An HTML5-Based Online Editor for Creating Annotated Learning Videos

Jan-Torsten Milde

Hochschule Fulda, Computer Science Department, Germany
milde@hs-fulda.de

Abstract. Creating multi media learning resources has become a common standard in university level teaching. We present an online video annotation editor, allowing to create time aligned annotations of video material. The editor is implemented using HTML5-technology and runs in a standard web browser. The annotations are used to generate searchable indexes, making it easy to quickly navigate in the video.

Keywords: online video annotation, learning videos, time aligned annotations.

1 Introduction

Creating multi media learning resources has become a common standard in university level teaching. Large amounts of video data are created as video recordings of standard classroom lectures, as screencasts produced to introduce into software systems, or as dedicated learning resources produced in professional tv studios. One of the central drawbacks of such recordings is their lack of accessible internal structure. When learning with electronic resources, students need to have access to the presented content, especially when preparing for exams. An automatic processing of the video content is still very limited, providing results that are far from satisfactory.

In our approach we developed an online video annotation tool allowing to annotate video content based on timing information. The annotations are stored in a separate XML-file. Based on this data, structural navigation and search can be perfomed by the students. From the XML-file we are generating specific HTML5-based online views that provide the students with a content outline, search menus and also generate an alphabetic index. These ressources are all linked to the original video data, making it easy to identify the relevant content.

2 TASX-Corpora

Over the last decade multiple tools for the creation of multimodal annotated media resources have been realized (see [1], [2], [3], [4], [5]). Especially the creation of multi level annotated data sets has been under investigation here. Our work so far has been focusing on the development of tools for the creation of multimodal

P. Zaphiris and A. Ioannou (Eds.): LCT 2014, Part I, LNCS 8523, pp. 172–179, 2014.

corpora (TASX-Annotator, later the Eclipse-Annotator, see [8], [10]), followed by a tool for the creation of parallel text corpora (SAM, see[11]). The TASX-Annotator has been used to create annotated language recordings, including annotated corpora of video recordings of german sign language (see [9]).

From the collected data an XML-annotated multimodal corpus has been set up. The XSL-T based transformation of the data allows to generate multiple output formats from a single data source. The TASX-environment supports the complete corpus setup procedure: XML-based annotation of raw video data, the transformation of non XML-data and the analysis and dissemination of the corpus.

2.1 The TASX Format

A central aspect of our research ist to explore up which point standard XML technology (XML, XSL-T, XSL-FO, XPath, SVG, XQuery) can be used to model multi media corpora, to transform, query and distribute the content of such corpora and to perform adequate search and usage analysis. As a result all annotation data in our system ist stored in an XML-based format called TASX: the *T*ime *A*ligned *S*ignal data e*X*change format. A TASX-annotated corpus consists of a set of sessions, each one holding an arbitrary number of descriptive tiers, called layers. Each layer consists of a set of separated events. Each event stores some textual information (e.g. explanations form the teacher, question of the students) and is linked to the pimary video data of the classroom recording. This is realized by defining two time stamps per event, denoting the interval the event. Events may also carry non speech data, including slide changing marks, pointing directions, mode changes, external references. Relations between events on different tiers can be encoded by defining links using the ID/IDREFS mechanism of XML. This approach is comparable to using stand-off markup in the creation of multimodal corpora. Finally, arbitray meta-data can be assigned to the complete corpus, each session, each layer and each event. It might be necessaryto extend the meta data description in a way, that tree structured data can immediatly be described by XML-annotaitons. Currently we rather use the simpler version describing meta data in a linear structure. The following DTD formalizes the structutre of the TASX format:

```
<!-- corpus data -->
<!ELEMENT tasx (meta*,session+)>
<!ELEMENT session (meta*,layer+)>
<!ELEMENT layer (meta*,event+)>
<!ELEMENT event (#PCDATA,meta*)>
<!-- meta data -->
<!ELEMENT meta (desc*)>
<!ELEMENT desc (name,val)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT val (#PCDATA)>
```

```
<!-- atributes -->
<!ATTLIST session
  s-id CDATA #REQUIRED
  day CDATA #REQUIRED
  ref IDREF #IMPLIED
  month CDATA #REQUIRED
  year CDATA #REQUIRED>
<!ATTLIST layer
  l-id CDATA #REQUIRED
  ref IDREF #IMPLIED>
<!ATTLIST event
  e-id CDATA #REQUIRED
  start CDATA #REQUIRED
  end CDATA #REQUIRED
  ref IDREF #IMPLIED
  mid CDATA #IMPLIED
  len CDATA #IMPLIED>
<!ATTLIST meta
  m-id CDATA #REQUIRED
  ref IDREF #IMPLIED
  access CDATA #IMPLIED
  level CDATA #IMPLIED>
```

