An Ontology Web Application-based Annotation Tool for Intangible Culture Heritage Dance Videos

Sylvain Lagrue, Nathalie Chetcuti-Sperandio, Fabien Delorme, Chau Ma Thi, Duyen Ngo Thi, Karim Tabia, Salem Benferhat

An Ontology Web Application-based Annotation Tool for Intangible Culture Heritage Dance Videos

Sylvain Lagrue
Heudiasyc CNRS UMR 7253
Université de Technologie de Compiègne
sylvain.lagrue@hds.utc.fr

Nathalie Chetcuti-Sperandio
CRIL CNRS UMR 8188
Université d'Artois
chetcuti@cril.fr

Fabien Delorme
CRIL CNRS - UMR 8188
delorme@cril.fr

Chau Ma Thi
University of Engineering and Technology
Vietnam National University Hanoi
chaumt@vnu.edu.vn

Duyen Ngo Thi
University of Engineering and Technology
Vietnam National University Hanoi
duyennt@vnu.edu.vn

Karim Tabia
CRIL CNRS - UMR 8188
Université d'Artois
tabia@cril.fr

Salem Benferhat
CRIL CNRS - UMR 8188
Université d'Artois
benferhat@cril.fr

ABSTRACT

Collecting dance videos, preserving and promoting them after enriching the collected data has been significant actions in preserving Intangible culture heritage in South-East Asia. Whereas techniques for the conceptual modeling of the expressive semantics of dance videos are very complex, they are crucial to exploit effectively the video semantics. This paper proposes an ontology web-based dance video annotation system for representing the semantics of dance videos at different granularity levels. Especially, the system incorporates both syntactic and semantic features of pre-built dance ontology system in order to not only use the available semantic web system but also to create unity for users when annotating videos to minimize conflicts.

CCS CONCEPTS

• **Information systems** → *Ontologies*; *Multimedia and multimodal retrieval*; Information retrieval query processing; • **Applied computing** → *Annotation*;

KEYWORDS

Video annotation; Ontologies; Knowledge Representation; Inconsistency-tolerant query answering.

ACM Reference Format:
Sylvain Lagrue, Nathalie Chetcuti-Sperandio, Fabien Delorme, Chau Ma Thi, Duyen Ngo Thi, Karim Tabia, and Salem Benferhat. 2019. An Ontology Web Application-based Annotation Tool for Intangible Culture Heritage Dance Videos. In *1st Workshop on Structuring and Understanding of Multimedia heritAge Contents (SUMAC '19), October 21, 2019, Nice, France.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3347317.3357245

SUMAC '19, October 21, 2019, Nice, France

ACM ISBN 978-1-4503-6910-7/19/10…$15.00
https://doi.org/10.1145/3347317.3357245

1 INTRODUCTION

Dance is a popular art of intangible culture which contains many historical and cultural features of the people performing it. In modern ages, dance resources are numerous and exist in many different forms: visual, audio and text. Among different forms of dance data, dance video is a valuable source, because it conveys information through many elements such as music, lyrics, clothing and especially dance movements that are well described in video data sources. In order to exploit the contents of dances and to recognize relevant information in dance, as well as to organize dance data, we need to build a system of semantics for dance effectively. Ontology, a data model representing a field and used to reason about objects in that area and the relationships between them, is a good choice for semantic dance representing.

In this paper, we aim to describe a dance video annotation system in order to enrich dance videos with annotation. The system is a web-based annotation tool connected to ontologies: the tool uses concepts of the ontologies to support users to annotate dance videos. The tool also collects information about users' confidence when performing annotation. This information could later be used for solving conflicts appearing in annotated contents from different users for a same video.

The paper is organized as follows. Section 2 reviews some related works. The overall system is presented in Section 3. Sections 4 and Section 5 describe different aspects related to the system.

