

Advene as a Tailorable Hypervideo Authoring Tool: a Case Study

Olivier Aubert
Université de Lyon, CNRS
Université Lyon 1, LIRIS,
UMR5205, France
olivier.aubert@liris.cnrs.fr

Yannick Prié
Université de Lyon, CNRS
Université Lyon 1, LIRIS,
UMR5205, France
yannick.prie@liris.cnrs.fr

Daniel Schmitt
Université de Strasbourg,
LISEC EA-2310, France
daniel.schmitt@etu.unistra.fr

ABSTRACT

Audiovisual documents provide a great primary material for analysis in multiple domains, such as sociology or interaction studies. Video annotation tools offer new ways of analysing these documents, beyond the conventional transcription. However, these tools are often dedicated to specific domains, putting constraints on the data model or interfaces that may not be convenient for alternative uses. Moreover, most tools serve as exploratory and analysis instruments only, not proposing export formats suitable for publication.

We describe in this paper a usage of the Advene software, a versatile video annotation tool that can be tailored for various kinds of analyses: users can define their own analysis structure and visualizations, and share their analyses either as structured annotations with visualization templates, or published on the Web as hypervideo documents. We explain how users can customize the software through the definition of their own data structures and visualizations. We illustrate this adaptability through an actual usage for interview analysis.

Categories and Subject Descriptors

H.5.4 [Hypertext/Hypermedia]: User Issues, Navigation;
H.5.1 [Multimedia Information Systems]: Video

Keywords

Hypervideo, Annotation, Video Analysis, Mediafragments, Active Reading

1. INTRODUCTION

As described in [1], video analysis is often carried out through active reading, an iterative activity based on the creation of structured metadata and the usage of various visualizations for this metadata, in an intermediary stage for video navigation or as a final publishing stage. Documents produced by this activity can take the form of *hypervideos*.

Many video annotation tools are often strongly tied to a specific application domain. This allows them to offer a customized experience, with dedicated tools and interfaces, at the expense of some difficulty to be used outside of intended uses. For instance, a tool like ELAN [8] does not allow overlapping annotations, which may make sense in the linguistic context but may not be convenient for other practices. Moreover, most specialized tools being aimed at analysis, they do not particularly insist on the final phase of the process: publishing documents. Tools like Anvil [4] or Elan [8] propose multiple predefined structured export formats but do not allow users to define their own visualizations. On the other hand, rendering environments like [5] do not focus on the analysis side.

The Advene project and its main application Advene [2] are trying to reach a larger usability spectrum by offering a number of generic tools and interfaces, that can be user-customized to fit the application domain. Users can also define their own visualizations, in order to accommodate their specific needs and possibly to generate publishable documents. We claim that it is necessary that tools adapt to some extent to the user's needs, especially in such an exploratory activity. Furthermore, this flexibility must always be effective, in order to accompany the continuous evolution of the user's ideas and needs.

In this article, we first present how Advene has been conceived to provide generic but customizable tools in order to accommodate a variety of usages. We briefly highlight the main features that provide this flexibility. Then we illustrate these capabilities by detailing a real use example and how Advene features were used to adapt to this particular situation.

2. ADVENE PRINCIPLES

The Advene software covers the whole range of active reading activities [1]: 1/ it allows to create annotations, either through dedicated interfaces, by importing data or through some feature extraction algorithms. Annotations can be structured through user-defined schemas. They also can be linked through n-ary relations, also user-defined, in order to express any kind of relationship. 2/ Once annotations are present, they can be used as a reference to navigate the audiovisual document, in order to investigate the material and possibly to update or extend the annotations. 3/ Annotations and possibly pieces from the video document can be used to generate various visualizations, as custom interfaces or as documents that may be published. The application offers a number of customizable interfaces for various

actions (annotation edition, visualization or search). It integrates a template-based document generation system, allowing users to specify various renderings for their data. Eventually, the application can be extended by developers either through the use of plugins, for dedicated developments, or by contributions, Advene being free software.

The principles and architecture of Advene have been described with more details in previous articles [2]. We will focus here briefly on the main aspects of Advene that allow more specifically its flexibility.

2.1 Data model

As described in [2], the Advene/Cinelab¹ data model features explicitly structured annotation data, as well as the definition of queries and visualization templates. In our audiovisual context, we define an annotation as a piece of data of any kind associated with a video fragment, i.e. with two timecodes designating beginning and ending instants in a given video.

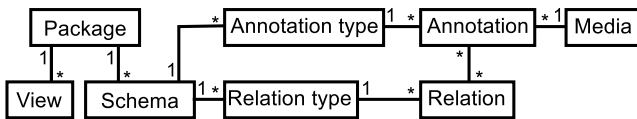


Figure 1: Global representation of the Advene/Cinelab data model

The annotation structure, summarized in figure 1, consists in user-defined annotation types and relation types that can themselves be grouped into schemas that represent a given point of view on the analysis. Views are used to transform annotations, along with the video, into visualizations appropriate for the user’s task.