Despite of it's simplicity,the TASX-format is powerful enough to encode most of the corpus annotation formats currently in use. Indeed a number of format transformation programms have been implemented. For example, in order to reconstruct the equivalent annotation graphs representation of a TASX annotated corpus, one only has to collect the time stamps encoded in the start and end attributes of the event tags, sort them and then produce the timeline. Finally the time stamps of the events have to be replaced by references to the timeline.

2.2 The TASX-Annotator

In order to create TASX-annotated corpora the TASX-annotator has been developed. The programm is very user friendly and can be used without a high level of computer skills. It is possible to completely control the tool by either mouse or by keyboard shortcuts. Video and audio playback can be controlled by a foot switch. Different data views are programmed (time-aligned partiture, word-aligned partiture, sequential text view) to make annotation as effective as possible. The time aligned view is organized as a two dimensinal grid of infinite size.

A layer is presented as a horizontal tier of events. The order of the layers is arbitrary and can be changed instantly. The user is able to define time intervals by dragging the mouse. Each time interval represents an ev ent. The event is displayed as a graphical box which can be selected and moved with the mouse. The content of an event is entered in an additional text field.

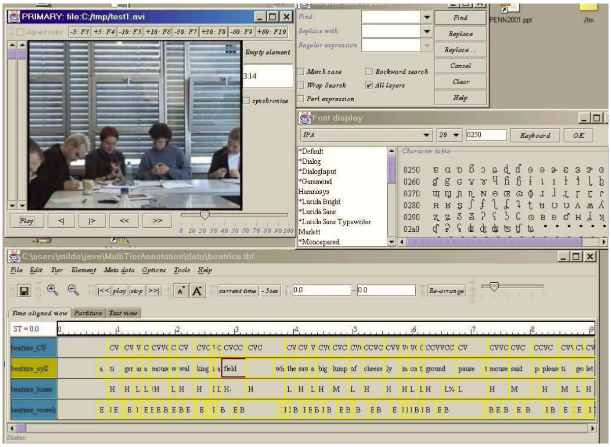


Fig. 1. A screenshot of the TASX-annotator. In the bottom half, the main panel is visible, where the time aligned tier view has been selected. On top of the main window, the font selection panel is visible (showing some IPA characters) and above it the find tool. In the upper left corner the video display can been seen.

Any (unicode) font (e.g. IPA fonts, HamNoSys fonts etc.) available for the operating system can be used for the transcription. The user can choose font and fontpage from a table displaying all characters of the selected font. It is also possible to define a virtual keyboard which maps the given keystrokes to arbitrary characters of the target font.

A separate video playback window will open up for each video file making it possible to e.g. display multiple perspectives of the same scene. The video playback is synchronized with the transcription. For audio transcriptions an oszillogram is calculated and is displayed inside the main window. In the text view the data can be manipulated in a standard text editor panel.

The content of the editor represents the layer and each line represents an event. A list selection box allows switching between different layers. It is possible to transfer text from standard text editors, e.g. Microsoft Word, by cut and paste operations. In order to additionally speed up the transcription process, a word completion function has been implemented for the text view. Entering the initial letter of a word will bring up all words starting with this letter. Once the text is tranferred into the TASX-annotator, the events still have to be aligned with the primary audio and video data.

Switching back to the time aligned view and moving the events with the mouse makes this task quite simple. In the partiture vie w the data cannot be edited. In practice this means that the data is transformed into an HTML table and then displayed to the user. A number of different HTML formatted views have been designed. The views can also be saved to external files and loaded into standard web browsers. One potential strength of the TASX-annotator is its manner of handling the export/import of XML based information. A standard way of solving this problem would be the implementation of a set of format specific XMLparsers which construct the internal representation of the XMLfile. While powerful integrated development systems make the design of such XML handlers simpler, it still remains a complex task to implement such a parser. In the TASX-annotatorwe follow a different approach. The system integrates an XSL-T processor (saxon), making it easy to perform on the fly data transformations. The import of an XML-file is split into two steps: first an XSL-T stylesheet transforms the XML file into TASX, second another XSL-T stylesheet will transform the TASX file into a simple text oriented format. This format can be loaded efficiently.

