Adobe Systems has found that its Flash plug-in is installed on over 95% of the personal computers around the world.[45]

Videos uploaded to YouTube are limited to ten minutes in length and a file size of 2 gigabyte.[47] When YouTube was first launched in 2005, it was possible for any user to upload videos longer than ten minutes, but YouTube's help section now states: "You can no longer upload videos longer than ten minutes regardless of what type of account you have. Users who had previously been allowed to upload longer content still retain this ability, so you may occasionally see videos that are longer than ten minutes."[46] The ten minute limit was introduced in March 2006, after YouTube found that the majority of videos exceeding this length were unauthorized uploads of television shows and films.

Video formats and quality

YouTube accepts videos uploaded in most formats, including .WMV, .AVI, .MKV, .MOV, MPEG, .MP4, DivX, .FLV, and .OGG. It also supports 3GP, allowing videos to be uploaded directly from a mobile phone.

They originally offered their videos in only one format, but now use three main formats, as well as a "mobile" format for the viewing on mobile phones. The original format, now labeled "standard quality", displays videos at a resolution of 320x240 pixels using the Sorenson Spark codec, with mono MP3 audio. This was, at the time, the standard for streaming online videos.

"High quality" videos, introduced in March 2008, are shown at a resolution of up to 860x480 pixels with stereo AAC sound. This format offers a significant improvement over the standard quality. In November 2008, 720p HD support was added. At the same time, the YouTube player was changed from an aspect ratio of 4:3 to a widescreen 16:9 resolution. 720p videos are shown at a resolution of 1280x720 pixels and encoded with the H.264 video codec. They also feature stereo audio encoded with AAC.

Collegerama components

A Collegerama lecture has screenshots of the PowerPoint slides and a video of the lecturer giving the lecture. On the web interface, these have been split up into separate parts. If the recorded lectures in Collegerama are to be published as a vodcast, the different elements need to be combined into a single multimedia file format.

In the current video system of Collegerama, the following elements are kept in sync:

- video of the lecturer
- audio
- PowerPoint slides
- closed captions/subtitles (not currently used at TU Delft)

Figure 3.2: The two main components of Collegerama, video and slides

Video of the lecturer

The video part of Collegerama usually shows the lecturer, but might occasionally be switched to a recording of the display screen for animations, movies etc. Collegerama publishes the video stream using the following quality settings:

Resolution: 320 x 240 (ratio 4:3)
Frame rate: 25 fps
Bit rate: 370 kb/s
Codec: wmv3
In short: Windows Media Video 9 / 320x240 / 25.00fps / 341kbps

Audio

Audio is an important part of the vodcast. It contains all the spoken text and explanations by the lecturer. A lecture can be followed by only having an audio recording without video, but not the other way around. This is shown by podcasts of lectures. A video stream without audio doesn't make any sense. Collegerama publishes the audio stream using the following quality settings:

Channels: 2 (Stereo)
Sampling rate: 22050 Hz (22 kHz)
Bit depth: 16 bits/sample
Bit rate: 20 kB/s
Codec: wma2
In short: Windows Media Audio 9.2 / 20 kbps / 22 kHz / stereo (1-pass CBR)

PowerPoint slides

The slides of a presentation contain the most detailed information. It's important for the viewers since it gives a guideline to the story. Fortunately the slides mostly contain keywords at a pretty decent font size, which means that the quality and resolution do not have to be high for it to be readable. Collegerama publishes PowerPoint slides using the following specifications:

Resolution: 1024 x 768 (ratio 4:3)
Bit depth: 24 bits/pixel (full color)
Codec: jpg

Closed captions / subtitles
There are different ways of publishing closed captions or subtitles on video. The most commonly used method is a text file containing the spoken sentences along with their corresponding timestamps. Closed captions and subtitles for Collegerama lectures are described elsewhere. For the production of a vodcast, the subtitle files are not relevant since they will be attached to the vodcast based on the internal timestamps of the movie file.

Publishing Collegerama on YouTube
A vodcast for YouTube should comply with the restrictions for resolution of YouTube. A general strategy for this is to develop a vodcast at the best video quality supported by YouTube, with the following considerations and constraints:

- movie size is limited to 2 gigabyte
- display size is limited to 10 minutes for the general public, unlimited for channel managers like YouTube-Edu
- YouTube gives the viewer the option to display at a lower quality when bandwidth is a limiting factor
- producing a vodcast at the highest quality enables the production of “child products” for other platforms with a lower quality, which results in smaller file sizes or bandwidth requirements
- YouTube converts movies with non-normalized resolution by downsizing to the nearest standard heights of 360, 480, 720 or 1280 pixels

Within these constraints, the best quality of a Collegerama vodcast for YouTube can be achieved by following these steps:

- reduce the size of the slides from 1024x768 to 960x720 (downsizing to 94%, keeping the display ratio 4:3)
- leave the video resolution at 320x240
- put both elements alongside each other, giving an overall size of 1280x720 (HD720, widescreen, display ratio 16:9)
- fill the remaining area with related info, navigation tools or leave them blank

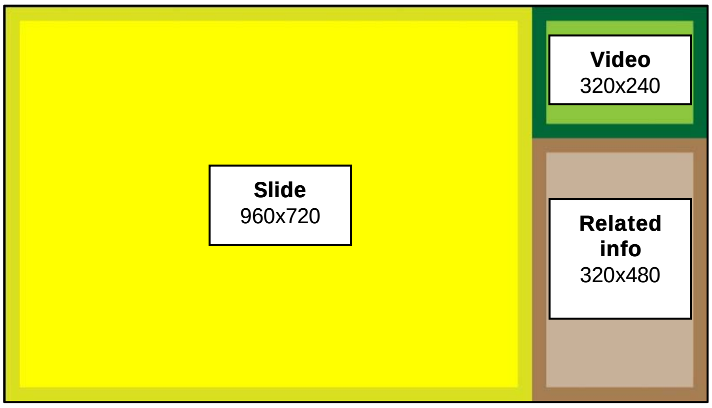


Figure 3.3: Layout of Collegerama elements within the resolution constraints for YouTube movies (1280x720)

A layout according to this setup is given in Figure 3.3. The video is located on the right-hand side of the slides, to give a more balanced overall picture for left to right reading. The overall view could be mirrored to obtain an overall picture which resembles the original Collegerama view, where the video is located on the left. A screenshot of a lecture converted to the resolution requirements and uploaded to YouTube can be seen in Figure 3.4.

Figure 3.4: Collegerama as a vodcast for YouTube (1280x720)

Vodcast production

Single audio files are often referred to as "podcast" files. The term podcast originates from the iPod, as iPod-broadcasting. In the slipstream of this term, single movie files are often referred to as "vodcast" files. Originally these were downloaded files since iPod and iTunes did not support streaming content. The meaning of these terms has later transferred into "audio on demand" or "video on demand (VOD)", in combination with an RSS feed. This audio or video can also be streaming audio or video, without actual distribution of a real file.

The most important step in the production of a downloadable vodcast out of a Collegerama recording is the conversion of the PowerPoint slides into a movie. This can be achieved with the help of screen capturing systems such as Camtasia Screen Recorder. These systems record an assigned part of the display screen into a movie file. By playing a Collegerama lecture, the slides can be recorded as a movie with the right time-framing. Figure 3.5 gives an impression of such a screen recording.

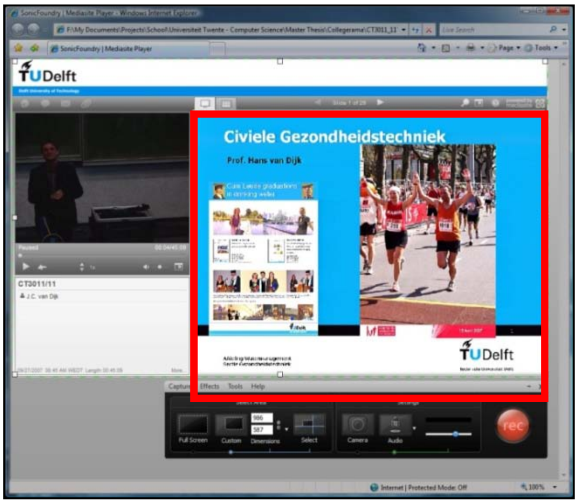


Figure 3.5: Converting PowerPoint slides into a movie file by recording the Collegerama slide-display

This screen recording resulted in a movie file of 39 MB (1024x768, 15 fps, wmv3), which is only 6.3 times the total file size of the 29 slides (1024x768, jpg). The wmv3 compression proves to be efficient when recording still pictures, since the original 29 pictures have been converted into over 40,000 picture frames. The captured slides movie and the Collegerama movie have been combined into a single HD movie file of only 88 MB (1280x720, 15 fps). This is only (88/117=) 75% of the original small sized Collegerama movie (320X240, 25 fps). The reduction is caused by a lower frame rate and the efficient compression of the wmv3 codec for still pictures. Converting this movie file into the H264 codec increases the file size to over 500 MB. This shows an inferior compression of the H264 codec over the wmv3 codec for this type of movie, typically including large areas with still pictures.

Scientific research on compression efficiency of these two codes shows less significant differences.[7][8] The common opinion is that the compression of these two codes is similar, but wmv3 (VC-1) would require less processor power for encoding and decoding. The differences in architecture might result in larger differences in specific situations. Moreover, the achieved compression with these codecs is also influenced by the efficiency of the encoding software. Wmv3 (VC-1) has more advanced features for motion compensation with a more flexible block sizing, which might be the main cause of the observed differences. The creation of a HD movie from a Collegerama recording increases the movie resolution with a factor of 4x4, allowing for much better display of subtitles as is shown in Figure 3.6.

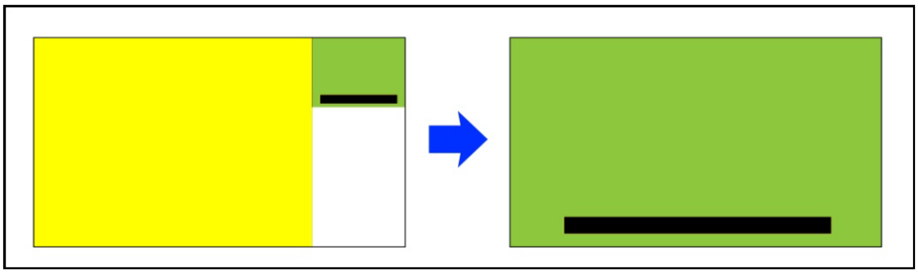


Figure 3.6: Vodcast of a Collegerama recording converts a small-sized video into a HD movie with room for proper subtitles

Above described production of a vodcast is rather labor and time consuming. A more or less similar result could be obtained by doing a one step recording session, where the overall Collegerama display is recorded by Camtasia.

3.2 iTunes

iTunes is an application that allows the user to manage audio and video on a personal computer, acting as a front-end for Apple's QuickTime media player. Officially, iTunes is required in order to manage the audio of an Apple iPod portable audio player (although alternative software does exist). Users can organize their music into playlists within one or more libraries, copy files to a digital audio player, purchase music and videos through its built-in music store, download free podcasts and encode music into a number of different audio formats. There is also a large selection of free internet radio stations to listen to.

Version 4.9 of iTunes, released on June 28th 2005, added built-in support for podcasting. It allows users to subscribe to podcasts for free using the iTunes Music Store or by entering the RSS feed URL. Once subscribed, the podcast can be set to download automatically. Users can choose to update podcasts weekly, daily, hourly or manually. It is also possible to select podcasts to listen to from the Podcast Directory, to which anyone can submit their podcast for placement. The front-page of the directory displays high-profile podcasts from commercial broadcasters and independent podcasters. It also allows users to browse the podcasts by category or popularity and to submit new podcasts to the directory.

Video content available from the store used to be encoded as 540 kbit/s protected MPEG-4 video (H.264) with a 128 kbit/s AAC audio track. Many videos and video podcasts currently require the latest version of QuickTime, version 7, which is incompatible with older versions of Mac OS (only v10.3.9 and later are supported). On September 12th 2006, the resolution of video content sold on the iTunes Store was increased from 320x240 (QVGA) to 640x480 (VGA). The higher resolution video content is encoded as 1.5 Mbit/s (minimum) protected MPEG-4 video (H.264) with a minimum of 128 kbit/s AAC for the audio track.

Video formats for iTunes

The main focus of iTunes is to distribute content to the Apple iPod and its successors. The original iPod was not provided with a video screen for movie display until October of 2005. The iPod Nano received a movie display in September 2007. The screen size of the iPod family is shown in Table 3.1.

Table 3.1: Screen sizes of the iPod and its successors

Type	Introduction date	Screen size	Aspect ratio
iPod video	October 2005		1.33 (4:3)
iPhone	June 2007	480 x 320	1.5 (3:2)
iPod Touch	September 2007	480 x 320	1.5 (3:2)
iPod Nano	September 2007	320 x 240	1.33 (4:3)
iPod Nano (new)	September 2009	376 x 240	1.57
Supported video (external screen)		640 x 480	1.33 (4:3)
HD movies			1.78 (16:9)

Over the years, the different iPod versions have evolved to larger screen sizes and wider screens (higher aspect ratio). If the iPhone aspect ratio is compared to the HD widescreen ratio used today, the iPhone is somewhere in between the traditional TV and the HD widescreen standards. All iPods support a video display of a maximum of 640x480 by use of an external screen. As widescreen HD video has become more or less the standard nowadays, it looks like Apple will someday also transform into larger video displays with HD specifications.

iPod constraints for Collegerama vodcasts

For the development of a Collegerama vodcast for iTunes (and iPods), the following aspects are of concern:

- the rather low resolution of the screen
- the different aspect ratio

These constraints have consequences for the following design aspects:

- the size of the display
- the size of the video component
- the location of the video component (upper/lower and left/right corner)

Low resolution

The resolution of the iPod is the same as that of the Collegerama video component. This would allow for simple distribution, using just the video stream as a vodcast and leaving out the presentation slides. Such a setup is used at MIT and many other universities. However, the slides in Collegerama provide a lot of the lecture content, since the keywords and a large part of the subject matter is on it.

In an alternative setup, the vodcast might include the slide part of Collegerama with the audio of the video component. This is only an adequate alternative if the slides are readable at this low resolution. Figure 3.7 gives an example of a typical PowerPoint slide at a normal iPod resolution. It shows that the smaller fonts in a presentation are no longer readable at the low iPod resolution, but the typical PowerPoint fonts can still be read quite well. The iPod resolution is around (320/1024=) 30% of the maximum slide size in Collegerama and (320/640=) 50% of the slide size in an overall Collegerama display.

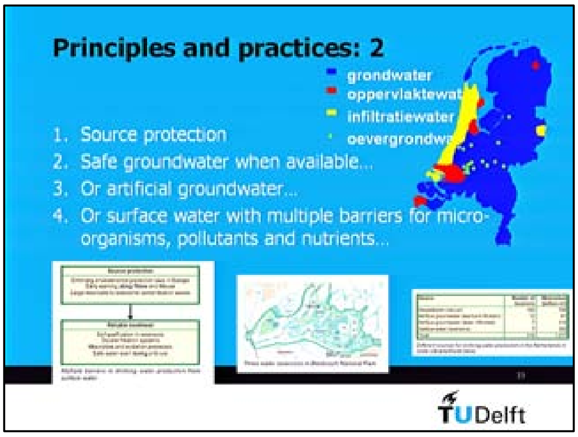


Figure 3.7: A typical PowerPoint slide at iPod resolution (320x240)

Different aspect ratio
The iPod aspect ratio is the same as both the slides and the movie components in Collegerama. Therefore combining these two components in a widescreen view, as is done in the previous YouTube vodcast, is not possible. Alternative solutions are:

- the slide components are not included (video only)
- the video component is not included (audio only)
- the video component is included at a smaller size (picture-in-picture)
- the video component is included at a smaller size (side-by-side) with unequal scaling

An example of these images is shown in Figure 3.8, which gives an impression of the latter three options.