2 RELATED WORKS

In 2000, Michael Kipp developed ANVIL [4] based on research in the field of gestures. ANVIL is a free video annotation tool that supports multi-layer annotations based on user-defined encoding schemes. ANVIL also supports many of the speech-based annotation features. In addition, ANVIL 5 (2010) supports annotation on 3D motion capture data in BVH format.

In 2005, Olivier Aubert and Yannick Prié introduced Advene [1], a project which provides a model and a format to share annotations about digital video documents. A video annotation tool, allowing annotating any video format read from DVD or stream, is included in the project. In Advene, annotated content can be displayed in text or on video. Annotated results being packed in packages can be saved, shared, stored on the server.

In [8], the paper presents the Dance Video Semantic Model (DVSM) to model dance video objects at multiple levels of detail. This taxonomy of multiple levels is determined by the components of the accompanying song. The model is built thanks to the close combination of the meaning of the song and movements. Besides, the author introduced Agent, used to express the spatial actions of dancers.

In [7], the authors have proposed an ontology-based web interface that allows the user to annotate classical ballet videos, with a hierarchical domain-specific vocabulary and provided an archival system for videos of dance. This work supports the search and browsing of the multimedia content using metadata (title, dancer features, etc.), and also implements the functionality of "searching by movement concepts", i.e. filtering the videos that are associated with particular required terms of the vocabulary, based on previous submitted annotations.

In [3], the authors investigated the accuracy of human dance motion capture and classification from selected Malaysian dances for Malaysian Dance Annotation (MDA). The goal of the study is to classify complex motion such as a dance rather than simpler motion such as walking or waving hand. The authors have applied classification algorithms on motion captured data aiming at automatically labeling specific segments of Malaisian folk dance sequences. The results of evaluations showed that despite the complex movements in dance, the proposed solution requires less human input effort and generates the proper accuracy for complex dance motion notes. [5] have applied a similar approach to Vietnamese folk dance.

In [6], a web-based annotation tool has been proposed. The tool is integrated in an archival system - the WhoLoDancE movement library (WML), including also other functionalities such as browsing and searching using metadata and annotations and personalization features. The movement library (WML) consists of a web-based interface, data, metadata – including title, genre, annotations, performer, dance company and date of recording, annotation management back-end as well as a user-management system. The annotation tool, embedded into the WML, enables manual annotation of performances with free text and controlled vocabularies based on the conceptual framework of WhoLoDancE, through a tabular and a timeline view. This tool is used to collect annotation contents from dance experts. To accomplish the goal of collecting a high amount of annotations (which will be used for the training of algorithms able to automatically describe dance performances) from a large community by making the annotation procedure sensibly lighter and easier, the authors proposed a movement quality annotation by comparison (MQA) tool. Given a movement quality, say fluidity, this tool displays a 3D representation of two short dance movements represented by a black and a white avatar, in a loop, and asks the user to make a comparison between them and to select the one which is expressing a higher level of fluidity. Users can choose between five levels, or even skip the comparison if they decide that the comparison is not meaningful or if they do not feel confident enough to make a decision.

These tools have been well known and widely used, however, they have not taken advantage and connected to the semantic network of the same field. In our research, the annotation tool is connected to an available dance ontology of which users could unify the terms and contents when annotating videos. Furthermore, we are also aiming to using annotated videos to enrich the dance ontology as well as to supply data for a query tool.

3 OUR ANNOTATION SYSTEM

We propose an overall system (Figure 1) that contributes to preserving dance art and in which the annotation tool is the part being useful for the community of dance field experts. This tool helps us to exploit dance video data more effectively and efficiently. To do so, we designed an ontology web-based dance annotation tool, where, in addition to dance video data, an available dance ontology is the essential foundation of the tool. Dance is a form of movement that is very different from other movements such as ordinary movements in daily life or sports. Due to the conventional rules of postures and movements, dance performs the duty of conveying informative content. Moreover, dance is influenced by history and culture. So, the available ontology is a source of dance-related concepts that orients and guides users when they annotate dance videos. The annotation tool is a web-based one, which helps a large number of users employing the tool simultaneously. In addition, using the tool is very flexible: the users can exploit the tool anywhere, any time and on any computer without depending on the hardware and the operating system of the computer.