2.3 Pause Tracker

To speed up the annotation process a pause tracking programm has been developed. The programm separates speech from pauses and generates a TASX annotated XML document with two tiers, one holding all pause events, the other one holding all speech events. The tracker uses Praat to perform the actual speech analysis. It simply calculates the pitch curve of the audio signal. If no pitch is detected, then non-speech is assumed, otherwise speech. In a second step, the results of this classification are combined to continuous stretches of pauses/speech. Finally the TASX conformant output is generated. The pause tracker has shown to work quite reliably on a set of recording in different languages (Japanese, English, German, Saterfriesisch, French, Ega).

Even if tracking is far from perfect, the annotator gets a good pre-segmentation of the signal. This allows to move very quickly through the file, possibly performing minor adjustments to the boundaries or combining a set of separated events of one speaker . While the pause tracker gives good results when processing lecture recordings it is not of much help overlapping speech.

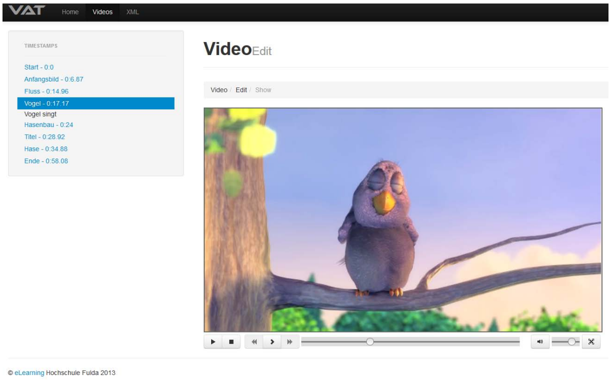
3 VAT: The Video Annotation Tool

Currently a large number of course recordings are created at the university. Students use these recordings as part of the preparation for the final exams. A big drawback of such recordings is their (long) duration and the lack of direct access to the content. Searching for slides, finding explanation of specific topics or keywords across all recordings is not possible.

In order to create more effective video learning ressources we have started to develop an online annotation editor for the creation of multi level annotated learning videos. The development is part of a cooperation between central elearning laboratory and the computer science department of Fulda University of Applied Sciences.

The annotation levels are synchronously linked to the video, each level containing an arbitrary number of events, holding a level specific description of the associated part of the video content. Events may overlap inside the levels and across levels.

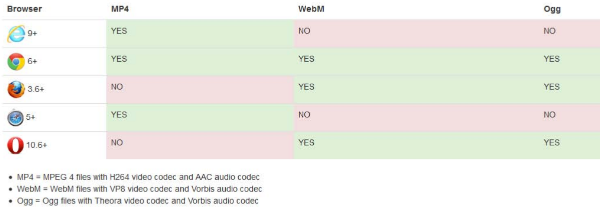
The annotation data is stored in the XML-based format TASX (Time Aligned Signal data eXchange format). The TASX files eventually carry all information



Start - 0:0
Anfangsbild - 0:6.87
Fluss - 0:14.96
Vogel - 0:17.17
Vogel singt
Hasenbau - 0:24
Titel - 0:28.92
Hase - 0:34.88
Ende - 0:58.08

© eLearning Hochschule Fulda 2013

Fig. 2. A short video annotated with the video annotation tool. The user is able to load video and annotation from separate files. The event list is presented to the user. A simple click moves the video to the corresponding position in the video. The complete system is realised as a HTML5 application, running in standard browsers on almost any platform.



Browser	MP4	WebM	Ogg
9+	YES	NO	NO
6+	YES	YES	YES
3.6+	NO	YES	YES
5+	YES	NO	NO
10.6+	NO	YES	YES

- MP4 = MPEG 4 files with H264 video codec and AAC audio codec
- WebM = WebM files with VP8 video codec and Vorbis audio codec
- Ogg = Ogg files with Theora video codec and Vorbis audio codec