Figure 3.8: Collegerama vodcasts with different options for the video component at iPod aspect ratio

From Figure 3.8, it is concluded that the most convenient option for including the movie component is the picture-in-picture layout. This is based on the following considerations:

- the slides should be shown at a maximum size for proper readability (no side-by-side)
- the movie component can be reduced to a small size (thumbnail) and still remain properly visible
- the audio component without the video component misses a focus point for the viewer (the movements of the lecturer give a better understanding of the lecture)

Size of display
An important aspect in the design of a vodcast for iTunes is the display resolution selected for the production and distribution. The design strategy for creating the smallest file size looks most promising, for the following reasons:

- vodcasts for iTunes should be downloaded to and stored on the iPod of the viewers (download time and storage capacity are relevant factors now, which is not the case in a streaming video setup)
- small file-sized vodcasts will minimize the requests for other small sized output options like podcasts (audio only), which would require additional production and distribution efforts (time, costs, organization)
- a small sized design gives a larger differentiation to the YouTube HD quality design

- iTunes-Uses the H264 codec, which is not as efficient in video compression as the wmv3 codec used in the YouTube design, so a smaller display size will be more relevant for a less efficient compression
- the smallest display design allows for viewing on the older iPods, which is still the majority of the iPods currently in use

For above mentioned reasons a vodcast for iTunes will be produced with a display size of 320x240 pixels.

Size of video component
The video in Collegerama shows the lecturer talking to the attendees. For this function a small video size is sufficient as the most important aspect of such a movie is its audio component (spoken words). This is shown in Figure 3.9, in which the original video resolution (320x240) is downsized to 10% of its original size. It shows that downsizing the Collegerama video to 20% (64x48) still gives supporting visibility of the speaking lecturer.

In some recorded lectures, the lecturer is writing text on the blackboard or is presenting experiments. Both circumstances require a larger display size for proper viewing and a full switch from the slide view to the video component might be useful. The downside is that this will require an extensive video-editing process which might also need the input of the lecturer. These constraints are not within the scope of a vodcast production out of a Collegerama recording. Production of a vodcast should be possible within a fully automated production process.

Figure 3.9: Collegerama video in original size (320x240) and reduced to 30%, 20% and 10%

The video component in a picture-in-picture design with the slides on the background will cover part of the slides, reducing its readability. This can be minimized by doing the following:
- selecting a small video component (10%-20%)
- making the video component (partly) transparent, still allowing for a background view (this setup might allow for a larger video size than a non-transparent movie, 20%-30% instead of 10%-20%)
- placing the video component in an area with the lowest disturbance of the slide view

Location of the video component
The video component should be located on the least disturbing part of the slide. Figure 3.10 gives an impression of these locations for a TU Delft PowerPoint slide at iPod resolution. It shows that the upper-left corner and the lower-right corner are unsuitable for movie insertion. The upper-left corner hides the important slide title, while the left corner hides the slide number. The lower-left corner hides the TU Delft logo and the upper-right corner might hide part of the slide title. Both locations can be deemed acceptable.

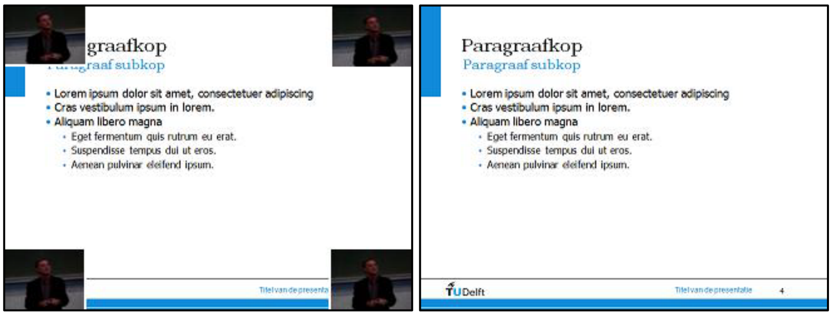


Figure 3.10: PowerPoint slide in TU design at iPod size, with and without inserted movie components (20%)

The lower-left corner might have a small advantage since this resembles the general lecture room layout at TU Delft, in which the lecture desk is in the front-left and the projection screen is located in the upper-center or upper-right part of the lecture room. This lecture room layout results in many Collegerama recordings showing the lecturer looking to his/her upper-left. With a movie component in the upper-right corner, the lecturer often seems to look up into the “sky”.

It should be noticed that not all lecturers use the standard TU Delft PowerPoint design. If the lecturer would have been made aware that his Collegerama recording is transformed into an iPod vodcast, he or she might adjust the slides to keep a certain corner of the slide empty. Therefore a uniform predesigned position of the movie component is important.

3.3 Portable Document Format (PDF)

Portable Document Format (pdf) is a file format created by Adobe Systems in 1993 for document exchange. It is used for representing two-dimensional documents in a manner independent of the application software, hardware and operating system. Each pdf file encapsulates a complete description of a fixed-layout of 2D document that includes the text, fonts, images and 2D vector graphics which compose the documents.[48] The great thing about pdf files is the fact that all the data of a document is frozen and “digitally printed”, so that it cannot be edited and all the layout properties are fixed. Over the years, it has become the standard medium for distributing and sharing documents online.

A new development at Adobe is the release of Adobe Acrobat Connect Pro (formerly called Macromedia breeze). It allowed for a new way of creating general presentations, online training materials, web conferencing, learning modules and user desktop sharing. The entire product is Adobe Flash based.[49] The module for creating lectures based on PowerPoint presentations is a plug-in called Adobe Presenter.

Figure 3.11: Adobe Presenter allows the creation of lectures based on PowerPoint

There are several advantages that come with the use of Adobe Presenter, as opposed to Collegerama:

- better navigation
- higher slide quality
- distributable through a single pdf file

Navigation

As you can see in Figure 3.12, the Adobe Presenter interface creates an automatic index based on the different slides. On the right side you can see each slide title, which is automatically extracted from the PowerPoint file. The great thing about this feature is the fact that there's a clear way of navigating through a lecture based on keywords taken from the lecture material. This is an option that Collegerama does not have.

Figure 3.12: Screenshot of lecture CT3011 implemented within Adobe Presenter

Adobe Presenter has a different navigation system compared to Collegerama. Instead of having a video stream that has several images of PowerPoint slides linked to it, it uses a different approach by placing the PowerPoint presentation at the heart of the interface. This means that there is no long video of 45 minutes with a main timeline. It splits the presentation up into separate timelines per slide. Each of these has its own short video attached to it with a separate timeline. As you can see in Figure 3.12, a video of 7 minutes and 25 seconds is playing along with the first introductory slide. The problem with such a system is that it requires the video recording of the lecturer to be split up into smaller segments and linked to each separate slide. This is a time consuming process.

Slide quality

Since the Adobe Presenter system makes use of the original PowerPoint presentation, it has all the slides digitally available at the highest quality. Once the lecture is converted to a shareable format, the quality of the sheets is no longer limited to a set resolution (1024x768 for Collegerama), but is stored as a vector oriented image. This means that the viewing quality is incredibly high compared to Collegerama.

Distributable through a single pdf file

There are two ways of distributing the recorded lectures with Adobe Presenter:

- server-based streaming
- single pdf file distribution

The obvious problem with the server-based streaming is the same as that of the current Collegerama system. It is not possible to distribute the lectures through a standard video-sharing and streaming medium such as YouTube or iTunes. This means that the distribution options are severely limited.

When choosing the single pdf file distribution, all the data that is required to view the lecture, the audio and video stream and the PowerPoint slides, are compacted within one single pdf file. It offers the option of playing it on an offline device that has the Adobe Reader installed. Once downloaded, it is also possible to play the lecture an unlimited amount of times, without having to be connected to the server.

Unfortunately the same distribution problem arises when choosing the pdf option. Currently none of the online streaming servers support the playing of pdf files. This means that for other distribution channels to be available, the lecture needs to be converted back to a single file video format.

3.4 Conclusions

Timeline

There are two approaches to creating recorded lectures:

- video-based
- slide-based

The difference between these two types is the timeline on which the lecture is based. The video-based system is the standard Collegerama method, where a video file of the lecturer exists and several screenshots are linked to the timeline of this video. An example of a slide-based system is Adobe Presenter. Here, the PowerPoint slides pose as a logical timeline for the whole lecture and audio and video streams are linked to each slide.

Navigation

The current navigation system within Collegerama does not work well. It relies on the screenshots of the PowerPoint slides that are displayed during the lecture. The problem is that during the recording of these lectures, a screenshot is taken every 1 to 4 seconds whenever a change on the screen has been detected. When the lecturer inadvertently moves the mouse or plays a video in his presentation, a lot of abundant screenshots are taken and the efficiency of navigation is greatly decreased.

Collegerama as vodcast

It is clear that if Collegerama lectures are going to be distributed through the current popular video-sharing mediums, it is required to convert the lectures to a single video file. This is the standard input that is required and accepted by all platforms. To do this, the two elements of a lecture need to be combined:

- video stream of the lecturer (wmv)
- screenshots of PowerPoint slides (jpg)

A lot of thought has to go into what screen resolution to use, where to place each element within the video stream and how to fill up any extra unused space in the newly created video. The size of the video used is dependent on the medium while sharing it. If a vodcast stream for an iPod or iPhone is being created, the resolution is obviously going to be a lot different compared to a video that is created for a high definition YouTube video.

It is concluded that a video-based system is a lot better for distribution, since virtually all popular online distribution channels do not offer support for pdf files or a server-based infrastructure to share lectures (YouTube, iTunes-U, Academic Earth etc). By creating high definition movies from the original Collegerama recordings, all other video versions with different pixel sizes can be derived (for instance, vodcasts designed to fit on mobile media players such as the iPhone or Blackberry). This HD movie has a smaller file size than the original Collegerama recordings, due to the efficient compression of still pictures (slides as movie). In the example lecture of 45 minutes, the file size is 88 MB instead of 117 MB.

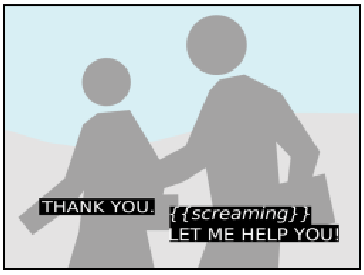
4. Subtitling

Subtitles form the foundation for a lot of extra functionality options, such as tag cloud indexing, searching and translation. In this chapter, the methods for creating subtitles, reasons for wanting to do so and ways of translating subtitles for foreign speaking students is discussed.
There are several reasons why the addition of subtitles for Collegerama lectures is useful:

- lectures are easier to follow
- lectures are available to foreign-speaking students
- lectures can be made searchable

Lectures are easier to follow
If a lecture contains subtitles during playback, it will be possible for the deaf and people with a hearing problem to understand what is being said. These special subtitles for the hearing impaired are called “closed captions” or are sometimes also referred to as “subtitles for the hard of hearing”. The term “closed” in closed captioning indicates that not all viewers see the captions, only those who choose to decode or activate them. This distinguishes from “open captions” (sometimes called “burned-in” or “hardcoded” captions), which are visible to all viewers.

Most countries in the world do not distinguish captions from subtitles. In the United States and Canada, these terms do have different meanings. Subtitles assume the viewer can hear but cannot understand the language or accent, or the speech is not entirely clear, so they only transcribe dialogue and some on-screen text. Captions aim to describe all significant audio content—spoken dialogue and non-speech information such as the identity of speakers and occasionally their manner of speaking—along with music or sound effects using words or symbols.



Lectures are available to foreign-speaking students
Subtitles are generally used to display the spoken words in a video on the screen. For every different language, a new subtitle track has to be created. Most DVD movies that are released in Europe contain at least the subtitle tracks for the languages German, French and English. During production these subtitles are mostly created by hand using professional translators.

An alternative for generating different subtitle tracks is to use an automated computer system. An example of such a service that is publically available is Google Translate. It is a beta service provided by Google Inc. to translate a section of text or a webpage into another language. In December of 2009 the system supports 52 different languages from around the world. Like other automatic translation tools, it has its limitations. While it can help the reader understand the general content of a foreign language text, it does not always deliver accurate translations. Some languages produce better results than others.[37]

Lectures can be made searchable
Every Collegerama lecture consists of a single video stream. Without some sort of indexing system, the only element offered is a 45-minute long video that has no possibility for skipping through relevant parts based on a certain topic.

For further background information about the research on this topic, see Annex D.

4.1 Subtitling process

Subtitles for translation and searching are only composed of spoken text. This is created from the audio track that has been extracted from the video stream. The creation method is shown in Figure 4.1.

Figure 4.1: Creation process for subtitles

There are several ways of creating subtitles:

- manual subtitling
- real-time subtitling
- speech recognition

Manual subtitling

Many different programs can be used to manually create subtitles for a movie, but the overall usage of them is generally the same. You start by typing in the lines of text that are spoken in the movie. Once these are finished, the transcript needs to be matched to the time sequences of the movie. For every line of text, a timestamp is added so that the subtitle generator can later show the appropriate text at the right timeframe.

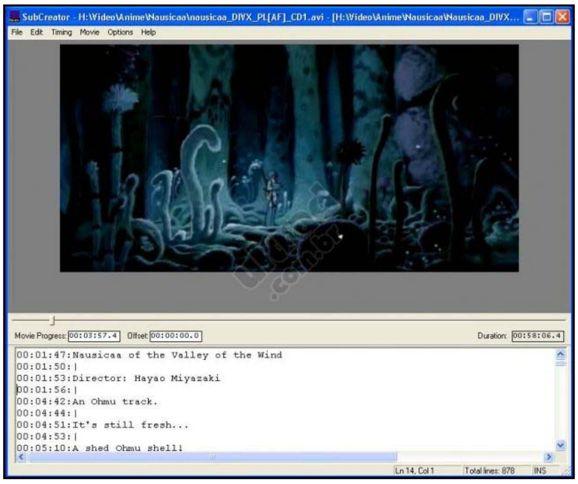


Figure 4.2: Screenshot of the program SubCreator
(Source: http://www.radioactivepages.com/index.php?section=software&docid=subcreator)

The advantage of this method is the easy editing of the subtitles. Everyone who can understand the language that is being spoken can write out the transcripts of a given video stream. Unfortunately, this process is very time consuming and therefore relatively expensive.

Real-time subtitling

Real-time subtitles have to be created within 2 or 3 seconds of the broadcast. There are people specializing in this sort of work, called Communication Access Real-Time Translation stenographers. They use a specialized keyboard that is specifically designed to support shorthand writing, called a stenotype or velotype typewriter.

Real-time stenographers are the most highly skilled in their profession. Stenography is a system of rendering words phonetically, and English, with its multitude of homophones (e.g. there, their, they're), is particularly unsuited for easy transcriptions. They must deliver their transcriptions accurately and immediately.[23]

Speech recognition (ASR)
At the moment, speech recognition technology or Automated Speech Recognition (ASR) is still a long way from achieving fully automatic subtitles for any program. There are still many errors in generating text and several challenges such as background noise, different accents and multiple simultaneous speakers make the process difficult. Speech recognition technologies do have their place in the world of modern subtitling. ASR systems are already used in live subtitling systems for sports, news and politics.

Translated subtitles
Previous described methods for creating subtitles can also be applied to the creation of subtitles in languages other than the spoken language. In general, two ways of creating translated subtitles can be distinguished:
- human translation of the spoken text (either offline or live)
- machine translation from subtitles of the spoken text

Figure 4.3: Translated subtitles improve the learning environment for non-native speaking students

At present, machine translation is not able to produce high quality subtitles. The produced quality is either accepted as an improvement over “no translation” or used as a starting point for human post-processing. Google Translate is a well known example of machine translation, but many other systems are presently available. Machine translation is a booming industry supported by an enormous amount of scientific research programs, executed at nearly every university in the world.

4.2 Subtitles from speech recognition

Automated speech recognition (ASR)
ASR is a sub-field of computational linguistics that investigates the use of computers to transfer spoken words into computer data, ranging from text (speech-to-text) to input control (voice-controlled machines). The fast development of stronger computers has boosted this field in the last decade, sometimes ironically leading to disastrous overrating, such as the Lernout & Hauspie collapse.[57]

Speech recognition systems have been and are being developed by universities as well as by commercial companies. Some major international institutions on ASR:
- LIMSI - Spoken language processing group (France)
- Speech research group at University of Cambridge (UK)
- Raytheon - BBN Technologies (USA)
- SRI - Speech Technology and Research (STAR) Laboratory (USA)

For recent research on ASR, a reference is made to publications of the International Speech Communication Association (ISCA). The most recent conference of the ISCA was held between September 6th and 10th 2009 in Brighton (UK). This 10th yearly conference (Interspeech 2009) included 38 oral sessions, 39 poster sessions and 10 special sessions, resulting in 762 reviewed and accepted papers.[9]

Performance evaluation of speech recognition

Speech recognition engines are developed for a certain language and most often a certain environment, such as telephone conversations, voicemails, news readings, movies etc. The performance of an ASR engine differs not only based on environment, but also on the different speakers (male/female voice, dialect, intonation etc).