The tool consists of three main functional parts (Figure 3): (i) functions for connecting to the dance ontology and displaying concepts from the ontology; (ii) functions for video processing; (iii) functions for collecting users' confidence.

The system output is in the form of web standard format, WebVTT format file containing a lot of information but being very simple to use. First of all, the information in the output file is directly used as subtitles for video presentation in web browsers. This is the most widely used mean of information transmission to the public in accordance with the goal of promoting dances. Secondly, the tool outputs can be exploited for updating and completing the dance ontology with assertions (Abox). In addition to WebVTT format, the outputs are parsed into OWL format and added to the ontologies to be queried by users using a querying tool (another tool of our traditional dance video tools suite). Besides, the system outputs are used as a source for building machine learning training datasets for our automatic annotation tool. Indeed, our aim is to annotate as many videos as possible, hence the use of machine learning for some dance videos.

4 TRADITIONAL DANCE ANNOTATION TOOL

We have built a semi-automatic annotation tool (called TDAT - traditional dance annotation tool) to assist dance experts (called users since now) in annotating videos based on pre-built ontologies. TDAT is now already online and used for annotating videos with different dance contents in a project of preserving and promoting

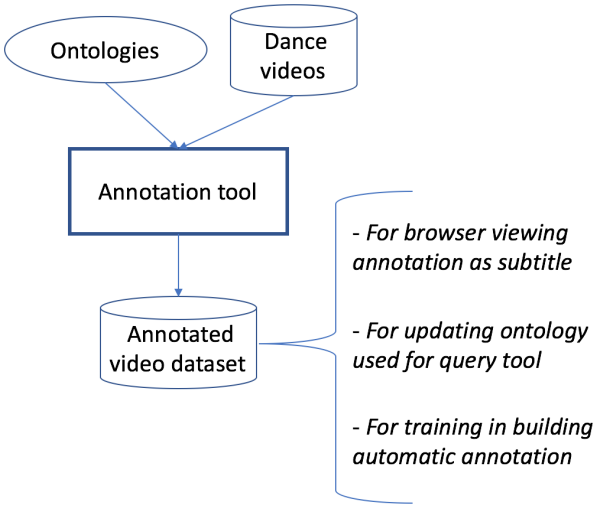

Figure 1: Our annotation system.

intangible culture. The interface of TDAT is illustrated in Figure 2. Elements of this interface will be described later, corresponding to the functional parts of the tool.

4.1 TDAT's input and output

The TDAT's input includes a (any) video that can be viewed through an Internet browser video player, and an ontology containing a concept taxonomy corresponding to the dance in the video. The output is the annotation content for the input video in WebVTT format. WebVTT (Web Video Text Tracks) is a W3C standard for displaying timed text tracks in connection with the HTML <track> element. This is a text based format, which must be encoded using UTF-8. The primary purpose of WebVTT files is to add text overlays to a video. These files provide captions or subtitles for video content, and also text video descriptions, chapters for content navigation, and more generally any form of metadata that is time-aligned with audio or video content. Therefore, the WebVTT format is very suitable for representing annotation content for dance videos, as well as facilitating further operations such as information processing and querying.

4.2 TDAT's functional parts

As mentioned in section 3, there are three main functional parts (Figure 3) in TDAT: (i) functions for connecting to the ontology and displaying concepts from the ontology; (ii) functions for video processing; (iii) functions for collecting users' confidence. The meaning and behavior of each part, corresponding to the elements on the interface of the tool, will be described in detail below.

