

Using Advene to bridge the gap between users and ontologies in movie annotation

O. Aubert¹ 

¹Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

Abstract

Feature movies and documentaries analysis has always relied on available access tools and possibilities: movie theaters required memorizing whole sequences, home video (VHS, DVD) brought new possibilities for analysis. Digital video tools now provide additional capabilities, such as video annotation, which is sometimes used in research contexts, from simple synchronized note-taking to more structured approaches.

The AdA project[†] of the Cinepoietics team of Freie Universität de Berlin aims at investigating the audiovisual rhetorics of affect in audiovisual media on the financial crisis. The analyses are framed by theoretical assumptions on the process of film-viewing, and one of the goals of the project is to study in what measure a systematic approach based on semantic annotations of the audiovisual corpus can bring a new light to the reflexions. Such an approach requires appropriate tooling for humanities researchers.

We will describe in this contribution how the Advene video annotation platform has been extended and used to produce and use semantic annotations, and to validate the underlying ontology, accompanying the humanities researcher practices in the AdA project.

1. Introduction - Context

The AdA project[‡] of the Cinepoietics team of Freie Universität de Berlin aims at investigating the audiovisual rhetorics of affect in audiovisual media on the financial crisis. The analyses are framed by theoretical assumptions on the process of film-viewing, and one of the goals of the project is to study in what measure a systematic approach based on semantic annotations of the audiovisual corpus can bring a new light to the reflexions. Such an approach requires appropriate tooling for humanities researchers.

We describe in this contribution how the Advene video annotation platform has been used and extended to produce and use semantic annotations, and to validate the underlying ontology, accompanying the humanities researcher practices.

Goals – Methodology The Cinepoietics team from Freie Universität Berlin previously worked on eMAEX [BGSS20], a standardized Method of Analyzing Qualities of Filmic Expressivity. This experience is one of the source of inspiration for the AdA project, which aims at evaluating systematic use of digital tools for movie analysis. This involves defining an ontology appropriate for the domain, evaluating data structured along this ontology in a movie cor-

pus, and studying the automation possibilities to produce some of the data through audio/video analysis. The computing side of the project was done by a HPI team, providing support for semantic data storage and manipulation [ARAHS18] as well as experimenting with automated extraction tasks.

Corpus The project video corpus is composed of 13 feature films, 76 documentaries and 315 TV news shows on the global financial crisis after 2007. It provides the context in which the expression of audio-visual rhetorics of affect is explored. Each video document has been described with semantic metadata[§] structured through an ontology based on existing vocabularies such as the DBpedia Ontology, Schema.org, and the Linked Movie DataBase.

Application requirements In this context, appropriate tooling is required to provide back and forth integration of semantic metadata (annotation ontology that provides the annotation structure, and annotation data) with the Cinepoietics scholar workflow. It especially needs to provide acceptable ways for users to annotate videos using constrained vocabularies, while retaining the possibility to extend the ontology.

Multiple challenges were then identified, and constituted the

[†] *Affektrhetoriken des Audiovisuellen* <http://www.ada.cinepoietics.fu-berlin.de/>

[§] available at <https://ada.cinepoietics.org/corpus/>

base requirements for identifying the appropriate tool. First, the tool needs to provide a usable annotation interface for humanities researchers, providing video annotation/navigation with structured annotations. It must provide appropriate video navigation and annotation features to facilitate in-depth annotation of the corpus. Second, the tool needs to provide a flexible data structure that can express an underlying ontology, either natively or through preservation mechanisms. Finally, the tool must offer some flexibility and extensibility in order to integrate data from external sources, especially algorithms developed during the project. Based on these constraints, multiple solutions were evaluated and the Advene application was selected.

2. Advene

The Advene software [AP05] is a free-software video annotation platform that allows to annotate digital videos and use these annotations for navigating inside the videos as well as author new documents called *hypervideos*. Among other video annotation solutions like Elan [WBR*06], Anvil [Kip01], Transana [Tho02], Lignes de Temps... , it features adaptability by design to a wide variety of contexts, through generic and user-defined data structures. It also has the capacity to generate custom, user-defined hypervideos combining information from the video and from the annotation structure. And it provides extensibility features through plug-ins covering import/export filters.