The standard evaluation metric used to measure the accuracy of an ASR engine is the Word Error Rate (WER). The word error rate is defined as the ratio of word errors over the total number of words in the correct reference transcript $Nref$. The number of word errors is the sum of the number of deletions D, insertions I and substitutions S:[12]

$$WER = \frac{D + I + S}{Nref} \cdot 100\%$$

Note that the word error rate can be higher than 100%. For example, when the result set contains more words then the reference transcript and all of these words are incorrect. In this case the number of substitutions would be equal to the number of words in the reference text. On top of that there would be insertion errors. For ASR, a WER of 50% is often considered as an adequate baseline for retrieval.[10] Modern ASR engines have a WER between 10% and 60%. Human-made transcripts have a WER between 2% and 4%.[11]

Word accuracy (WA) is defined as the supplement of the word error rate:[12]

$$WA = 100 - WER$$

The word accuracy is not just the fraction of words correctly recognized, because the latter does not include insertions.

Determining the WER value requires a reference transcript. By absence of such a transcript, the quality of ASR can be indicated by the Word Correctness (WC). The WC value is defined by the ratio of the number of correct words Nc over the number of output words $Nout$:[12]

$$WC = \frac{Nc}{Nout} \cdot 100\%$$

Word accuracy and word correctness can be used interchangeably in case the ASR engine does not produce deletions ($D = 0$) and insertions ($I = 0$), or only in a negligible number (less than 5% to 10%). This is often true for modern ASR engines with a good performance. In this case, the ASR output only includes correct and incorrect words and the number of words in the reference transcript is equal to the number of words in the ASR output ($Nout = Nref$), so:

$$WC = WA = 100 - WER$$

Speech recognition for recorded lectures

The 28 recorded lectures of course CT3011 (TU Delft) have been used as input for ASR (see Annex E). These lectures were given in the Dutch language. Speech recognition was done with SHoUT, a speech recognition engine for the Dutch language developed at University of Twente by Marijn Huijbregts, as part of his PhD research.[13] SHoUT is an acronym for the Dutch project name “Spraak Herkennings Onderzoek Universiteit Twente” (in English: Speech recognition research University of Twente).

Table 4.1 gives some data on the lectures and the quality assessment. Figure 4.4 gives the word correctness per lecturer.

Table 4.1: Quality assessment of word correctness by speech recognition on lectures

Item	Range	Average per lecture	Total
Number of recorded lectures			28
Duration of lectures (hh:mm:ss)	23:26 – 53:51	41:20	19:17:33
Number of words in output	3,581 – 9,392	6,748	188,957
Sample size for assessment	4.2% – 11.0%	6.1%	6.1%
Word correctness	23% – 73%	50%	50%

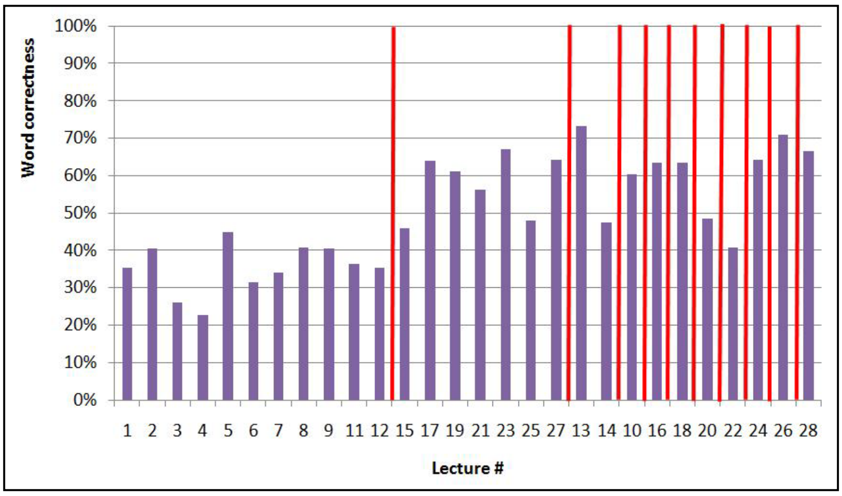


Figure 4.4: Word correctness of SHoUT for the CT3011 lectures, clustered by speaker

The results of the quality assessment of the SHoUT output can be discussed for the following items:

- number of words
- word correctness

Number of words

For one of the 28 recorded lectures, human-made subtitles have been manually created. These subtitles contain 6.970 words. SHoUT produces 7.351 words for the same lecture. This 5% increase is probably due to the rather low speaking-rate in lectures, for which SHoUT divides long words into smaller words.

Word correctness

The average word correctness of SHoUT amounts to 50%, with a variation between 23% and 73%. The word correctness differs significantly for different lecturers. For the word correctness, no correlation was found with either the gender of the lecturer (male or female voice) or the age (lowered voice).

Subtitles from speech recognition output

For one lecture the SHoUT output was used for creating subtitles. This required substantial input to cluster the individual words and sentences into proper subtitles. The result of this conversion is shown in Figure 4.5.

Figure 4.5: Subtitles created from the SHoUT transcript

In the produced subtitles, no word correction was done. Such a correction for real lectures is essential for using SHoUT output for subtitles. In the result set of Figure 4.5, only 7 out of the 48 test sentences (15% of the subtitles sentences) have a word correctness of 100%. Speech recognition engines like SHoUT might be extended with a statistical post-processor which clusters the generated words into subtitle sentences. This post-processor will be fed by a huge collection of subtitle sentences, in a way similar to collections used for machine translation (see paragraph 4.3). Statistical post-processing will not only produce sentences instead of isolated words, but might also increase the word correctness of the ASR engine by analyzing these complete sentences. In this way, statistical post-processing will reduce the efforts required for the production of high quality subtitles.

ASR time-coding of transcript

An alternative approach of using ASR for the creation of subtitles is the time-coding of human-made transcripts. The ASR engine will analyze the entire transcript and try to match the known words to similar words that are picked up. These words are then linked to the proper time-code. The website Radio Oranje[51] shows a demo of this method for a speech broadcast on the radio by Queen Wilhelmina during World War II. The existing transcripts of the broadcast were available and each individual word has been time-coded by SHoUT.

4.3 Machine translation for subtitles

Machine Translation (MT)

Machine translation, sometimes referred to by the abbreviation MT, is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another. At its basic level, MT performs simple substitution of words in one natural language for words in another. Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition and translation of idioms, as well as the isolation of anomalies.

Machine translation can be diverted into two main approaches:

- rule based translation
- statistical translation

Rule-based machine translation relies on countless built-in linguistic rules and millions of bilingual dictionaries. It focuses on translating separate words and afterwards correcting the grammar by using dictionary grammar rules. The translation is predictable, but the translation results may lack the fluency readers expect.

Statistical machine translation utilizes statistical translation models whose parameters come from the analysis of monolingual and bilingual collections of texts. Building statistical translation models is a quick process, but the technology relies heavily on existing multilingual documents. A minimum of 2 million words for a specific domain and even more for general languages are required. Statistical MT provides good quality when large and qualified text data is available. The translation is fluent, meaning it reads well and therefore meets user's expectations. However, the translation is neither predictable nor consistent.[1]

Phrase-based statistical machine translation has emerged as the dominant paradigm in machine translation research.[1] In order to obtain the benefit from both approaches, existing rule-based translation systems are presently extended by statistical post-processing.[2] For further recent research on machine translation, a reference is made to publications of the Association for Computational Linguistics.[50] The ACL is the most prominent international scientific and professional society for people working on problems involving natural language and computation. Regional associations related to the ACL include:

- EACL: The European Chapter of the ACL
- NAACL: The North American Chapter of the ACL

The most recent conference of EACL was held between March 30th and April 3rd 2009 in Athens, Greece. The 12th conference included several special workshops. Many international researches on this subject were presenting their most recent findings at the conference and workshops.[3][4] Research on machine translation in Europe is heavily funded by the European Union. Their research programs on machine translations are:

- EuroMatrix Project (Sept. 2006-Febr. 2009)[52]
  Project motto: Statistical and Hybrid Machine Translation Between All European Languages
- EuroMatrixPlus (March 2009-February 2012)[53]
  Project motto: Bringing Machine Translation for European Languages to the User

A special aspect of machine translation is machine transliteration. Transliteration is the conversion from one writing system to another, with different scripts. Translation from English to Chinese is an example of this. The most recent workshop on this subject (2009 Named Entities Workshop: Shared Task on Transliteration) was held in Singapore on August 7th 2009 as part of Association for Computational Linguistics.[5] Proceeding of this workshop can be found on the Internet.

Many machine translation engines have been developed by universities as well as by commercial companies. These machine translation engines compete in the quality of the produced translation. Table 4.2 gives some examples of these engines.

Table 4.2: Some popular machine translation engines

Product	**Owner**	**Type (*)**	**Start year**	**Languages**
SYSTRAN	SYSTRAN	R	1968	21
Babelfish	Yahoo	R (+S?)	1990	13
Translate	Google	S	2004	53
MOSES	Open source	S	2006	toolkit
Asia Online	Asia Online (MOSES)	S	2006	516
Bing	Microsoft	S	2009	20

Type: R= rule based S=statistical
(Source: http://en.wikipedia.org/wiki/Comparison_of_machine_translation_applications)

Performance of Machine Translation

The performance of statistical machine translation depends strongly on the size and quality of its data (corpus). This performance might differ with the direction of translation. Translation from Dutch to English might differ from the translation of English to Dutch, even though it is produced by the same translation engine. Moreover, the performance will be different in districted domains. Translation of news bulletins might be different to translation of scientific articles, produced by the same translation engine.

Several automatic metric scores have been developed for evaluating machine translation performance, such as Blue, Meteor, TER (Translation Error Rate), HTER (Human-targeted Translation Error Rate) MaxSim, ULC, and many others. However, automatic measures are considered to be an imperfect substitute for human assessment of translation quality.[6] The performance of some English to German machine translation engines is shown in Figure 4.6. These results were obtained from a quality assessment by 160 translators for English and five other languages (German, Spanish, French, Czech and Hungarian).[6] The translators were asked to rank the outcome of 26 MT engines on 38,000 sentences (1,500-7,000 per language pair). They were also asked to edit about 9,000 isolated sentences, coming from the MT engines, into fluent and correct sentences without looking at the original source. This should reflect the people's understanding of the output. The edited output was used in the evaluation, even in instances where the translators were unable to improve the output because it was too incomprehensible. The edited output was given a value for the percentage of the time that each MT system was judged to produce an acceptable translation. This value can be considered as a value for "understandability", not as a real measurable value, but as a relative figure for comparison of different systems. The reference system is an online human-made translation. Around 20% to 50% of the time, adequate edited translations were obtained with machine translation.

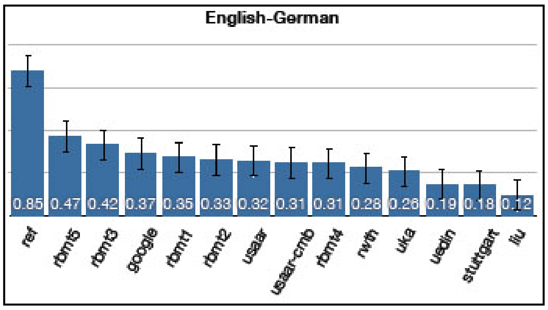


Figure 4.6: The performance of some English to German translation engines compared to human translation (=ref) [6]

Assessments showed that languages for which large and reliable language pairs are available are better translated.[5]

Differences in evaluation of ASR and MT

Under the present state of development the values for word accuracy of ASR engines are in the same order as the values for understandability of MT engines. However, the first is regarded as "far from sufficient for subtitling", while the latter is often considered as "adequate for subtitling". This phenomenon can be explained by the big difference in awareness of the viewer. Hearing the speaker while watching the ASR subtitles, put a lot more emphasis on differences between the two. Most of these differences are noticed by the viewer and can be seen as a serious shortcoming. Bad subtitles will also result in bad translated subtitles through the use of MT.

ASR can be used for searching, since mistakes in sentences won't be visible and aren't a big problem for the search functionality. A word accuracy of 50% is considered as suitable and is obviously a lot better than having no reference data for search. For MT, the criteria aren't as demanding, since the reference situation is that of a student who is trying to find a spoken word in a bilingual dictionary to his native language, while the lecturer is continuing with his lecture.

Google Translate in YouTube

If there is at least one subtitle track available, YouTube provides a translation service that can automatically convert the subtitles to another language. This is done through the Google Translate service mentioned. On the bottom-right of the YouTube interface, a button with the CC logo (the official logo which stands for Closed Captions) is available to turn the subtitles on or off. It also opens a submenu from which you can access the translation menu. When the translation menu has been opened, the user can choose from 52 different languages that are available under the dropdown menu. Once a language has been chosen, the subtitles will be automatically sent to the Google Translate engine and YouTube will display the results.

Figure 4.7: Translated subtitles from Dutch to English in YouTube

Google Translate coverage has been expanded dramatically. It now supports the translation between any of the following languages: English, Arabic, Bulgarian, Chinese, Croatian, Czech, Danish, Dutch, Finnish, French, German, Greek, Hindi, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Romanian, Russian, Spanish and Swedish. Google Translate now supports 56 language pairs and has become the most comprehensive online translation tool available for free. In November 2009, YouTube announced that they will expand their services for translating subtitles. Users will be able to post-process the subtitles generated by Google Translate. This service also includes the use of Google's ASR system for generating time-tagged subtitles for YouTube-Edu channels (initially only available in English). As part of this service it will be possible to upload transcripts which will be time-tagged by Google's ASR system.

Human post-processing

The present automatic translation by Google Translate results in translated subtitles which are readable for 20% to 80% of the time. It is expected that this quality will improve significantly over the next few years. This quality improvement is obtained with the help of larger and better data sets.

For recorded lectures at TU Delft, the following translation pairs are most significant:

- Dutch to English (BSc courses for non-Dutch-speaking MSc students)
- English to Dutch (MSc courses for Dutch professionals, as life-long learning material)
- English to any other language (MSc courses for non-native English speaking students)

For these target areas the present quality of machine translation might be considered to be insufficient. Manual post-processing might be used in these cases for improving the quality of the machine translation output.

4.4 Text-to-speech for translated subtitles

Having proper translated subtitles opens the door for spoken subtitles in the native language of the student. Dubbing of lectures is possible by using text-to-speech engines. In the chain from spoken words to speech recognition (speech to text) to machine translation (text to translated text) to spoken translated words (text to speech), this part has been most developed. IBM's ViaVoice Text-To-Speech is an example of such a service, which is available online. It should be noted that "Real-Time Translation Service" will be a major research goal for the near future. Another example is MeGlobe, which is an instant messaging service with real-time translation to and from over 15 languages (see Figure 4.8). For educating foreign speaking students, such developments will be a serious boost. This futuristic development is not further elaborated within the scope of this research project.

Figure 4.8: Will automatic real-time translation engines become available within the next decade? (Source: http://www.meglobe.com)

4.5 Conclusions

It is concluded that producing subtitles for a video lecture opens up a lot of new possibilities. Having the option of turning subtitles on in the same language as the spoken text could make lectures easier to follow for certain students. For Dutch students who follow an English master course, it adds to their learning experience if those lectures are subtitled in Dutch. Subtitles can also be useful as a basis for searching of lecture content.

The present state of development in speech recognition for producing subtitles, and machine translation for producing translated subtitles, has been investigated in this research project. The current speech recognition technology has also been evaluated for the generation of proper subtitles. For this, the speech engine created at University of Twente, called SHoUT, has been used.

With this ASR engine a word correctness of 25% to 75% was observed for the 28 Dutch spoken lectures that were tested. It is concluded that this system is not yet sufficient to generate proper subtitles and manual post-processing is always required. Machine translation allows for a decent translation, which is always better than having no translation at all. Using it professionally in the education program still requires substantial post-processing.