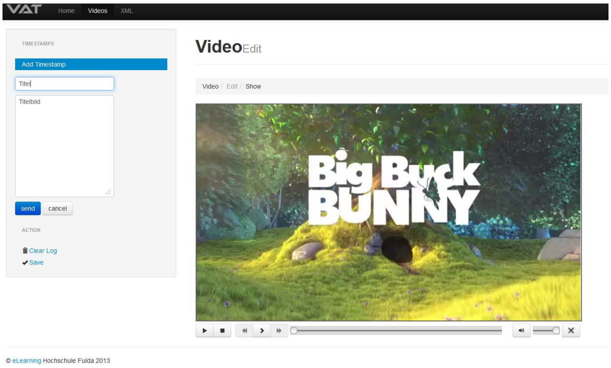
Fig. 3. The video annotation editor supports a number of different video formats. These depend on the support of the underlying web browser.

about a specific digital artifact, including the video file meta data, the level structure and of course all description events with their respective temporal information.

The implementation of the tool is based on HTML5 technologies. As such, it runs in most of the current web browsers and is therefore platform independent. The video material is streamed by the central video server of the university.

We tried to make the user interaction with the annotation editor as simple and effective as possible. Moving through the video data, adding layers and events and entering annotations is straight forward and fast. Upon key press, the editor will generate the XML-file, which can be stored locally or put onto the central elearning server.

Once the annotations are completed, the TASX file can used to generate various elearning resources. A linked index of all annotation is automatically created, making it possible to directly jump to specific topics. Further processing of both video and annotation data is realized with XSL-T programms. We have implemented programms to generate book like HTML structures out of the video



Titel

Titelbild

Fig. 4. Inserting an annotation with video annotation tool is simple. The annotator is moving to video to a definite position and enters the describing text. This also works, while the video is running. As soon, as text is entered, an annotation event is created and stored.

content. This process combines the textual annotation with screenshots of the slides and at same time provides the relevant (small) parts of the video. The students love these enhanced video view and use it very effectively during the exam prepration.

We have started to experiment with speech recognition and OCR in order to automatically extract annotations from the video. Results are promising, but further tests have to be performed.

4 Conclusions

We presented the development of a video annotation tool used for the creation of annotated learning videos. The underlying data is stored in an XML-file using the TASX-format. TASX provides a general format for the exchange of time aligned data, thus being specifically useful in the context of lecture annotations. While already powerful annotation tools existed, we chose to create a simple to learn online video annotation tool. Using standard HTML5-technology makes it possible to quickly create annotated version of lectures, both by students and teachers.

References

[1] Kipp, M.: Anvil - A Generic Annotation Tool for Multimodal Dialogue. In: Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech), pp. 1367–1370 (2001)

[2] Schmidt, T.: Transcribing and annotating spoken language with EXMARaLDA. In: Proceedings of the LREC-Workshop on XML Based Richly Annotated Corpora, Lisbon. ELRA, Paris (2004)

[3] Sasaki, F., Wegener, C., Witt, A., Metzing, D., Pönninghaus, J.: Co-reference annotation and resources: A multilingual corpus of typologically diverse languages. In: Proceedings of the 3nd International Conference on Language Resources and Evaluation (LREC 2002), Las Palmas (2002)

[4] Broeder, D., Offenga, F., Willems, D., Wittenburg, P.: The IMDI Metadata set, its Tools and accessible Linguistic databases. In: IRCS Workshop, Philadelphia (2001)

[5] Bird, S., Liberman, M.: A Formal Framework for Linguistic Annotation. Speech Communication 33(1-2), 23–60

[6] Milde, J.-T., Gut, U.: The TASX environment: an XML-based toolset for time aligned speech corpora. In: Proceedings of the third International Conference on Language Resources and Evaluation, Las Palmas, pp. 1922–1927 (2002b)

[7] Milde, J.-T., Gut, U.: A Prosodic Corpus of Non-Native Speech. In: Bel, B., Marlien, I. (eds.) Proceedings of the Speech Prosody Conference, pp. 503–506. LPL, Aix-en-Provence (2002a)

[8] Behrens, F., Milde, J.: The Eclipse Annotator: an extensible system for multimodal corpus creation. In: Proceedings of the Fifth International Conference on Language Resources and Evaluation, Genova (2006)

[9] Geilfuss, M., Milde, J.-T.: SAM - ein Annotationseditor für parallele Texte. In: Berliner XML Tage, pp. 71–78 (2005)

[10] Sippel, T.: Eine Anwendung zur touchgesteuerten Annotation elektronischer METS/MODS-Dokumente in Silverlight C#. Masterthesis. Hochschule Fulda (2011)