Advene already provided a number of features necessary for the project. During the project, its user interface has been extended and streamlined to facilitate manual annotation with ontology-conforming data. A data validation framework has been added, with multiple validation modules, in order to help identify conformance issues. RDF interoperability has been implemented through a OWL import filter for defining the data structure, and JSON-LD import/export filters to integrate the data into a semantic triple store. In addition, visual feature extractions algorithms have been integrated into Advene. This article will describe these extensions and their usage in the project.

3. Ontology integration

The semantic web extensions developed during the project are two-fold. First, an ontology import filter has been developed. It allows to import an ontology defined with the OWL standard into Advene, defining an annotation structure that preserves references to the original ontology URIs so that appropriate information can be generated afterwards.

The AdA Filmontology [BSSAR20] has been defined and refined during the project. As presented in figure 2, the ontology defines the `AnnotationType` owl:Class. Instances of this type can be grouped in more generic `AnnotationLevel` owl:Class, that define consistent domains of data such as `ImageComposition`, `Acoustics` or `Camera`.

An `AnnotationType` can be linked through a `predefinedValue` property to multiple `AnnotationValue` elements, that can express a

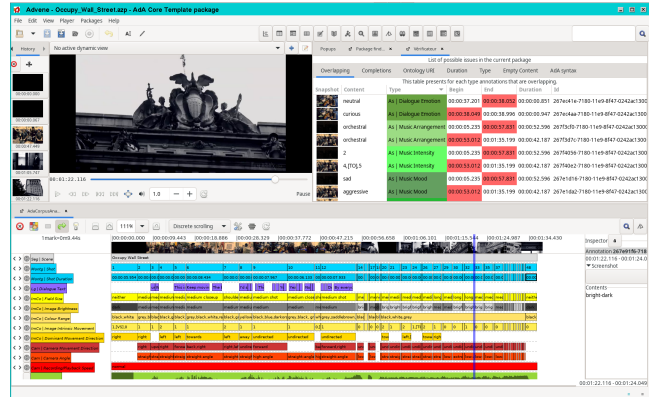


Figure 1: Advene in the AdA project - this screenshot presents the Advene timeline below the video, and the data validation module on the right.

set of standardized values with additional metadata. For instance, the `FieldSize` annotation type references values such as `FieldSize_closeup`, `FieldSize_long_shot`, etc.

Additional constraints on the `AnnotationType` can be defined through subclasses. For instance, `EvolvingAnnotationType` indicates that its annotations may describe an evolving value, such as a `Field Size` going from `long shot` to `close-up` during a shot.

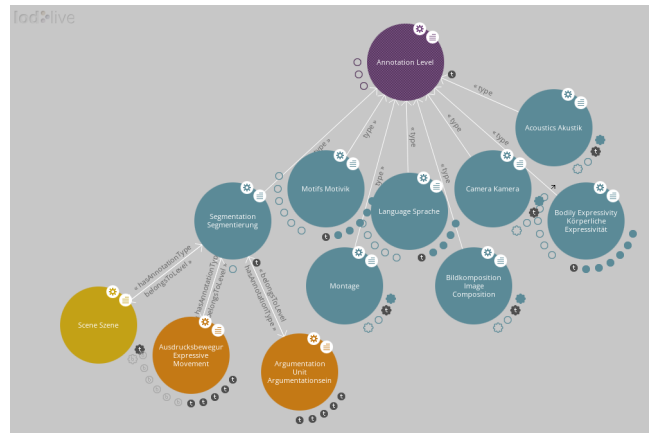


Figure 2: LODlive view of the Annotation Levels defined in the AdA filmontology

The mapping of ontology information into operational Advene concepts was relatively straightforward. The data structure of Advene is composed of `Annotation Types` and `Relation Types` that can be grouped into meaningful `Schemas`. The Annotation Type ontology class can then be directly mapped to Advene Annotation Type, while the Annotation Levels can be mapped into Advene Schemas.

The Advene OWL importer handles these definitions and maps them into appropriate data/metadata so that the interface can offer optimized behaviours, such as one-key shortcuts for quickly inputting pre-defined terms. The keyboard keys 1 to 9 in the Advene

<https://www.advene.org/>

timeline can be used to toggle the predefined keyword corresponding to the given number, based on their definition order.