A problem that most universities currently have is that certain master courses have a prerequisite bachelor course that is given in Dutch. Foreign speaking students who are only going to do a master need to know the subject matter of these courses, but aren't able to look back through those lectures. With subtitles and MT technology, it becomes possible for them to at least follow part of the lecture (dependant on the quality of translation).

5. Navigation and searching

Presently, Collegerama does not provide any form of search functionality. The Collegerama catalog shows an overview of all recorded lectures in a course in a crude form. An example of this catalog is shown in Figure 5.1.

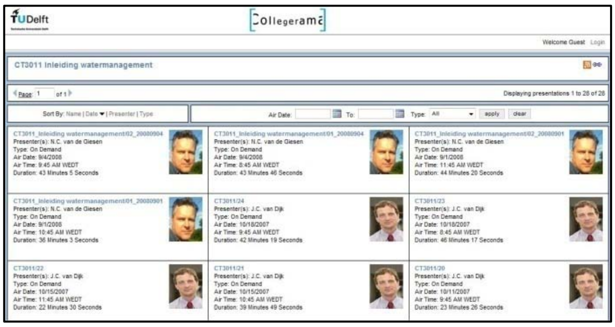


Figure 5.1: Catalog of recorded lectures in a course
(Source: http://collegerama.tudelft.nl/mediasite/Catalog/?cid=16b5f5fa-0745-4b8b-9f02-f79a03abf50a)

The lecture titles and the name of the lecturer are usually wrong. The only correct metadata of a lecture are the recording date and time (announced as air date and time) and the duration of the recording. Searching for a particular lecture in Collegerama can only be done by sorting on this improper metadata. This form of searching seems far from sufficient. Due to this inadequate metadata, the lecturer usually creates an URL-link of a particular lecture recording within Blackboard, the digital learning environment. In Blackboard the lecturers have full access to the published course material.

Within a lecture, the only navigation and/or search facility of Collegerama is the overview of slides. Using this thumbnail table during playback hides the view of the current slide. The main drawback of the Collegerama navigator is the disturbance caused by screen actions either by mouse movements or by screen actions, due to a PowerPoint animation or by writing on an electronic blackboard. This enormous amount of screenshots makes this slide-based system completely unsuitable for navigation (see Figure 2.6).

This description clearly shows the need for improvement of navigation and searching facilities in the Collegerama environment. In this chapter the possibilities for searching in and browsing through recorded lectures in a course will be presented. Initially, navigation in movies and DVD's is presented, as well as the scientific research on multimedia retrieval systems.

In the next paragraph, the following sources of information are presented:
- lecturer (lecture titles, lecture chapters)
- slides (slide titles, slide content, slide notes)
- spoken words (transcripts, subtitles and/or speech recognition output)

Afterwards, the different products are presented:
- search engine on lecture data in a course
- tables of content (for courses and lectures)
- tag cloud presentations of lecture content

Finally the results will be evaluated in order to determine a proposal for searching facilities, i.e. required sources and proposed output. For further background information about the research on this topic, see Annex E and F.

5.1 Meta-data for navigation and search

Elaborating the improvement of navigation and search within lectures recorded by Collegerama might be preceded by investigating these aspects in parallel environments or disciplines. For navigation of videos, the navigation within DVD and Blu-Ray movies can be evaluated. These movies are considered to be the most commonly accepted development in user accessibility. For search, the latest developments in multimedia retrieval have been studied.

Selecting of and navigation in DVD movies

The selection and navigation process for (recorded) lectures could be compared to the selecting (buying/hiring) and viewing of DVD movies. The movie box sets containing movies from a TV series can be considered as comparable to courses containing recorded lectures. To make a proper selection, the potential viewer requires further information on the actual content of the movie box set and its movies. This metadata is normally printed on the movie box set and on the cover of the individual movies. With this concept in mind, the primary metadata of courses, lectures and lecture content is presented in Table 5.1.

Table 5.1: Primary metadata for selecting of and navigating in recorded lectures

Course	Lecture	Lecture content
University	Lecture title	Table of contents
Course name	Name of lecturer	Tag cloud of content
Responsible teacher	Course name (and year)	Screenshots (picture story)
Course code	Date of recording	Short description
(Academic) Year	Initial slide (picture)	
Academic discipline	Tag cloud of content	
Faculty	Screenshots (picture story)	
Logo	Short description	

Not all metadata is text. Screenshots, logos and tag clouds are pictures which give a better impression on the movie box (course) and its movies (lectures) than text in titles and descriptions. For navigation within a movie itself, Table 5.2 gives the analogy for recorded lectures.

Table 5.2: Analogy of navigation in DVDs and recorded lectures

Element	DVDs	Recorded lectures
Main menu	Chapters	Chapters
Submenu per chapter	Scenes	Slides

Search in movies

Searching in movies is studied in the research discipline of computational multimedia information retrieval.[14] Such video information retrieval focuses on searching in video collections by using various methods of abstracting information from video recordings. The abstraction of spoken text (speech recognition) for data retrieval or the detection of shot changes for segmenting can be mentioned as examples of these methods. Figure 5.1 gives an overview of a multimedia information retrieval system, as described in the book Multimedia Retrieval.[14]

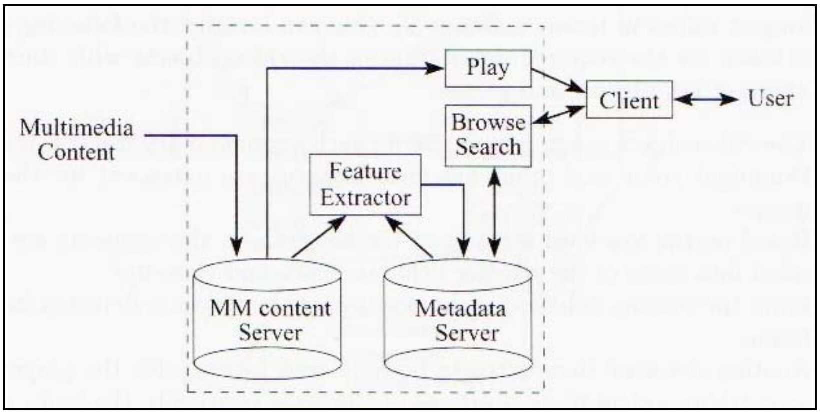


Figure 5.2: Schematic view of a multimedia information retrieval system [14]

Specific elements of multimedia information retrieval with relevance for recorded lectures are:

- languages for metadata[15]
- presentation of search results[16]
- evaluation of Multimedia Retrieval Systems[17]

An important element in searching within multimedia data is the relation between the video content and the metadata. For recorded lectures, this relation can be fixed by using time-tagging. With time-tagging, the metadata is related to a certain time interval in the multimedia content. Subtitles with a particular begin and end time is a typical example of this. Other items such as slide views (pictures/scenes) and chapters can be time-tagged. Figure 5.3 gives an impression of searching in multiple parallel metadata of recorded lectures.[18]

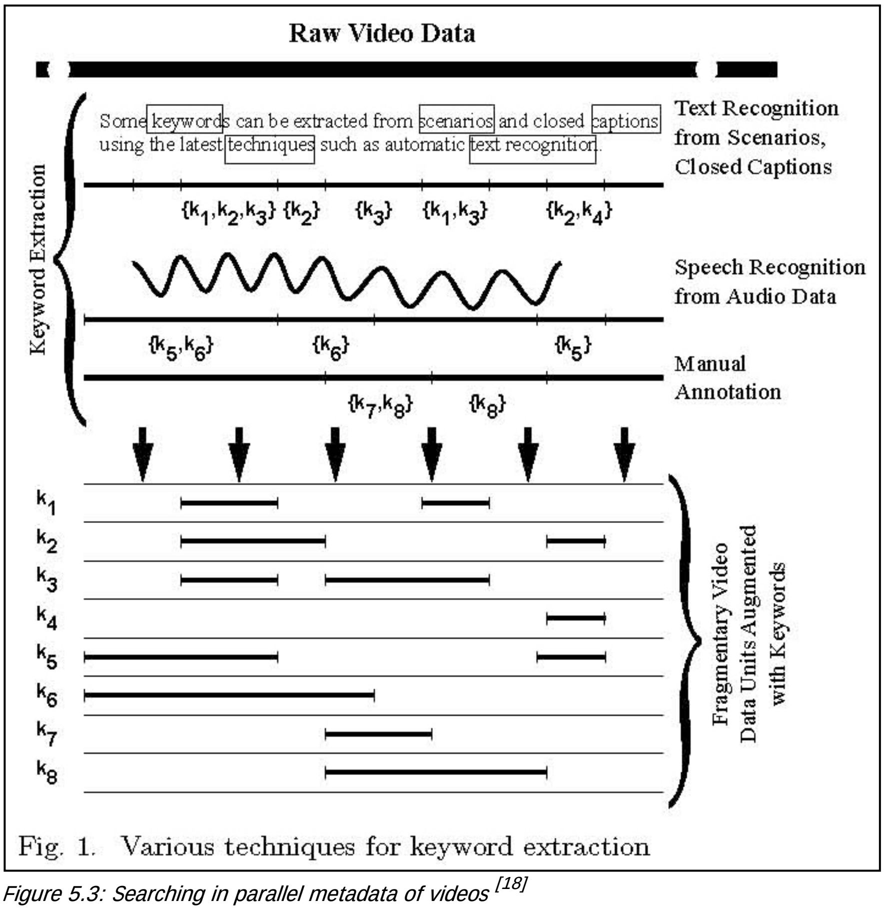


Figure 5.3: Searching in parallel metadata of videos [18]

Searching in a multimedia system will give a result set. The user will be confronted with this result set in order to further select one or more of the results for actual viewing. For this, selection it might be essential that the user is able to see the context of the result element. As a user looks for the keyword "water", the result set will show multiple occurrences of this word. Information about the context of the search result or the source type that the data came from might be relevant for evaluating the search results. This constraint requires context-preserving information retrieval.[19]

5.2 Metadata sources

Input from lecturer

The Collegerama recording system is based on input from a video camera and input from screenshots at the display-computer. These screenshots should be regarded as a low level of screen recording with a maximum frame rate of 1 fps. Thumbnails of screenshots are used for navigation in Collegerama/Mediasite. For this the individual screenshots can be clustered in a group showing only one thumbnail in the navigation screen. This clustering is done automatically during recording. At TU Delft, this results in the generation of far too many thumbnails. Further clustering can be done in a manual post-processing session, but this is currently never done. The lecture recording department is understaffed to handle this task and the lecturer does not have access to the Collegerama server. The ultimate result is that the recorded lectures often lack a proper navigation.

The lecturer should get access to the Collegerama server so that the overhead of screenshots in the recorded lectures can be corrected. As an alternative approach, the recording department may develop an offline tool (or web based data collection system) in which this clustering can be done. Such a system could be used for collecting all data from the lecturer, such as:

- proper lecture title
- accurate name of the lecturer or lecturers
- time based chapter titles of a lecture
- time based correlation between recording and original PowerPoint slides
- original PowerPoint presentation (either as ppt or as pdf file)

The main purpose of the data collected from the lecturer is to create a proper table of contents for the recorded lecture. To accomplish this, the lecture should be divided into 3 to 8 "chapters" for a 45 minute lecture. This provides each lecture with chapter durations of approximately 5 to 15 minutes. The lecturer should at least create a "text slide" per chapter in case the lecturer does not use a PowerPoint presentation or equal presentation tools (such as electronic blackboards etc). This text slide is used as an equivalent to a presentation slide and is shown during the playback of the whole chapter. The collected data can be incorporated into a database per course (Collegerama data system) and might also be used to improve the original Collegerama recording/navigation. This database can also be used to generate a proper table of contents (TOC), containing all recorded lectures of a course. This might replace the original Collegerama catalog. The word correctness of text information from data collected from the lecturer is estimated at 90% to 100%. The text itself has completely been recovered from the PowerPoint slides, but the lecturer might have made mistakes while creating them.

Input from slides

Text on PowerPoint slides form a rich source of data for recorded lectures. The text data from slides can be divided into:

- slide titles
- slide content
- slide notes

The text data of PowerPoint slides can automatically be retrieved from a digital file, either from the ppt/pptx file and/or the “printed” pdf file. This data can then be inserted into the Collegerama data system. The slide titles form a table of contents (TOC) of the lecture based on the timing input of the lecturer. Every word in the text itself is automatically retrieved; however the text that is shown in pictures requires a different technique (Optical Character Recognition). In this research project, OCR has not been used to accomplish this.

Input from spoken words

The spoken text in recorded lectures might be available in one of the following forms:

- transcript (full text, without timestamp)
- subtitles (time-stamped per sentence)
- words (time-stamped per word)

For the sample course CT3011, the following sources are available:

- subtitles and transcript of the sample lecture #15 of this course (transcript generated from human-made subtitles)
- words of all lectures retrieved through speech recognition (SHoUT)

For the speech recognition by SHoUT, the word correctness of all recorded lectures has been determined (see Annex E). The mean word correctness is 50%, with values between 23% and 73% (standard variation 14.6%).

5.3 Metadata storage

All the collected metadata can be incorporated into a Collegerama data system. For this research project this database is restricted to only the recorded course. The database shown consists of 2 tables:

- lectures, containing all metadata related to the lecture as a whole
- content, containing all metadata within the course, on a time-based level (start and end time in milliseconds)

A visual representation of each table, its columns and their corresponding data type is given in Table 5.3 and Table 5.4.

Table 5.3: Database table Content

Field name	Data type
Content_id	int
Lecture_id	int
Start_time	int
End_time	int
Text_type	int
Text	nvarchar(MAX)

Table 5.4: Database table Lectures

Field name	Data type
Lecture_id	int
Lecture_nr	int
Title	nvarchar(100)
Lecturer	nvarchar(50)
Air_date	datetime
Collegerama_id	nvarchar(50)

For this project, only text data has been included into these tables. A future addition could be the adding of thumbnails per record, so that a characteristic screenshot preserves the context of information. This screenshot might be taken at a certain time moment Tss in the time interval ($Tbegin$ to $Tend$) at a fixed elapsed time interval:

$$Tss = Tbegin + aRef \cdot (Tend - Tbegin)$$

In the latest YouTube movies, the typical screenshot for a movie shown at selection is taken at 33% of the length ($aRef = 0.33$). The screenshot might be replaced by storing the value for $aRef$ in the metadata tables, in case the movie and metadata are stored in a multimedia retrieval system. The value of $aRef$ per record could be flexible giving additional selection freedom to the lecturer.

Figure 5.4 and Table 5.5 give an impression of the data collected in the Collegerama data system.

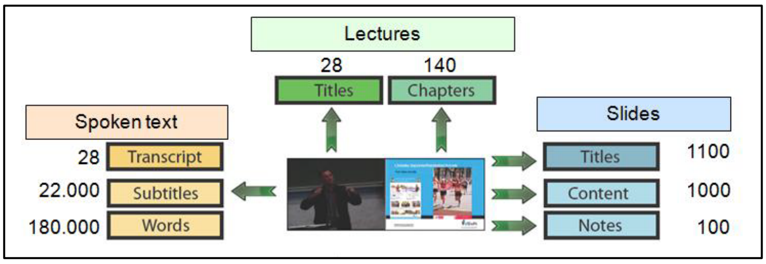


Figure 5.4: Source and number of records in the Collegerama data system for the course CT3011 (assuming subtitled for all lectures)

Table 5.5: List of Text_types and the amount of records and words in the database for course CT3011

ID	Text type	Nr of records	Nr of words	Nr of characters
1	Lecture title	28	129	917
2	Lecture chapter	116	300	2,526
3	Slide title	1,183	3,900	28,741
4	Slide content	1,042	15,943	129,195
5	Slide notes	280	22,512	142,856
6	Transcript (lecture)	28	* 179,480	* 768,058
7	Transcript (slide)	1,183	* 179,480	* 768,058
8	Transcript (sentence)	21,812	* 179,480	* 768,058
9	Transcript (word)	118,926	188,926	808,482

* 95% of the total number of words generated by SHoUT, based on the comparison between the human-made subtitles and the SHoUT subtitles

5.4 Course and lecture navigation

Tables of content

The Collegerama data system can be used as a generator for a table of contents (TOC) for:

- an overview of recorded lectures in a course
- an overview of content in a recorded lecture

For this research project, two prototypes of TOC's have been evaluated:

- a static TOC (list)
- an interactive TOC (based on Macromedia Flash technology)

Static table of contents
Figure 5.5 gives an impression of a static TOC generated from the Collegerama data system.