The model has been defined to be as expressive and flexible as possible. The definition of annotation types gives a way to categorize annotations, which is useful both for the user and for the computer. Relations, classified by relation types, may be used to express any graph structure, or more basically to link annotations according to the user’s needs.

The annotation structure (annotation types and relation types) is defined by the user to fit his or her needs. It can be defined iteratively, and interactively, in order to foster experimentation with various analysis structures. New annotation types can be created on the fly, and annotations copied into this new type, seamlessly.

2.2 Visualization - hypervideo

Advene provides three main types of visualizations: ad-hoc views, dynamic views and static views.

Ad-hoc views are graphical components present in the software, such as a timeline or a tabular view. Such components provide great interactivity, and can be customized to some extent by users. For instance, users can define which annotation types are displayed in a timeline, and can save this configuration for future invocations.

Dynamic views use the annotation data to modify the rendering of the video while it is being played. Users can modify the display (caption textual or graphical data over the video, etc) or the time (automatically navigating from

one annotation to another one), or trigger different actions such as using text-to-speech or a Braille table [3] to render annotation data while the video is playing. Users can define dynamic views through rulesets, comparable to the filtering rules that can be found in e-mail client software.

Static views are visualizations specified by XML-based templates, combining information from the annotation structure and from the video. Using an XML-compatible template [9], users can define any kind of visualization according to their needs. The use of a template language is intended to lower the barrier to conceiving such visualizations. The Advene application embeds a webserver that dynamically serves the rendered templates to any web browser, or through an embedded HTML component. Links into the video control the Advene video player, offering the ability to directly play fragments related to annotations. This gives users the possibility to modify the underlying metadata (annotations) and get instant updates of the generated documents.

2.2.1 Hypervideo publishing

However, this approach implies that the Advene software has to be used to access visualizations and to fully experience their hypervideo nature by synchronizing with the Advene video player. In order to provide a publishing functionality, we integrated the possibility to statically export a set of defined visualizations.

The web export feature is basically a specialized version of a website copier, that walks through the defined visualizations (which are most often HTML documents with some specific links to control the Advene video player), captures the HTML content and injects javascript code to embed an HTML video player. Links to the Advene video player are converted into corresponding Media Fragment [7] URLs, to be able to address the embedded HTML video player.

Some features can be activated in the exported version, through the definition of HTML classes in the template. For instance, specifying the `transcriptHighlight` class will highlight elements of the class `transcript` synchronously with the video player.

We have seen that in addition to the annotation creation and navigation tools, user-definable visualizations allow Advene to cover the whole annotation lifecycle, from annotation creation to publishing. Besides, user-defined data structure and views offer a high level of tailorability. Their definitions represent the specialization of the tool, and they can be themselves shared and reused. We will now present how these features have been used through a real use case.

3. USE CASE

3.1 Context

Advene is used by Daniel Schmitt for his analysis of the visitors perception of museum exhibitions². His experiment consists in fitting on visitors a small video camera mounted on glasses. The camera records a *subjective view* of the visit. At the exit of the exhibition, the subject and the researcher watch together the recorded video, and the researcher invites the subject to comment his or her behavior. For instance,

¹The Advene data model has been refined and used in the context of the Cinelab project, see <http://www.advene.org/cinelab>.

²More explanations (in french) and public examples of interview analysis hypervideos are available on the <http://www.museographie.fr> website.