- (i) Concept displaying functions

In the first part, TDAT connects to the selected input ontology through queries to retrieve the taxonomy of dance related concepts. These concepts are then organized into a menu of options which users can use by selecting them when annotating video. The section surrounded by the dashed line with the symbol *(i)* in Figure 2 illustrates this menu on the overall interface of the tool. A more detailed example of this menu with specific concepts (shown when the user clicks on the menu) is shown in the section with a dashed line in Figure 4. Attributes and descriptions of the concepts will be displayed when videos are being played if the user requests them. The information obtained from the ontology (dance related concepts organized into a menu) will effectively support users during the video annotating phase in the second part of the tool.

- (ii) Video processing functions

In the second part of TDAT, users can perform tasks to annotate dance videos based on the information that the tool has retrieved from the ontology before.

Firstly, the tool allows users to upload unannotated videos to the system (by clicking on the button *(ii-1)* in Figure 2), to select one video from a list of videos (by clicking on the menu *(ii-2)* in Figure 2), to play and to view the selected video (by manipulating buttons in the region *(ii-3)* in Figure 2), and to annotate the selected video.

To perform annotation for a selected video, users mark video segments (called clips) by selecting the start and the end frames for each segment of video being annotated. Each clip is marked by clicking on the *start*, *end*, and then *insert* buttons in the section *(ii-4)* in Figure 2. When users select the start and the end frames for each clip by clicking the *start* and the *end* buttons, the time points corresponding to the start and the end frame of the clip will be displayed in the section *(ii-5)* in Figure 2. When a clip is completely marked, it will be visualized by a line segment in the section *(ii-6)* in Figure 2. In this section, each line corresponds to one level in which the video is annotated; the first line goes with the first level, followed by (optional) lines representing the next levels. Each line can consist of one or more line segments, each representing a clip that has been marked at that level. Users can annotate videos at any level; and a clip in any level could be segmented into smaller clips in the next level. When users mark a clip in a video segment that has been marked in one or several levels, the clip being marked will belong to the next level. The levels of the video segmentation, closely linked to the content of each dance, could be suggested as in Table 1.

To perform annotation for the selected video, users can mark all clips at all levels, then make annotation for each clip. Or users can also mark one clip, make annotation for this clip, and then continue to do the same for other clips. To annotate one clip, users first select the clip (by manipulating the slider in the section *(ii-3)* in Figure 2 to select the time point) then switch to the first function group to select the recommended levels and optional concepts from the input ontology. When the time point is selected, the clips at that point will be displayed on the right side of the interface of the tool; each clip has information about the level to which it belongs, and information of start, end time displayed on its right side (section *(ii-7)* in Figure 2). To add annotation content for one clip, users choose one concept from the menu containing the information retrieved from the input ontology (section *(i)*), and then (optionally) enter the value for this concept. The newly added annotation information will be displayed at the top of the menu (section *(ii-8)*). Users can repeat the above annotating operations, resulting in annotation for the video at different levels. Since annotation can be a time-consuming and