#	Lecture title	Lecturer	Length	
1	Water in de wereld (1 van 2)	van de Giesen	0:36:03	
2	Water in de wereld (2 van 2)	van de Giesen	0:44:20	
3	Water in de bodem (1 van 2)	van de Giesen	0:43:46	
4	Water in de bodem (2 van 2)	van de Giesen	0:43:05	
5	Polders (1 van 2)	van de Giesen	0:45:56	
6	Polders (2 van 2)	van de Giesen	0:39:19	
7	Irrigatie (1 van 2)	van de Giesen	0:39:19	
8	Irrigatie (2 van 2)	van de Giesen	0:48:19	
9	Kunstwerken en operationeel waterbeheer (1 van 2)	van de Giesen	0:44:49	
10	Kunstwerken en operationeel waterbeheer (2 van 2)	van Overloop	0:10:28	
11	Reservoirs en review (1 van 2)	van de Giesen	0:36:50	
12	Reservoirs en review (2 van 2)	van de Giesen	0:38:35	
13	Hydrologie - Het hart van Water Resources Management	Savenije	0:40:50	
14	Hydrologie van extreme gebeurtenissen	Savenije	0:33:43	
15	Civiele gezondheidstechniek	van Dijk	0:45:09	
16	(Kleinschalige) drinkwaterzuivering in ontwikkelingslanden	van Halem	0:32:25	
17	Waterkwaliteit 1: Eisen/microbiologie	van Dijk	0:46:28	

Figure 5.5: Table of contents for recorded lectures in course CT3011, generated from the Collegerama data system

The generated TOC lists all lecture titles, the lecturer and the duration. The TOC also contains a hyperlink to the related Collegerama recording. This generated TOC is an improvement over the TOC generated by the lecturer, created as an improvement over the Collegerama catalog, for the following reasons:

- a uniform layout for the whole university
- possibility for automatically updating after modification of the content within the Collegerama data system

Interactive table of contents
Figure 5.6 gives an impression of an interactive TOC generated from the Collegerama data system. In this example a Flash movie is generated containing:

- time slider for all chapters, including chapter titles
- time slider for all slides, including slide titles
- screenshots of HD movie ($aRef = 0.1$)

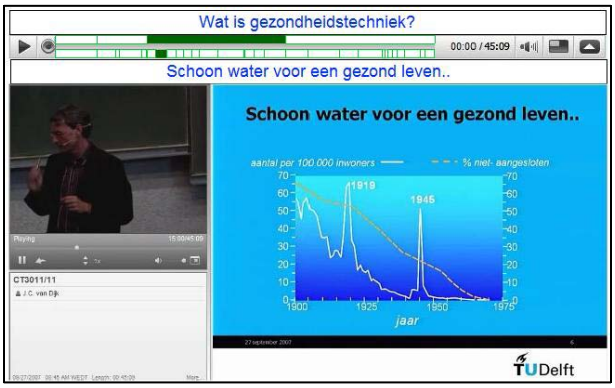


Figure 5.6: Interactive TOC for recorded lecture #15 in course CT3011, generated from the Collegerama data system

The generated TOC shows the screenshot of the HD movie whenever the users' mouse goes over the related time slider section. This is synchronized with the related chapter. The chapter slides does the opposite, showing the first slide in the chapter. This TOC gives a proper viewing of the content of the lecture and is a great improvement over the Collegerama thumbnail navigation.

A similar interactive TOC can be generated for each course. This interactive TOC might show additional metadata such as:

- lecture name
- date and time of recording (air date/time)
- short description of the lecture
- tag cloud of the lecture

The Flash technology allows for relatively large amounts of text information. Flash movies contain vector based text which keeps it sharp at all magnifications (for example at full screen display).

Tag clouds

A tag cloud is a selection of tags or a list of relevant words from a document, in which the size of each tag is based on its frequency of occurrence. The Collegerama data system can be used as a generator for Tag clouds for a certain lecture. These are considered to be a useful representation of the content of a lecture. Annex F evaluates different forms of tag clouds. A basic relationship between frequency and word size is:

$$S = \frac{C - Cmin}{Cmax - Cmin} * Rat + Bas$$

In which:

S = font size of word (pixels)
C = frequency count for the word (or tag)
Cmin = frequency count for the least popular word (or tag)
$Cmax$ = frequency count for the most popular word (or tag)
Rat = largest font size minus smallest font size for words (pixels)
Bas = smallest font size for words (pixels)

In practice, more sophisticated relations are also applied, such as logarithmic or different non-linear relations as well as all kinds of clustering algorithms. Tag clouds have been studied on various other aspects, such as order of words, layout of words, color usage etc.[22] Tag clouds are often produced using specialized websites, such as MakeCloud, Wordle or ToCloud.[54][55][56] The tag clouds for this research project have been produced via the website Wordle. Figure 5.7 gives an impression of such a web-generated tag cloud.

Figure 5.7: Tag cloud for recorded lecture #15 in course CT3011, generated by Wordle, with and without deleted words by prof J.C. van Dijk [54]

Evaluation of tag clouds

In this research project, tag clouds have been produced from different data sources or text types (subtitles, ASR output, slide titles, slide content). These tag clouds were evaluated in order to determine rules for creating the best tag clouds that could best represent the content of a lecture. All tag clouds have been produced in black and white with the same font face, in order to have only the font size as a distinctive element. In most cases the words for the tag clouds have initially been “cleaned” by removing “common Dutch words” (common according to wordle.net) or by selecting only the nouns. These tag clouds and the assessment experiments are reported in more detail in Annex E and Annex F.

This assessment was done in 2 steps:

- quality assessment of 10 tag clouds with 15 to 100 words from different sources
- quality assessment of 10 uniform tag clouds with 15 words from the same sources

The second step was based on the remarks by the lecturer on the first step:

- tag clouds with 100 words are always unacceptable since these are unreadable
- tag clouds with 25 to 35 words contain too many irrelevant words

In the second step, the lecturer was asked to assign a sequential ranking of the 10 tag clouds (actually 9, as #6 was identical to #7) and to mark irrelevant words in each tag cloud for deletion, in order to obtain a better representation for the lecture. In Table 5.6, the results of this second assessment are shown. It contains two rankings, the first is the ranking as given by the lecturer, the second is this ranking combined with a ranking based on the number of deleted irrelevant words.

Table 5.6: Tag cloud assessment of modified tag clouds (all 15 words)

ID	Source	Cleaning method *	Lecturer assessment results			
			General appearance	**Rank**	**Nr of deleted words/rank**	**Total rank**
1	Slide titles	1	Many same sized (small) words	5	4 (=1)	6
2	Slide content	1	Too many same sized (small) words	7	7 (=6)	13
3	Slide titles and slide content	1	Too many same sized (small) words	8	5 (=2)	10
4	Slide notes	1		6	9 (=7)	13
5	Human subtitles	-		9	15 (=9)	18
6/7	Human subtitles	1		4	11 (=8)	12
8	Human subtitles	2		1	5 (=2)	3
9	Human subtitles	2		3	5 (=2)	5
10	SHoUT output	2	Word “chloor” is missing	2	6 (=5)	7

* 1 = after removing common Dutch words; 2 = nouns only

Table 5.6 shows that the two tag clouds from nouns in the subtitles (#8 and #9) have the best overall ranking. These two tag clouds contain the same words, but differ in letter font and layout of the words. The best readable font (Coolvetica) was preferred by the lecturer over a less readable font (Vigo). The lowest number of deleted words was obtained from the slide titles. However, the produced tag cloud contains a low variance in font size, so there isn’t a large distinction in occurrence. The variance in word count in subtitles is much larger giving a more pronounced picture. The tag cloud from the SHoUT output has a lower ranking than the human subtitles, because it misses an important word (“chloor”) and has more deleted words. The other produced tag clouds were significantly less appreciated.

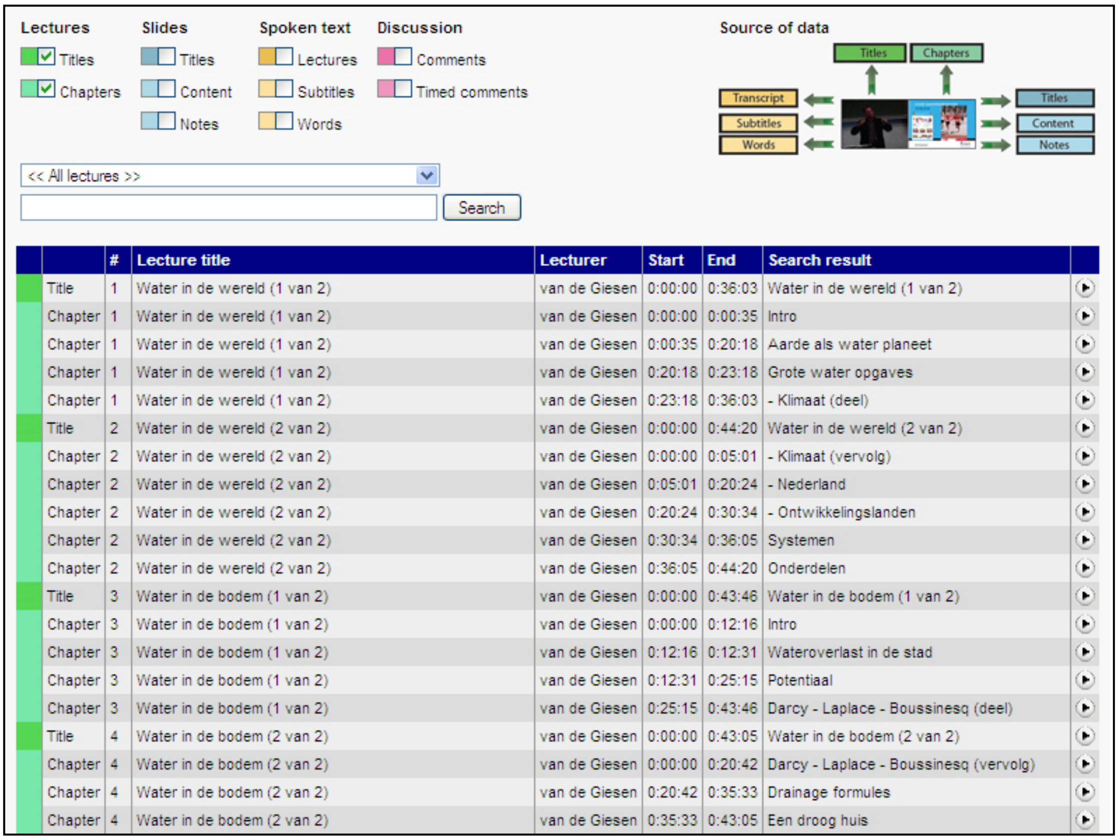
The following conclusions have been made from these results:

- tag clouds should contain less than 15 words
- tag clouds should be obtained from “nouns only”
- tag clouds from subtitles (or speech recognition) are preferred over tag clouds from slide titles (or slide content / slide notes), because of their larger variance in font size
- tag clouds need a “best readable font”
- tag clouds could be improved by removing bad words chosen by the lecturer

The use of colored tag clouds is not evaluated, since this might be largely dependent on the personal preference of a lecturer.

5.5 Collegerama lecture search

The collected data is the source for the Collegerama lecture search engine. Figure 5.8 gives an impression of this.



	#	Lecture title	Lecturer	Start	End	Search result
Title	1	Water in de wereld (1 van 2)	van de Giesen	0:00:00	0:36:03	Water in de wereld (1 van 2)
Chapter	1	Water in de wereld (1 van 2)	van de Giesen	0:00:00	0:00:35	Intro
Chapter	1	Water in de wereld (1 van 2)	van de Giesen	0:00:35	0:20:18	Aarde als water planeet
Chapter	1	Water in de wereld (1 van 2)	van de Giesen	0:20:18	0:23:18	Grote water opgaves
Chapter	1	Water in de wereld (1 van 2)	van de Giesen	0:23:18	0:36:03	- Klimaat (deel)
Title	2	Water in de wereld (2 van 2)	van de Giesen	0:00:00	0:44:20	Water in de wereld (2 van 2)
Chapter	2	Water in de wereld (2 van 2)	van de Giesen	0:00:00	0:05:01	- Klimaat (vervolg)
Chapter	2	Water in de wereld (2 van 2)	van de Giesen	0:05:01	0:20:24	- Nederland
Chapter	2	Water in de wereld (2 van 2)	van de Giesen	0:20:24	0:30:34	- Ontwikkelingslanden
Chapter	2	Water in de wereld (2 van 2)	van de Giesen	0:30:34	0:36:05	Systemen
Chapter	2	Water in de wereld (2 van 2)	van de Giesen	0:36:05	0:44:20	Onderdelen
Title	3	Water in de bodem (1 van 2)	van de Giesen	0:00:00	0:43:46	Water in de bodem (1 van 2)
Chapter	3	Water in de bodem (1 van 2)	van de Giesen	0:00:00	0:12:16	Intro
Chapter	3	Water in de bodem (1 van 2)	van de Giesen	0:12:16	0:12:31	Wateroverlast in de stad
Chapter	3	Water in de bodem (1 van 2)	van de Giesen	0:12:31	0:25:15	Potentiaal
Chapter	3	Water in de bodem (1 van 2)	van de Giesen	0:25:15	0:43:46	Darcy - Laplace - Boussinesq (deel)
Title	4	Water in de bodem (2 van 2)	van de Giesen	0:00:00	0:43:05	Water in de bodem (2 van 2)
Chapter	4	Water in de bodem (2 van 2)	van de Giesen	0:00:00	0:20:42	Darcy - Laplace - Boussinesq (vervolg)
Chapter	4	Water in de bodem (2 van 2)	van de Giesen	0:20:42	0:35:33	Drainage formules
Chapter	4	Water in de bodem (2 van 2)	van de Giesen	0:35:33	0:43:05	Een droog huis

Figure 5.8: Collegerama search engine produced for the course CT3011

The produced search engine allows for selecting each individual data source. The user might search for a certain word or word combination in the selected sources. Along with this, the user might also search over all lectures or within a particular lecture. It is also possible to look through all the available content leaving the keyword empty, which results in:

- a table of contents (TOC) of the course (by selecting only the lecture titles)
- a table of contents (TOC) over the lecture (by selecting only the slide titles in a particular course)

The output of the Collegerama Lecture Search, shown in Figure 5.8, presents the following context-preserving data:

- data source (subtitles, slide titles, etc)
- lecture number (ID)
- lecture title
- lecturer
- time interval (begin, end)
- queried keyword, with 30 preceding and 30 subsequent letters

Evaluation of search engine

The performance of search engines on recorded lectures is studied in the research discipline of Spoken Document Retrieval (SDR).[20] SDR involves the retrieval of excerpts from recordings of speech using a combination of automatic speech recognition and information retrieval techniques. Movies and videos form a sub domain of Spoken Documents. Special workshops on evaluation of information retrieval systems for movies and videos have been organized under the name TRECVid (Text Retrieval Conferences on Videos).[21]

For this research project, an analysis on the results of certain important keywords has been evaluated. The following tests have been done:
- comparing query results from ASR output versus human-made subtitles
- comparing query results from all data sources
- analyzing the video length of search results, based on different data sources in Collegerama lecture search
- "precision and recall" measurement[17]
- analyzing multiple keyword queries

Comparing query results from ASR versus human-made subtitles

The query results of the 15 most-used nouns on both data sources are presented in Table 5.7. The data has been abstracted from lecture #15. In determining the query results of the word "water", compounds such as "drinkwater", "drinkwatervoorziening", "grondwater", "oppervlaktewater" have not been included (as is the case for "drinkwater" in "drinkwatervoorziening"). This table also shows the 5 deleted words that are marked by the lecturer as less relevant in the assessment of tag clouds (see chapter 5.4), leaving the ten most important words (marked by "ok" in the table) as selected by the lecturer.