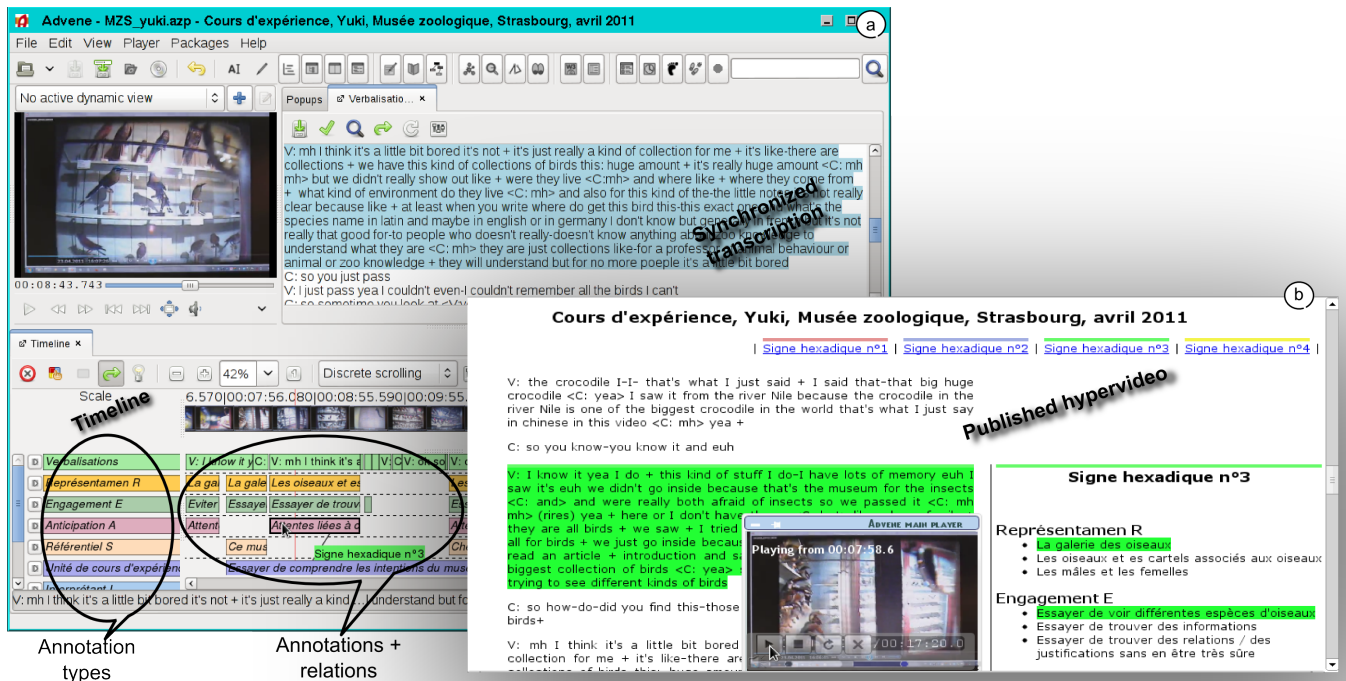


Figure 2: (a) Advene application GUI (b) Example of published hypervideo

he or she can be asked about his or her feelings when he or she stared for so long at a given piece of art, or sighed when arriving in front of another.

The interview process is itself recorded, and this recording constitutes the analyzed material: the audio track records the exchanges between the subject and the researcher, while the video track captures the video player that is used to replay the subjective view.

Once the interview is over, the researcher can later analyze the recorded material. He used the course-of-action [6] methodology, which is situated in the enaction paradigm. It involves recording a subject activity, and inviting the subject to verbalize explanations during a self-confrontation phase. The course-of-action methodology then distinguishes 6 kinds of components in the discourse, named *hexadic signs*: Involvement in the situation, Potential actuality, Referential, Representamen, Unit of course of experience and Interpretant. The researcher's aim is to identify in the discourse of the subject various *courses of experience*, in order to find out how the visitor builds up knowledge in the course of her or his action. For instance, from the following interview (extract):

Yuki and her friend, both students in biology (Master) are visiting the bird gallery in a zoological Museum. They look at different species and compare the size or the color of birds. This methodology shows that in fact, Yuki is disappointed not to find more indications.

Visitor: Before we came to this museum, I read an article... introduction and it said this museum has one of the biggest collection of birds [...] I think it's a little bit bored... we have this kind of collections of birds... this huge amount... it's really a huge amount but we didn't really show out like, where they live, where they come from, what kind of environment do they live and also for this kind of little notes, it's not really clear because [...] they are just collections for a professor of animal behavior or animal or zoo knowledge [...] I'm very

interested maybe just for the color... or maybe for... oh it's cute! [...] there are too many.

the researcher may identify the following hexadic signs:

Representamen The birds in the gallery
Involvement in the situation Try to find information
Potential actuality Would like to know where the birds live, where they come from and their environment
Referential This museum presents one of the biggest collection of birds / (I am student in biology)
Interpretant This gallery doesn't provide enough information
Unity of course of experience Try to understand what the museum intends to do without success

3.2 Methodology implementation

The implementation of this methodology in Advene was done in 4 steps, that include the definition of the annotation structure and of the visualization template that were done only once. After the structure and visualization were defined, the necessary steps were then reduced.

1. A transcription of the interview is carried out, using Advene dedicated interfaces – mainly the synchronized notetaking view. It appeared later that thanks to the interaction features of Advene, the exact transcription was not always strictly necessary for the analysis: some keywords used for navigation, accompanied with direct access to the audiovisual material could sometimes be enough to carry on the task of identifying the hexadic signs. However, transcription, though costly, is always useful for other uses such as publication.

2. The annotation structure is defined, based on the course-of-action [6] methodology. This consists in finding out how to best translate hexadic signs into data structures. The flexibility of Advene in the definition of data structures allowed rapid experimentation with different setups before finalizing an appropriate schema, which was then used for

all analyses. The final structure defines one annotation type for each hexadic sign, and uses relations to identify related hexadic sign occurrences that take part in a single course of experience.