In addition, some of the ontology derived constraints have been implemented in the Advene constraint checker framework. The extensible framework provides standard checkers such as an overlapping checker, that identifies annotations that may overlap temporally, or a duration checker that identifies annotations that have no duration, and may thus be hard to identify in some views. Using the additional information from the imported ontology, new checkers have been implemented in Advene, and can be activated on demand. For instance, if an annotation type is linked to predefined values, the checker indicates if some of its annotations contain terms that do are not defined in the ontology. Another checker validates the presence of URI metadata needed for RDF generation.

This non-enforcing checking approach is important for allowing users to experiment with new terms. Terms that are non-conforming with the ontology can easily be identified (through the checker) and be discussed for inclusion in following versions of the ontology.

4. WebAnnotation interoperability

WebAnnotation is a W3C defined standard for annotation representation and interoperability. It defines a data model, as well as a standard representation using JSON-LD syntax. Advene does not use WebAnnotation as a native data format, but provides interoperability through import and export filters.

Both import and export filters for WebAnnotation have been extended and made more robust during the AdA project, so that annotations produced with Advene could be integrated into a Triple Store. The export filter has been designed as a generic WebAnnotation filter, that can be used for any kind of data. This generic filter has then been extended with an AdA-specific export filter that can use the additional metadata and conventions defined in the project. For instance, the ontology import discussed above defines URIs for concepts and predefined values, that are preserved as metadata in Advene and used for exporting data with appropriate references. This two-level implementation is also meant to serve as an example for extending and customizing the WebAnnotation export filter for other domains/ontologies.

5. Feature extraction plugins

One of the goals of the AdA project was to experiment with automatic extraction of specific audiovisual features. Some of the data produced by the manual annotation process was used to train deep-learning algorithms, but the resulting models are not yet fully operational. However, the definition of the ontology lead to identify other relevant features, that can be extracted through less demanding algorithms.

Two feature extraction filters have been contributed by HPI to Advene during the AdA project. Figure 3 presents a visualisation of their output in the Advene timeline. The Motion Dynamics Extractor calculates an indicator related to the amplitude of movement in a video. It outputs a series of numeric values, which can be displayed as bargraphs in the timeline, thus providing a broad overview of the dynamic profile of a sequence.

The Dominant Color filter takes as input a sequence of annotations and extracts the dominant color shades for each annotation, that is to say the colors appearing most importantly in terms of number of pixels. As all automatically extracted features, the quantity-based approach has obvious shortcomings, such as not considering the perception of a bright colored small spot in a unified background. But it nevertheless provides interesting insights for color analyses. The manually annotated corpus produced during the project helps in evaluating the relevance of the algorithms.

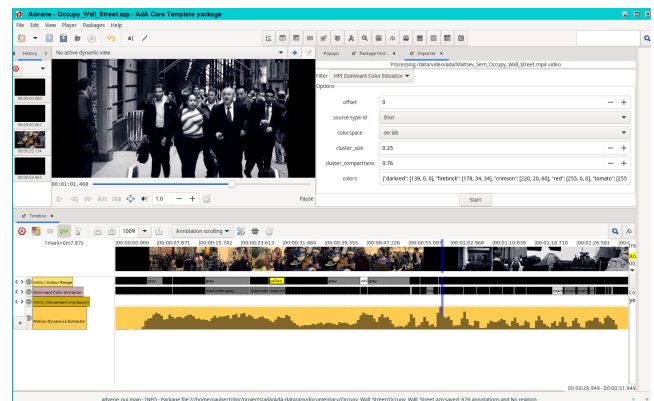


Figure 3: Visualisation of the information extracted by Dominant Color and Motion Dynamics filters - the first line in the timeline displays the human-produced data, the second line displays the dominant color detected by the filter. The third line presents a bargraph with the output of the Motion Dynamics filter.

6. Timeline visualisation

The timeline present in previous figure is the standard Advene timeline, which provides advanced and detailed interaction facilities. One of the issues identified during the AdA project is that it is not fit for the global visualisation of about 20 thousand annotations, which was typical for a feature-length movie in the video corpus.

One of the workarounds for this shortcoming was to use the Advene ability to split data according to some reference annotation (*Scene* for instance) which limited the number of annotations that were handled by a session and facilitated collaboration. A movie was then first annotated globally with *Scene* annotations, then split into multiple packages - one per scene - for edition with the whole AdA ontology. These *Scene*-specific packages could then be merged into a single one in a postprocessing phase. The process already existed and has been streamlined during the AdA project.