Table 5.7: Occurrences of the 15 most used nouns from ASR versus human-made subtitles

Keyword	**Lecturer check**	**ASR**	**Human-made subtitles (ref)**	**WA for single word**
		(occurrences)	(occurrences)	(%)
chloor	ok	0	10	0%
drinkwatervoorziening	ok	4	16	25%
boek	ok	5	15	33%
oppervlaktewater	ok	7	15	47%
plaatje	deleted	6	11	55%
vragen	ok	6	10	60%
soort	deleted	7	9	78%
water	ok	33	39	85%
stoffen	ok	8	9	89%
grondwater	ok	20	21	95%
Nederland	ok	35	36	97%
dingen	deleted	17	16	106%
keer	deleted	16	13	123%
drinkwater	ok	16	13	123%
jaar	deleted	28	16	175%
Total		**208**	**249**	**84%**
Total ok words		**134**	**184**	**73%**

The most remarkable result in the occurrences is the word "chloor", which has been indicated by the lecturer as one of the ten most important words. This word has not been recognized by SHoUT as being an uncommon word in the Dutch language. This word or item is therefore not retrieved from the lecture if no correct subtitles are available.

The word accuracy (WA) for "jaar", "keer" and "drinkwater", shows that for searching composed words in the ASR output, it is better to search for word components instead of full words. This is illustrated by the low WA of the word "drinkwatervoorziening". A WA of above 50% is expected from the ASR output as the accepted or expected quality level for ASR engines. The word "boek" has a lower WA in the ASR output, which shows that for SHoUT, this word is difficult to decode. This word has also been indicated by the lecturer as one of the ten most important words.

Comparing query results from different data sources

In order to evaluate the different data sources, the search results of the ten most important keywords of lecture #15 (as determined by the lecturer) have been compared. The results are shown in Table 5.8.

Table 5.8: Occurrences of the 10 most important keywords (as determined by lecturer) from different data sources

Keyword (lecturer)	**Sub-titles**	**ASR**	**Slide titles**	**Slide cont.**	**Slide t+c**	**Slide notes**	**Lecture title**	**Lecture chapter**
water	39	33	5	2	7	0	0	0
Nederland	36	35	0	0	0	3	1	0
grondwater	21	20	0	1	1	1	0	0
drinkwatervoorziening	16	4	0	0	0	2	1	0
boek	15	5	0	1	1	0	0	0
oppervlaktewater	15	7	0	1	1	2	0	0
drinkwater	13	16	1	0	1	2	0	0
chloor	10	0	0	0	0	0	0	0
vragen	10	6	0	0	0	0	0	0
stoffen	9	8	0	0	0	0	0	0
Total occurrences	**184**	**134**	**6**	**5**	**11**	**10**	**2**	**0**
Nr retrieved keywords	**10**	**9**	**2**	**4**	**5**	**5**	**2**	**0**
% retrieved keywords	**100%**	**90%**	**20%**	**40%**	**50%**	**50%**	**20%**	**0%**

The results of Table 5.8 show that for searching in lectures, the lecture chapter titles are of no importance, since none of the important keywords are retrieved. The slide titles and the lecture titles only retrieve 20% of the keywords. These three text types are particularly suitable for navigation, but clearly not for searching. To a lesser extent, the same holds true for slide content and slide notes, which retrieve 40% to 50% of the keywords.

The overall ASR word correctness of this lecture is 46%, as shown in Annex E. The word correctness over the keywords is 73% (= 134 / 184). When comparing the keywords themselves, 90% of them are retrieved by ASR. These results show that ASR gives a drastic increase in search results over slide data. Having human-made subtitles will further increase the search results to an assumed 100% value. The results of Table 5.8 can partly be explained by the fact that transcripts, either from ASR, subtitles or other, contain around ten times more words than the slide content. This is shown in Table 5.5.

Video length per data source

The query results indicate how many of the items are found in a search, but not how long the accompanying video length is for each item. Searching an item in (non time-tagged) transcripts may indicate the lecture in which the item is used, but the user has to watch/listen to the whole lecture to actually come across the correct video segment. Assuming a constant speaking rate might give a best guess to jump to the equivalent time-frame, but in most cases this is not suitable for the user. The time correctness of a search is related to the video length or duration (end time minus start time) of the related video fragment. The video length per data source in Collegerama lecture search is shown in Table 5.9.

Table 5.9: Video length per data source in Collegerama lecture search for course CT3011

Data source (text type)	**Description**	**Minimum (sec)**	**Maximum (sec)**	**Mean (sec)**
Lecture title Transcript (lecture)	Lecture recording	1,351	3,231	2,451
Lecture chapter	Chapters by lecturer	15	2,197	592
Slide title Slide content Slide notes Transcript (slide)	Slide data	2	611	55
Transcript (sentence)	Subtitles	0.6	6.0	3.4
Transcript (word)	ASR output	0.0	3.4	0.3

Table 5.9 shows that the video length for slides may vary between 2 seconds and 7:28 minutes, with a mean value of 58 seconds. This means that on average the user has to wait for nearly 1 minute to encounter his searched item. This video length might be acceptable for recorded lectures, as most spoken text has a relevant surrounded text. In general, all spoken text belongs to that particular slide, as the lecturer explains the slide content. More detailed searching for a specific sentence can be achieved by searching in subtitles or time-tagged words (such as the ASR output of SHoUT). With time-tagged words, it is possible to show a karaoke-type subtitling, with sentences and coloring of the spoken word. An example of this can be seen at the website for Radio Oranje, in which old transcripts have been time-tagged by ASR (SHoUT).[51]

Precision and recall measurement

The effectiveness of an information retrieval system is often measured by the combination of "precision" and "recall".[17] Precision is the fraction of retrieved objects that is relevant. Recall is the fraction of relevant objects that is retrieved. These values can be defined in the following formulas:

$$Precision = \frac{r}{n}$$

$$Recall = \frac{r}{R}$$

In which:

r = number of relevant documents retrieved

n = number of documents retrieved

R = total number of relevant documents

The measurements require a set of objects or documents for which the number of relevant objects is known. For searching in recorded lectures, slides can be considered as documents since they give an overall overview of the subject matter. For the Collegerama lecture search engine, these test can be executed on the data of lecture #15. The slides of the lectures can be used as an object for these tests. A slide is regarded to give a completed subset of a lecture in which the related subject matter is explained. A detailed description of this test is given in Annex F, while the results are shown in Table 5.10. The test was done on 3 of the 10 "important words" of lecture #15: "stoffen", "grondwater" and "chloor".

The words "stoffen" and "grondwater" were selected because of their high ASR accuracy and their low occurrence in all lectures. It is assumed that these 2 keywords will also give a high ASR accuracy for the other lectures, despite the fact that most of these lectures were given by other lecturers. The low occurrence will result in more profound results. The word "chloor" was selected because of the missing of this word in ASR.

Table 5.10: Precision and recall measurement for different data sources on 3 important words of lecture #15

Item	Data source	Keyword stoffen	grondwater	chloor
Occurrences	subtitles	9	21	9
Number of related slides (R)	subtitles	4	5	2
Number of slides retrieved (n)	ASR	4	6	0
	slide titles	0	0	0
	slide content	0	1	0
	slide notes	0	1	0
Number of related slides retrieved (r)	ASR	4	5	0
	slide titles	0	0	0
	slide content	0	0	0
	slide notes	0	0	0
Recall (r / R)	ASR	100%	100%	0%
	slide titles	0%	0%	0%
	slide content	0%	0%	0%
	slide notes	0%	0%	0%
Precision (r / n)	ASR	100%	83%	-
	slide titles	-	-	-
	slide content	-	0%	-
	slide notes	-	0%	-

Table 5.10 shows that for the keyword “stoffen”, the slide recall and slide precision for ASR are both 100%, despite the fact that the retrieval rate for this keyword was only 89% (8 out of 9, according to Table 5.8). Although ASR has missed 1 occurrence, on an object/slide/document level this was not relevant. Slide data does not give any recall for this keyword and consequently no precision. For the keyword “grondwater”, the slide recall is 100% (20 out of 21, according to Table 5.8). The precision is 83% since 1 additional slide is retrieved by ASR. Again, for recall and precision the slide data is of no importance. The keyword “chloor” was missed by ASR and is not shown on the slides. Consequently the recall is 0%. The above mentioned recall and precision measurements show higher values on slide level than were obtained by a previous a known item/keyword search.

Multiple-keyword search

Multiple-keyword searching on individual subtitles will not give a positive result, as these keywords are never used in one particular sentence and won’t be retrieved as one record in the database. The same holds true for searching on individual words from ASR. A solution to this problem is offered by storing all spoken text belonging to a slide, called a slide transcript. The time-code contains a start and end time for the slide. The same is done for an entire lecture. This will allow for the searching of combined keywords. The student can use the slide or lecture timeframe as the starting point for further viewing.

Searching within spoken text per slide is included in the database but not implemented in the prototype for the web interface. Evaluation of this feature has been done directly on the database. This approach results in the storing of the same data in multiple records. Transcripts per lecture could be searched by a search engine using the transcript per word (ASR output). The approach used gives additional flexibility in the layout of transcripts, which enables more sophisticated output options. A lecture transcript can be printed in a more convenient way if additional line breaks are included. This option is not available if lecture transcripts are automatically abstracted from word transcripts.

If a multiple-keyword search is done on the ASR data for the words “stoffen” and “grondwater” in lecture #15, 8 results are returned. When clustering this result set by slide, there are only 2 slides out of a total of 29 slides that contain both keywords. The slide timeframe 24:07-25:09 gives 1 paired result and the slide timeframe 29:47-33:35 gives 4 paired results. The total viewing time for the combined results is reduced from the lecture duration of 45:09 minutes to only 4:50 minutes.

Table 5.11: Occurrence of combinations of 2 important words in all lectures

Lecture #	**stoffen**	**grondwater**	**stoffen + grondwater in lecture**	**stoffen + grondwater in slide**
15	8	20	8	1 + 4
17	11	12	11	1 + 1 + 1
19	11	26	11	1 + 1 + 2 + 1 + 2
20	3	2	2	0
21	4	75	4	1 + 1 + 1
23	15	21	15	1 + 2 + 1
25	1	10	1	1
Total occurrence	54	223	52	23
Number of lectures	8	18	7	6
Number of slides	-	-	270	17
Total duration	-	-	5:22:05	20:57

These 2 words can be searched in all lectures. Both keywords are present in 7 lectures. Without a multiple-keyword search per slide, this will require a total viewing time of 5:22:05 hours in order to see all results. If a search is done on slide level, only 6 lectures will be retrieved, with a total of 17 slides in which the combination of keywords is found. This reduces the viewing time to only 20:57 minutes. Searching on slide level reduces the total viewing time to 6.5%, or a reduction of 93.5%.

Ranked search results (implementation)

The search results have to be ordered according to a certain norm. In this research project, two of these options have been evaluated:

- time-based
- rank based

Time-based

In this order method, all the results are sorted in chronological order. This makes sense for recorded lectures, assuming the sequential explanation of key items in lectures. Later in the course, the key items are explained in further detail. In SQL Server, this can be accomplished by ordering the query results on Lecture_nr and Start_time. The query that can be used for this is shown below:

```
SELECT *
FROM Content
INNER JOIN Lectures ON Content.Lecture_id = Lectures.Lecture_id
WHERE CONTAINS (Text, 'stoffen')
AND Lecture_ID = '15'
ORDER BY Lectures.Lecture_nr, Start_time, Content.Text_type
```

Rank based

SQL Server has a function ranks search results based on several factors:

- text length
- number of occurrences of search words/phrases
- proximity of search words/phrases in proximity search
- user-defined weights

The query that can be used for this is shown below:

```
SELECT *
FROM Content AS FT_TBL INNER JOIN
CONTAINSTABLE(Content, Text, 'stoffen') AS KEY_TBL
ON FT_TBL.Content_id = KEY_TBL.[KEY]
WHERE Lecture_ID = '15'
ORDER BY KEY_TBL.RANK DESC;
```

Evaluation
With the rank based approach, the ASR results (text_type = 7) are higher ranked than the other types, because each record only contains one single word. According to the relevance ranking system Okapi BM25[58], these will be evaluated as being of high relevance. Similar results can be expected for subtitles and slide titles in comparison with slide notes and slide transcripts. These effects might be corrected by using user-defined weights for different text types. However, this has not been tested in this research project. The current search engine uses time-based ordering.

5.6 Conclusions

Table of contents
A table of contents (TOC) of a recorded lecture is an important element in the navigation of recorded lectures. Such a table of contents should be drafted by the lecturer or the assisting staff. It's useful to prevent such a TOC in an interactive way based on screenshots from the HD movie, the timeline and the available text data. This could be generated automatically into a Flash movie by accessing the database content and the related HD lecture movie.

Tag clouds
The accessibility of a recorded lecture could be further enlarged by creating tag clouds per lecture (limited to 15 words). The following conclusions have been made from these results:

- tag clouds should contain less than 15 words (nouns only)
- the best source of information for tag clouds are human-made subtitles
- tag clouds from subtitles (or speech recognition) are preferred over tag clouds from slide titles (or slide content / slide notes), because of their larger variance in font size
- tag clouds need a "best readable font"
- tag clouds could be improved by removing bad words chosen by the lecturer (in our examples, 25%-40% of the words were removed)

Search engine
A search engine on the course database is a useful element for enlarging the accessibility of the course and its lectures. It forms as an additional component over the navigation tools such as table of contents and tag clouds. The following conclusions can be made:

- the best source of information for searching are human-made subtitles followed by ASR output
- chapter titles and slide content has a low importance for searching
- chapter titles and slide titles are only relevant for the generation of table of contents
- by clustering subtitles or ASR per slide, multiple-keyword searching is largely improved because of shorter viewing times in the search results (in our example lecture, it was reduced from 5.3 hours to 21 minutes)

For the proper operation of a search engine, the output of a speech recognition system with sufficient word correctness is required. Better retrieval rates can be obtained with full subtitles. In view of the other beneficial elements of subtitles (for machine translation, and for better following of the lectures) these subtitles are considered as an essential part of all recorded lectures.

Future extensions
A text-based database per course can also form as a basic container for a course discussion board, using time stamped remarks ("questions and answers", discussions). A further extension to the database and search engine could be the adding of the other course material, such as readings (books, lecture notes), activities (assignments, tests, lab tests) and practice exams.

6. Proposed improvements

From the information and knowledge derived in this research project, as described in the previous chapters, it can be concluded that the usability of recorded lectures can be expanded. However, to increase the usability, it will be necessary to improve and extend the existing lecture recording and storage system. These improvements and extensions can be divided into these four categories:

- improved lecture accessibility
- improved navigation and searching
- addition of online discussion
- re-using recorded lectures to increase the course frequency

In paragraph 6.1 through 6.4, each of these elements will be discussed and the accompanying recommendations for improvement are mentioned. These improvements are a combination of conclusions from this research project, as well as suggestions and recommendations for future developments.

Paragraph 6.5 will give the outline of a pilot project for further development of these proposed improvements. This project can be regarded as a practical approach for implementing the conclusions and recommendations of the previous paragraphs.

6.1 Lecture accessibility

Improving the accessibility focuses on giving more students access to recorded lectures, independent of their location, computer device or operating system. The ultimate goal is to offer all lectures in several different video formats, as well as a small sized version that is designed specifically for mobile devices like the iPhone or Windows Smartphones.

All lectures need to have subtitles of the spoken language, as well as translated subtitles for the most common foreign languages such as English, Spanish, French, German and Chinese. This will support the student exchange programs that are available at most universities in the Netherlands.

We can divide the improvement of accessibility into these three general categories:

- vodcast distribution
- subtitling
- translation

Vodcast distribution

Since TU Delft likes to offer lectures to any student, no matter what his or her location is, several vodcast versions need to be produced. At the moment the only way to watch recorded lectures is by having access to a broadband Internet connection that has enough bandwidth to support the online streaming of videos. This makes it impossible to watch lectures while being in the train or bus, where a fixed high-speed Internet connection isn't available (mobile GPRS and EDGE data networks do not suffice).

A prototype for the integration between streaming and downloadable recorded lectures within the Blackboard environment is shown in Figure 6.1. This figure shows the different downloadable video formats in which this sample lecture is available, as well as the related course items.

Figure 6.1: Online viewing (YouTube) and available downloads and links for a recorded lecture (Source: http://blackboard.tudelft.nl CT3011-OpenCourseWare – Lecture new – demo-version)

Subtitling and translation
Subtitling has proved to be a substantial improvement to the online viewing experience of lectures. It is therefore recommended to display the subtitles of the spoken language for all different lectures in Collegerama. Furthermore, Dutch lectures (in the BSc phase) should be subtitled in proper English whenever the course is regarded as a useful resource for English speaking MSc students. For this goal, an automated translation as offered by Google Translate might be of insufficient quality. Additionally, English spoken courses could be subtitled in the Dutch language as a service to people who have trouble understanding English.