3. Once the structure is stabilized, the video can be analyzed along the different hexadic signs. Fragments of the discourse are identified by the researcher as being instances of a given hexadic sign. A new annotation is then created in the type corresponding to the sign, with an explanation of the analysis. Advene ability to copy and move elements from one type to another was used. For instance, verbalization timecodes could be simply used as references for the creation of hexadic sign annotations.

The timeline view, shown in figure 2(a), proved to be an interesting and inspiring visualization for these hexadic signs, providing visual time indications and some kind of temporal grouping. The Advene transcription view, visible in figure 2(a), presents verbalizations as an interactive text document, usable to navigate the video. It was also used to provide contextual information and as a navigation interface.

4. In order to publish on the web, an Advene static view is defined as a template (that is reused for all analyses). It generates a transcription of the verbalizations synchronized with the video, and aligned with the hexadic signs in a second column, also synchronized with the video. This visualization brings another perspective on the same data, by visually grouping elements (verbalizations and hexadic signs) that may be dispersed in other views. It offers another anchor to allow the emergence of the meaning.

An example output is presented in figure 2(b). An embedded video player is synchronized with the transcription, which is sided with the hexadic signs. Highlighting is used to indicate the parts of the transcription and analysis that are related, and which ones concern the active video moment. The view can be visualized dynamically through the embedded web server, and an Advene-independent version can be extracted through the website export feature of Advene, in order to be put online on any website. The document presented here is publicly accessible on the <http://www.museographie.fr> website.

3.3 Evaluation results and lessons

More than 40 interviews have been carried out and analyzed using this setup. The availability of different visualizations for the annotations proved to be precious for the analysis, sometimes opening new perspectives on the data. For instance, the graphical timeline offers an overview of the temporal distribution of hexadic signs over the whole interview, that allows to distinguish temporal patterns in the discourse, while the specialized HTML visualization provides another perspective by grouping verbalizations and related hexadic signs.

The possibility to add on-the-fly new structures, new annotations or new visualizations, allows the researcher to build his own dedicated tools during the course of his research. Advene plasticity accompanies the researcher in the emergence of ideas.

Eventually, the possibility to share the analyses, either as published web documents like the one presented in figure 2(b) or as structured data (a package containing the definition of the annotations, of their structure and of the

defined views) makes it possible to subject the analysis to a critic, but also to constitute communities of experience.

4. CONCLUSION

Advene is a generic video annotation platform offering explicit means of customization (schemas for data structure, parameters for visualization customization, templates for document generation) in order to be adaptable to specific tasks. The definitions of the specialized data structure and visualizations express the adaptation of the tool, and can be shared and reused. After presenting these features, we illustrated this adaptability through a real-world example, where Advene proved to be an appropriate companion in the development of an exploratory research.

The intention of Advene is to lower the barrier to conceiving dedicated structures and visualizations. The collaboration of researchers and programmers was still needed to design some of the templates, but we intend to gather more examples to be able to build a library which could serve as inspiration, and also make directly available some of the most used visualizations.

Acknowledgements. This work has been partially funded by the French FUI (Fonds Unique Interministériel) - CineCast project.

5. REFERENCES

- [1] O. Aubert, P.-A. Champin, Y. Prié, and B. Richard. Active reading and hypervideo production. *Multimedia Systems Journal, Special Issue on Canonical Processes of Media Production*, page 6 pp., 2008.
- [2] O. Aubert and Y. Prié. Advene: active reading through hypervideo. In *ACM Hypertext '05*, pages 235–244, Salzburg, Austria, Sep 2005.
- [3] B. Encelle, M. Ollagnier-Beldame, S. Pouchot, and Y. Prié. Annotation-based video enrichment for blind people: A pilot study on the use of earcons and speech synthesis. In *13th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 123–130, Dundee, Scotland, Oct 2011.
- [4] M. Kipp. ANVIL - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of Eurospeech 2001*, pages 1367–1370, Aalborg, Sep 2001.
- [5] Mozilla. *Popcorn.js The HTML5 Media Framework*, 2012. Website <http://www.popcornjs.org/> accessed 2012/01/30.
- [6] J. Theureau. *Handbook of Cognitive Task Design*, chapter Course of action analysis and course of action centered design. Lawrence Erlbaum Ass., 2003.
- [7] R. Troncy, E. Mannens, S. Pfeiffer, and D. V. Deursen. Media Fragments URI 1.0. Technical report, W3C, 2011.
- [8] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. Elan: a professional framework for multimodality research. In *In Proceedings of Language Resources and Evaluation Conference (LREC)*, 2006.
- [9] Zope Corporation. *Zope Page Templates reference*, 2004. http://www.zope.org/Documentation/Books/ZopeBook/2_6Edition/AppendixC.stx.