Another approach has been to design a new timeline visualisation interface, trading-off editing capabilities for the ability to display the tens-of-thousands annotations that were obtained after merging for a single movie. The definition of this new visualisation took place in an interdisciplinary dialogue between humanities scholars and computer science scholars [ASS21] to determine the appropriate representations as well as the trade-offs between convenience and development time.

The web-based timeline uses the underlying ontology to specify

appropriate representation for elements. It is customizable through a text-based specification language, in order to save developers time from having to develop an interface for the settings edition and instead focus on the data visualisation interface. It also gives the ability to easily copy/paste and share visualisation definitions, in order to keep them for future reference or to share them with other scholars.

Figure 4 presents an example of a produced timeline, generated from the same package as previous illustrations. It presents the shot duration as a bargraph, a soundwave representation of the sound levels, and different annotation types. Annotations can be filtered or highlighted based on their content.

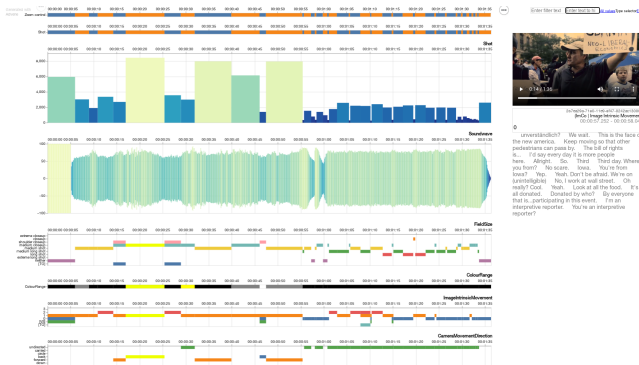


Figure 4: New timeline visualisation - web-based presentation of the data, dynamically generated from the Advene application. The video is displayed on the right, accompanied by a transcription based on a specified AnnotationType.

7. Conclusion

The AdA project has produced multiple valuable research outputs, such as a Film Ontology that builds upon previous ontology definitions, and an annotated corpus. The evolutions of the Advene software can also be considered as an output of the project, considering the integration of the instrument and the research process, as developed in [ASS21].

References

- [AP05] AUBERT O., PRIÉ Y.: Advene: active reading through hypertext. In *ACM Hypertext '05* (Salzburg, Austria, Sep 2005), pp. 235–244. [2](#)
- [ARAH18] AGT-RICKAUER H., AUBERT O., HENTSCHEL C., SACK H.: Authoring and publishing linked open film-analytical data. In *Proceedings of the Posters and Demos Track of the 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW 2018)* (Nancy, France, Nov 2018). [1](#)
- [ASS21] AUBERT O., SCHERER T., STRATIL J.: Instrumental genesis through interdisciplinary collaboration - reflections on the emergence of a visualisation framework for video annotation data. [3, 4](#)
- [BGSS20] BAKELS J.-H., GROTKOPP M., SCHERER T., STRATIL J.: Matching computational analysis and human experience. performative arts and the digital humanities. *Digital Humanities Quarterly* 14, 4 (2020). [1](#)
- [BSSAR20] BAKELS J.-H., SCHERER T., STRATIL J., AGT-RICKAUER H.: Ada filmontology – a machine-readable film analysis vocabulary for video annotation. In *Digital Humanities Conference 2020 (DH2020)* (Ottawa, Canada, July 2020). [2](#)
- [Kip01] KIPP M.: ANVIL - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of Eurospeech 2001* (Aalborg, Sep 2001), pp. 1367–1370. [2](#)
- [Tho02] THORN C.: Creating New Histories of Learning for Math and Science Instruction: Using NVivo and Transana to manage and study large multimedia datasets. In *Conference on Strategies in Qualitative Research* (2002), Institute of Education, University of London. [2](#)
- [WBR*06] WITTENBURG P., BRUGMAN H., RUSSEL A., KLASSMANN A., SLOETJES H.: Elan: a professional framework for multimodality research. In *Proceedings of Language Resources and Evaluation Conference (LREC)* (2006). [2](#)