Subtitles available in one or two languages enables automated subtitling in other languages. Such an automated subtitling system could be convenient for non-native English-speaking students. This service reduces the need for using a dictionary in order to understand the lecture, which is common for Chinese students in their first MSc year.

Subtitles in the original spoken language can be created with the help of an ASR system such as SHoUT. The word-error rate of SHoUT is rather high (30%-70%), however these systems do provide an accurate timing of the spoken words. Human post-processing should improve the generated text and should divide the text in sentences, as is needed for proper subtitling. Figure 6.1 gives an impression of subtitles in the spoken language of a lecture. Subtitling and translated subtitles are further described in chapter 4.

6.2 Navigation and searching

Recorded lectures have an average duration of 30 minutes for a short lecture and 100 minutes for a double lecture session (discounting the break time). For first-time viewing this might be considered as acceptable, resembling the live course environment. However for reviewing lectures at a later time, better browsing, navigation and search capabilities are required. This is especially true for students who are studying for the exam and are browsing through the course material and/or doing course assignments.

Students also need a much better indication of the content of a certain lecture. The only available piece of metadata available is the lecture title. Searching for specific course content is not possible. The following improvements are recommended:
- browsing the lectures and its content through a course navigator and/or table of contents
- searching the course content (online search engine)
- indication of the course content by presenting a tag cloud for each lecture

The contours of a search engine and the creation of tag clouds have been described in chapter 5. A course and slide navigator could be produced from the content of the search engine. Such navigators function as an interactive table of contents. Figure 6.2 shows the improved navigation and searching within the Blackboard environment.

Figure 6.2: Tools created from the Collegerama database (slide navigator, tag clouds and search application) will significantly improve the accessibility of recorded lectures
(Source: http://blackboard.tudelft.nl CT3011-OpenCourseWare – Lecture (new) – demo-version)

6.3 Student interaction

Live lectures given in a lecture room allow for a direct form of communication between students and lecturer. This communication is two-way. The lecturer might ask the students some questions and receive feedback in order to test his educational performance. The rest of his lecture will then be based on this response. When a recorded lecture is used, this form of communication is no longer available.

A similar kind of discussion can be achieved by employing an online message board linked to each recorded lecture. Students will be able to ask questions, discuss events and topics during the lecture and receive feedback from the lecturer. During a live lecture the frequency of these questions is very low when the student attendance is very high. They are either too far away from the lecturer and/or students dislike interrupting a large classroom and drawing a lot of attention to them. Such an online messaging system also promotes student-to-student discussion and interaction that is not possible during a live lecture, since it will hinder the other classmates. In general, an online discussion board linked to recorded lectures could greatly increase and promote frequency of students asking questions.

An online discussion board will have even more value when the discussions are moderated by the lecturer or someone from the teaching staff. This moderation could include the answering of questions and the removal of silly unrelated remarks. This form of discussion can be complemented by adding the option to post time-tagged questions and comments. This means that the student can ask a question based on a certain timeframe within the lecture to which the question is relevant. With such a form of time-lined discussion, other students might look for specific remarks. These time-based discussions could be accessed by means of a search engine and/or a time slider that gives a popup whenever a discussion is related to that moment within the lecture. Figure 6.3 gives an impression of such a time-based discussion for online poker lectures.

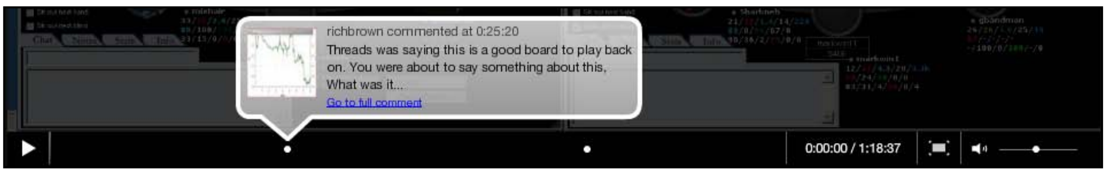


Figure 6.3: Time-lined online discussions on recorded lectures are common practice for the online educational poker community
(Source: http://www.deucescracked.com/videos/1210-Episode-Seven)

6.4 Increasing course frequency

Recorded lectures with improved accessibility and provided with online communication facilities could allow for the repeating of a course in the same academic year. These recurring courses might be of importance in the following situations:

- students following a minor program in another faculty (all scheduled in the first academic semester) might miss courses in their own faculty
- students with deadlines for BSc or MSc exams might encounter problems when preferred courses are not available in the current and/or next course period

These students can now be given the option of following and trying to pass the course through self-study, since all recorded lectures and accompanying material can now be shared. It could facilitate better study results and shorter study durations. A moderating lecturer can provide students with the required assistance and help by answering questions via the online communication facilities. Figure 6.4 gives a visual representation of these recurring courses.

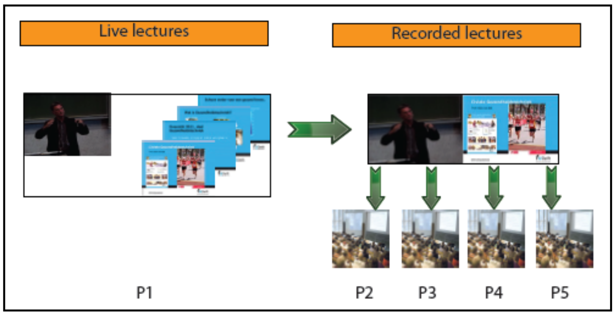


Figure 6.4: Multiple scheduling of courses with recorded lectures and online/moderated assistance by a lecturer

Time-critical courses

If TU Delft wants to apply this program of recurring courses within the same academic year, then this multiple scheduling is beneficial to the following types of time-critical courses:

- last year BSc courses
- minor-program courses (inside/outside faculty in first semester)
- courses for exchange students (Erasmus Mundus exchange in 1 semester)
- courses in cooperation with other universities (unparalleled scheduling)
- intensive courses (3 full weeks instead of 10 weeks of 30%)
- courses in graduate school (for starting PhD students, multiple starting moments)

Giving the students more freedom in choosing when to follow a certain course within an academic year, should have a positive influence on the time it takes for them to complete their education. Often times, a student will have to wait several months before he or she can follow a specific course that is required for them to finish their curriculum.

Figure 6.5 shows a visual representation of the current lecture situation, along with 3 possible ways to execute such a recurring course system. The green bars represent a live lecture that is given in front of students in a classroom. The yellow bars represent a course that is given primarily online, in which no live lectures are available. The red dots constitute the moments of examination.

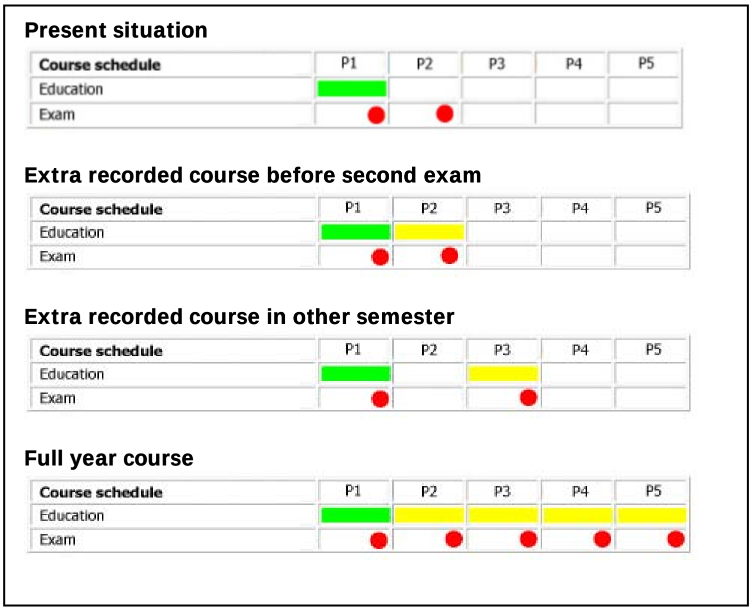


Figure 6.5: Examples of multiple scheduled courses

These additional online courses without live lectures should be provided with an online discussion board, to allow for the input of students by asking questions and comments of the lecturer by answering them. This further promotes students helping each other and starting a dialogue about the presented course material. The lecturer also acts as a moderator for this discussion board.

Scheduling

When all lectures are pre-recorded and available, it is easy to simply allow students access to all the lectures. In that fashion, they can decide whenever they want to watch a lecture. Another option is to create scheduled releases of pre-recorded lectures. This means that all lectures are made invisible, but are released at set intervals (for instance, every week). Such a system simulates the experience of following a live course in which students go to the classroom every week.

This form of scheduled releasing of lectures might give the following advantages:

- improving the weekly attendance by students (fixation in calendars of students)
- increased concurrent attendance by concentrating students into virtual classrooms
- allowing for moderation by lecturers (supporting the virtual classroom)

In the online poker teaching community, such a system is already employed. They offer the recurring releasing of pre-recorded lectures on a weekly basis. An impression of such a schedule that is offered at a poker instruction website called Deuces Cracked is shown in Figure 6.6.

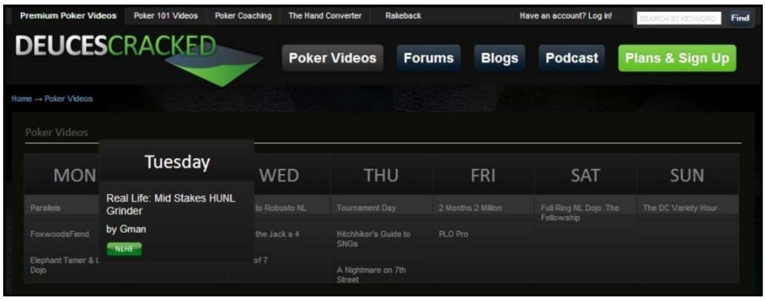


Figure 6.6: Online poker courses are scheduled on specific days, in order to enlarge the attendance and to promote live online discussion
(Source: http://www.deucescracked.com/)

6.5 Pilot project for further development

Goals

The above described improvements can best be developed in a pilot project under a real educational environment. The goals for the pilot project are summarized in Table 6.1. This table shows both the required short term improvements (1-3 years) as well as the long term goals (5-10 years).

Table 6.1: Current situation and goals for future academic courses

Current situation	Short term improvements	Long term goals
1 time a year	2 times a year (each semester)	5 times a year
1 location	between 1-3 locations (3TU)	3-10 locations (associated universities)
1 language	Dutch and English (subtitled)	plus 1 or 2 other local languages

The developments in Table 6.1 are based on two alternative approaches:

- classroom courses, with a live moderating lecturer
- scheduled self-study courses, with an online moderating lecturer

It is recommended that these are developed within the scope of a pilot project and run alongside the ongoing TU Delft OpenCourseWare project. A similar concurrent pilot project could also be done at University of Twente. The project should include about 5 to 10 courses, giving enough content to apply for a YouTube-Edu account and/or an iTunes-U account. These platforms require a minimum volume of around 100 video lectures organized in 5 to 10 courses.

The pilot project should focus on expanding the scheduling of courses from once a year to at least once per semester (repeated courses with recorded lectures) and the expansion of the course locations from only in Delft or Twente to at least one other location (simultaneous distant learning, with live streaming and the playing of recorded lectures). This approach covers a classroom environment. A classroom approach is preferred for this demo since it gives the smallest deviation to the current curriculum and it allows for the maximum amount of feedback from the students.

In a second phase, the focus could be shifted more towards individual self-learners. In this phase it should be established whether a scheduled organization gives better results over a free agenda approach.

Developments of new products

Different additional new products have to be developed in order to achieve the above mentioned goals within this pilot project. Table 6.2 gives an overview of these products, for which Figure 6.7 gives the relations.

Table 6.2: Additional products for expanded usability of recorded lectures

Item	Addition to / Replacement for	Responsible	Number per course
Videos - HD-video (YouTube) - Mini-video (iTunes) *	 Collegerama online view Collegerama online view	 MMS MMS	 5 - 30 5 - 30
Table of contents - Course (Flash) - Lectures (Flash)	 Collegerama catalog Collegerama slide navigator	 Lecturer Lecturer	 1 5 - 30
Subtitles - Course language - NL / EN (optional)	 Lectures NL in EN / EN in NL	 MMS MMS	 5 - 30 5 - 30
Search - Course search - Tag clouds	 TOC course/lectures TOC lectures	 MMS MMS	 1 5 - 30
Discussion board - Course discussion board - Lecture discussion board	 Course Lectures	 MMS / Lecturer MMS / Lecturer	 1 5 - 30

* if design differs from HD-videos

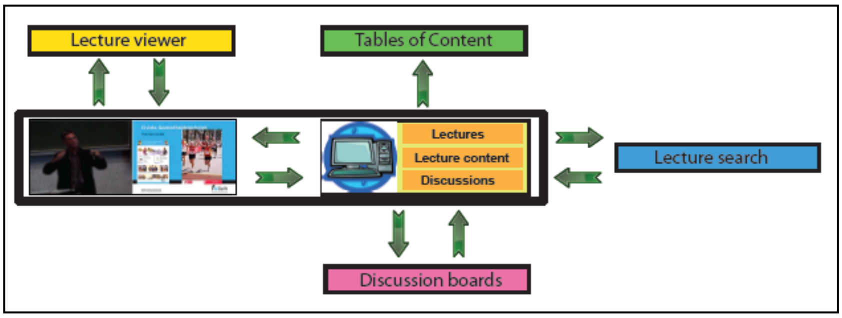


Figure 6.7: Recorded lectures are embedded in a Multimedia Information Retrieval System, containing multimedia content and structured course and lecture metadata

Requirements

The usability of recorded lectures can be expanded with the following requirements and/or additional provisions:

- proper recording
- HD movie creation
- post-processing by lecturer
- post-processing by data creator

Proper recording

It is concluded that a better system for recording slides needs to be developed. Looking at the future of education and the increasing developments in technology, it seems clear that presentations are going to be supported by animation and video. This means that an old screenshot recording system will no longer be sufficient to properly record PowerPoint slides. Re-using recorded lectures requires proper recording of a lecture. For this the following guidelines can be given:

- record lectures in a natural classroom environment (“recorded for a live audience”, no “talking head” recording)
- no slides, no recording (if not, creation of slides is required during post-processing)
- use full audio recording (minimum of 1 extra microphone, for introducing speaker and/or for the lecture room audience)
- add full screen recording options to Collegerama, for animations, electronic drawing boards, movies, computer demos (minimum of 5 fps, preferably 10-25 fps)
- original Collegerama camera (small size movie) should follow the lecturer at all times, never the projected slide or PowerPoint material

HD movie creation

The creation of a HD movie from a Collegerama recording will allow for the distribution of recorded lectures via YouTube, iTunes and Blackboard. For the creation of this HD movie, the following conclusions and guidelines can be presented:

- Collegerama recordings can be used as a basis for the creation of a HD movie (minimum of 1280x720)
- a HD movie is preferred for streaming and distribution
- a uniform design of HD movies is proposed
- several LQ movies can be derived from this HD movie for the distribution on alternative platforms (mobile phones, mobile media players, iPod/iPhone)
- the HD movie is prepared for subtitles (no hard-coded subtitles, always as separate subtitle text files)

Post-processing by the lecturer

Recorded lectures require post-processing with the following guidelines:

- provide lectures with proper lecture titles, speaker names etc
- divide lectures into 2-10 chapters (5-15 minutes per chapter)
- connect the video time-frame to the original PowerPoint slides
- eventually improve the slides and/or add slides (explaining text)

This post-processing should be done either within the Collegerama system (by using special login access for lecturers) or in a new recorded lecture data system.

Post-processing by Collegerama services

Recorded lectures require post-processing by a data creator with the following guidelines:

- import the slide data into a database (slide titles, slide content, slide notes)
- create tag clouds based on subtitles or slide titles for each lecture
- create subtitles for the lectures (at least in the spoken language, preferably in the additional Dutch or English language)
- create interactive tables of contents (both for the lectures in a course as well as for the chapters and slides in the individual lectures)
- create a search engine for course content and lectures
- create a discussion board for the course and the individual course lectures
- provide these elements within the Blackboard environment of the course

The post-processing for creating subtitles might be largely reduced when better performing ASR systems become available, which includes statistical post-processing of the result set produced by the word decoder of the system.

7. Conclusions

At present, Delft University of Technology records around 10% of their lectures. This number is expected to increase in the following years. Having these recorded lectures opens the door to all kinds of new ideas and improvements for their educational programs. At this moment they employ a video streaming system called Collegerama, which allows viewers with an active Internet connection to watch their lectures online. It combines a video stream of the lecturer with a series of screenshots of the accompanying PowerPoint slide.

In this thesis, a broad spectrum of possibilities for expanding the usability of recorded lectures has been examined and evaluated. The main research question for this project is:

How can we efficiently and effectively present recorded lectures and course material to students at universities?

This main research question has been divided into three sub-questions, which are discussed below.

How can we increase the accessibility and availability of the recorded lectures in Collegerama?
To increase the availability of the lectures, it is recommended to create a single video file from the Collegerama recordings. This will allow for the distribution over many other popular online multimedia platforms, such as YouTube-Edu and iTunes-U. A single video file distribution allows for offline viewing without an active broadband Internet connection (for example, while sitting in the train or lying at the beach). This is not possible within the current Collegerama system.

In this research project, a Collegerama lecture has been converted into a single video stream, after careful review of several layout designs and technical specifications. This lecture has been published on YouTube. Several other technical formats have been created, so that the lecture can also be distributed elsewhere. This includes a smaller sized version, created specifically for mobile devices and has been tested on Apple's latest iPhone.

How can we make recorded lectures easier to follow, especially for foreign speaking students?
To make lectures easier to follow, it is concluded that the creation and displaying of subtitles is useful. These subtitles can automatically be translated using machine translation. For this research project, Google Translate has been used which currently supports translation to 52 different languages. Although the quality of these has not been tested on Collegerama, evaluations in EACL show that around 20% to 50% of the time, adequate edited translations was obtained with machine translation. If necessary, this generated text can be enhanced by manual post-processing. The current speech recognition technology has also been evaluated for the generation of proper subtitles, using the speech engine created by University of Twente called SHoUT. It has an average word error rate of 50% and it's concluded that this system is not yet sufficient to generate proper subtitles and manual post-processing to improve the output is always required.

How can we effectively and efficiently navigate and search within recorded lectures?
This research project has shown that to properly navigate through the available recorded lectures, the input from teachers is important. They need to provide the lecture title and divide their lectures into several chapters with a proper chapter title, based on separate timeframes (start time and end time). These chapters together with the slide titles and slide content form the foundation for navigation and searching. The search element can be further expanded by the available subtitles. For the purpose of this research project, all lecture titles and chapters provided by the lecturer, slide titles and content and the generated SHoUT transcripts for all 14 lectures (28 lecture videos) have been collected. The slide metadata has been digitally and automatically extracted from the original PowerPoint files.

All this new information and metadata has been stored in a multimedia database, so that the retrieval options for the lecture content could be researched. This database will serve as the source for all the additional options for navigation and searching:

- generating a static and/or interactive table of contents for each lecture (based on the lecture chapters)
- generating tag clouds
- displaying subtitles in several different languages
- searching within lecture material

To demonstrate its functionality, a prototype for a Collegerama lecture search engine has been developed. This is an online web application that can be accessed from any location with an active Internet connection and searches within all the above mentioned data linked to a lecture. Every search result provides a link to Collegerama, so users can immediately see the related part of the lecture. The following conclusions can be made:

- the best source of information for searching are human-made subtitles followed by ASR output
- chapter titles and slide content has a low importance for searching
- chapter titles and slide titles are only relevant for the generation of table of contents
- by clustering subtitles or ASR per slide, multiple-keyword searching is largely improved because of shorter viewing times in the search results (in our example lecture, it was reduced from 5.3 hours to 21 minutes)

For the proper operation of a search engine, the output of a speech recognition system with sufficient word correctness is required. Better retrieval rates can be obtained with full subtitles. In view of the other beneficial elements of subtitles (for machine translation, and for better following of the lectures) these subtitles are considered as an essential part of all recorded lectures.

Future developments

It is concluded that a better system for recording slides needs to be developed. Looking at the future of education and the increasing developments in technology, it's clear that presentations are going to be supported by more animation and video. This means that an old screenshot recording system will no longer be sufficient to properly record PowerPoint slides.

To further increase the usability of the recorded lectures, a new interactive way to discuss lectures with the teacher and other students needs to be introduced. It promotes the asking and answering of questions, not just by the teacher but also by fellow classmates. This can be done through the use of a dynamic message board that is linked to the timeline of each lecture. Students can comment and discuss on the different topics in the lecture. To support such a system, an extension of the current multimedia database is required, so that the messages along with their optional timeframes can be stored.

With these recommendations, it is possible to use recorded lectures as a foundation for future online-given courses without the need for live lectures.

List of references

[1] Improving Mid-Range Reordering using Templates of Factors
Hieu Hoang and Philipp Koehn
In: Proceedings of the 12th Conference of the European Chapter of the ACL, pages 372–379, Athens, Greece, 30 March – 3 April 2009. Association for Computational Linguistics 2009

[2] SMT and SPE Machine Translation Systems for WMT'09
Holger Schwenk and Sadaf Abdul-Rauf and Loïc Barrault
In: Proceedings of the 4th EACL Workshop on Statistical Machine Translation , pages 130–134, Athens, Greece, 30 March – 31 March 2009. Association for Computational Linguistics 2009

[3] Proceedings of the 12th Conference of the European Chapter of the ACL
Athens, Greece, 30 March – 3 April 2009. Association for Computational Linguistics 2009

[4] Proceedings of the 4th EACL Workshop on Statistical Machine Translation
Athens, Greece, 30 March – 31 March 2009. Association for Computational Linguistics 2009

[5] Proceedings of 2009 Named Entities Workshop (NEWS2009): Shared Task on Transliteration
Singapore, 7 August 2009, The Association for Computational Linguistics and The Asian Federation of Natural Language Processing, 2009

[6] Findings of the 2009Workshop on Statistical Machine Translation
Chris Callison-Burch, Philipp Koehn, Christof Monz and Josh Schroeder
In: Proceedings of the 4th EACL Workshop on Statistical Machine Translation , pages 1-28, Athens, Greece, 30 March – 31 March 2009, Association for Computational Linguistics 2009

[7] An Overview of VC-1
Sridhar Srinivasan and Shankar L. Regunathan
In: Visual Communications and Image Processing 2005, Proc. of SPIE, 2005

[8] The VC-1 and H.264 Video Compression Standards for Broadband Video Services
Jae-Beom Lee and Hari Kalva
Springer 2008 (Google books: http://books.google.nl/books?id=MKhbLPHRb78C page 114-122)

[9] Conference Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009 Brighton, UK) (including Abstract book and conference programme)
International Speech Communication Association, 2009 (partly available at http://www.interspeech2009.org/conference/)

[10] Speech indexing
Roeland Ordelman, Franciska de Jong and David van Leeuween
In: Multimedia retrieval, Henk Blanken, Arjen P. De Vries, Henk Ernst Blok and Ling Feng, Springer 2007

[11] The Rich Transcription 2007 Speech-To-Text (STT) and Speaker Attributed STT (SASTT) Results
J. Fiscus and J. Ajot
In: Presentation at NIST's Rich Transcription 2007 Meeting Recognition Workshop, National Institute of Standards and Technology, 2007

[12] Evaluation of Text and Speech Systems
Laila Dybkjæ , Holmer Hemsen and Wolfgang Minker, Springer 2008

[13] Segmentation, Diarization and Speech Transcription: Surprise Data Unraveled (PhD thesis)
M. A. H. Huijbregts
University of Twente, Nov. 2008

[14] Multimedia Retrieval
Henk Blanken, Arjen P. de Vries, Henk Ernst Blok and Ling Feng
Springer 2007

[15] Languages for Metadata
Ling Feng, Rogier Brussee, Henk Blanken and Mettina Veenstra
In: Multimedia Retrieval, Henk Blanken, Arjen P. de Vries, Henk Ernst Blok and Ling Feng, Springer 2007

[16] Interaction
Erik Broertjes and Anton Nijholt
In: Multimedia Retrieval, Henk Blanken, Arjen P. de Vries, Henk Ernst Blok and Ling Feng, Springer 2007

[17] Evaluation of Multimedia Retrieval Systems
Djoerd Hiemstra and Wessel Kraaij
In: Multimedia Retrieval, Henk Blanken, Arjen P. de Vries, Henk Ernst Blok and Ling Feng, Springer 2007

[18] A Query Model to Synthesize Answer Intervals from Indexed Video Units
Sujet Pradhan, Keishi Tajima and Katsumi Tanaka
IEEE Transactions on Knowledge and Data Engineering, Volume 13, p. 824 - 838
IEEE 2001

[19] Towards a Unified Framework for Context-Preserving Video Retrieval and Summarization
Nimit Pattanasri, Somchai Chatvichienchai, and Katsumi Tanaka
In: Proceedings of 8th International Conferences on Asian Digital Libraries (ICADL 2005), Springer 2005

[20] TREC-6 1997 Spoken Document Retrieval Track - Overview and Results
John S. Garofolo, Ellen M. Voorhees, Vincent M. Stanford and Karen Sparck Jones
In: Proceedings of the 6th Text REtrieval Conference (TREC 6), National Institute of Standards and Technology (NIST), 1997

[21] Evaluation Campaigns and TRECVid
Alan F. Smeaton, Paul Over and Wessel Kraaij
In: Proceedings of the 8th ACM international workshop on Multimedia information retrieval (MIR 2006), Association for Computing Machinery (ACM), 2006

[22] Tagging: people-powered metadata for the social web
Gene Smith
New Riders (Pearson Education), 2008

[23] Submissions to the captioning standards review
Department of Communications, Information Technology and the Arts; Australian Caption Centre

List of URL’s

[24] MIT Facts 2009: Faculty and Staff [online] [Cites 20 October, 2009]
Access: http://web.mit.edu/facts/faculty.html

[25] Enrollment Statistics: MIT Office of the Registrar [online] [Cites 7 November, 2009]
Access: http://web.mit.edu/registrar/stats/yrpts/index.html

[26] Research – The Center for Measuring University Performance [online] [Cites 7 November, 2009]
Access: http://mup.asu.edu/research_data.html

[27] Kauffman Foundation study [online] [Cites 7 November, 2009]
Access: http://web.mit.edu/newsoffice/2009/kauffman-study-0217.html

[28] MIT [online]. Instructor Profile: Walter Lewin [Cites 20 October, 2009]
Access: http://ocw.mit.edu/OcwWeb/web/courses/instructors/lewin/lewin.htm

[29] Our History – MIT OpenCourseWare [online] [Cites 20 October, 2009]
Access: http://ocw.mit.edu/OcwWeb/web/about/history/index.htm

[30] Home Page - OpenCourseWare Consortium [online] [Cites 20 October, 2009]
Access: http://www.ocwconsortium.org/

[31] About – Academic Earth [online] [Cites 21 October, 2009]
Access: http://academicearth.org/about

[32] About VideoLectures.Net [online] [Cites 21 October, 2009]
Access: http://videolectures.net/site/about/

[33] Alexa Top 500 Global Sites [online] [Cites 21 October, 2009]
Access: http://www.alexa.com/topsites

[34] YouTube Blog: Higher Education for All [online] [Cites 21 October, 2009]
Access: http://www.youtube.com/blog?entry=uvxBVPuf4A8

[35] Apple Announces iTunes-U on the iTunes Store [online] [Cites 7 November, 2009]
Access: http://www.apple.com/pr/library/2007/05/30itunesu.html

[36] TAG - definition of TAG by the Free Online Dictionary [online] [Cites 7 November, 2009]
Access: http://www.thefreedictionary.com/TAG

[37] Google Translate [online] [Cites 7 November, 2009]
Access: http://translate.google.com/

[38] Collegerama.nl [online] [Cites 28 October, 2009]
Access: http://www.collegerama.nl/

[39] Paul Copier – LinkedIn [online] [Cites 28 October, 2009]
Access: http://www.linkedin.com/in/paulcopier

[40] Open Course Ware: Home [online] [Cites 28 October, 2009]
Access: http://ocw.tudelft.nl/

[41] Inleiding Videolectures [online] [Cites 28 October, 2009]
Access: http://www.utwente.nl/videolecture/

[42] Collegerama Etalage [online] [Cites 28 October, 2009]
Access: http://collegerama.tudelft.nl

[43] Silverlight-op-Linux plugin Moonlight 1.0 beschikbaar [online] [Cites 2 November, 2009]
Access: http://webwereld.nl/nieuws/55091/silverlight-op-linux-plugin-moonlight-1-0-beschikbaar.html

[44] YouTube Community Guidelines [online] [Cites 7 November, 2009]
Access: http://www.youtube.com/t/community_guidelines

[45] Adobe – Flash Player Statistics [online] [Cites 2 November, 2009]
Access: http://www.adobe.com/products/player_census/flashplayer/

[46] Learn More: Longer videos. - YouTube Help [online] [Cites 2 November, 2009]
Access: http://help.youtube.com/support/youtube/bin/answer.py?hl=en&answer=71673

[47] Getting Started: Video length and size - YouTube Help [online] [Cites 2 November, 2009]
Access: http://www.google.com/support/youtube/bin/answer.py?hl=en&answer=55743

[48] PDF Reference – sixth edition [online] [Cites 7 November, 2009]
Access: http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

[49] Adobe Acrobat Connect Pro [online] [Cites 7 November, 2009]
Access: http://www.adobe.com/products/acrobatconnectpro/

[50] The Association for Computational Linguistics [online] [Cites 25 November, 2009]
Access: http://www.aclweb.org/

[51] Radio Oranje demo [online] [Cites 7 December, 2009]
http://hmi.ewi.utwente.nl/choral/radiooranje.html
In:
Choral (http://hmi.ewi.utwente.nl/choral/index.html)
University of Twente, 2005-2010

[52] The EuroMatrix Project (Sept. 2006 – Febr. 2009) [online] [Cites 14 December, 2009]
Access: http://www.euromatrix.net/

[53] EuroMatrixPlus – Bringing Machine Translation for European Languages to the User [online] [Cites 14 December, 2009]
Access: http://www.euromatrixplus.net/

[54] Wordle – Beautiful Word Clouds [online] [Cites 14 December, 2009]
Access: http://www.wordle.net

[55] Make A Tag Cloud [online] [Cites 17 December, 2009]
Access: http://www.makecloud.com

[56] Keyword Cloud Generator [online] [Cites 17 December, 2009]
Access: http://www.tocloud.com/

[57] How High-Tech Dream Shattered In Scandal at Lernout & Hauspie [online] [Cites 14 December, 2009]
Access: http://web.archive.org/web/20060420133237/http://www2.gol.com/users/coynerhm/how_high.htm

[58] Integrating the Probabilistic Model BM25/BM25F into Lucene [online] [Cites 18 January, 2009]
Access: http://nlp.uned.es/~jperezi/Lucene-BM25/

Annexes

Accompanying material

DVD 1. Lectures CT3011

Lectures CT3011 28 lectures
Maps per Lecture, each map contains the following material:

Item	Description	Files	Total size (MB)
Lectures (video)	Collegerama video's of all lectures	28	3.140
Slides	Powerpoint presentations and abstracts	622	76
SHoUT	SHoUT- output and trancripts	117	272
Total		767	3.488

Lectures CT3011 Lecture no.15 Civiele Gezondheidstechniek
Maps for additional produced material from Lecture no. 15 Civiele Gezondheidstechniek:

Item	Description	Files	Total size (MB)
Lecture presentation	Powerpoint-file (plus pdf)	2	19
Collegerama_CT3011_11	Collegerama files	296	125
Vodcast	Vodcast try-outs and results....	42	3.081
Vodcast Camtasia	Screen captures Camtasia	3	148
Vodcast_TU	Vodcasts by TUDelft	2	229
Presenter	Presenter try-outs and results	59	269
Flash	Try-out for Flash	31	981
Subtitles	Subtitle files and try-outs	24	122
Speech recognition	Initial try-outs and evaluation SHoUT	12	13
Total		471	4.987