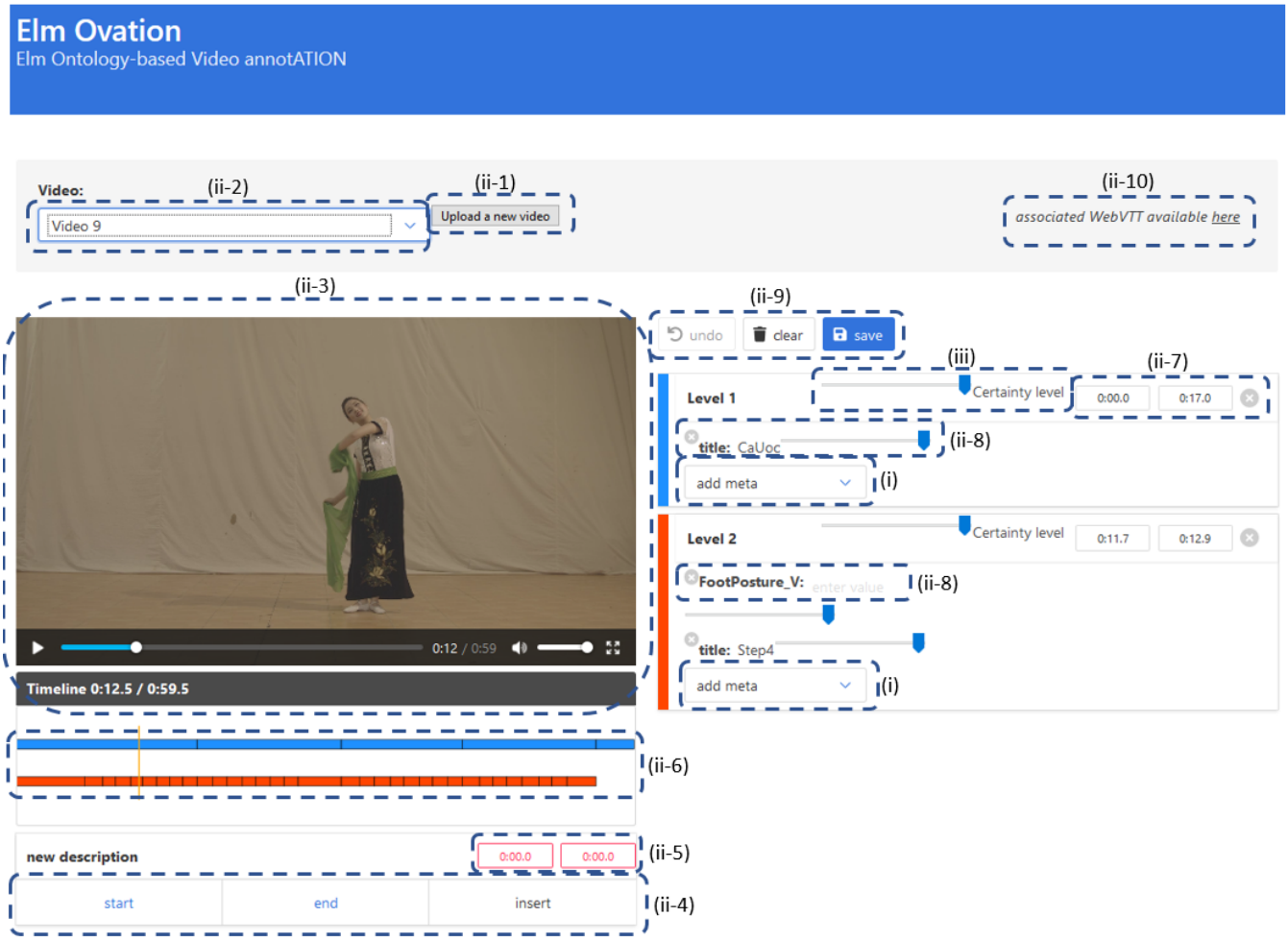

Figure 2: The interface of the annotation tool.

complex process, the system allows users to revisit the annotations at a later time, to edit and refine them. The annotation content for the video is saved in WebVTT format; users can get this WebVTT file in the section *(ii-10)*. Figure 5 shows an example of annotation in WebVTT format.

- (iii) Confidence collecting function

In the third part, TDAT collects information about users' confidence when performing annotation. In the former part, when annotating each clip, users provide information about their confidence in their annotated contents (section *(iii)* in Figure 2). The third part of the tool will collect this information so that it could later be used for solving conflicts appearing in annotated contents from different users for the same video. The tool allows many users to annotate a same video, so conflicts may appear in annotated contents from different users for the same video because of their different backgrounds, knowledge and perspectives. Users' confidence information collected by the tool will support conflict resolution in the step of querying information from the ontologies later.

In summary, after a user uses the tool to annotate a selected video, the video will be divided into clips; annotated information for each clip is described by a tuple (video, user, start-end-times, concepts, confidence-degree). In this tuple, "start-end-times" represents the start frame (start time) and the end frame (end time) of the clip, "concepts" shows concepts from the input ontology and their corresponding accompanying values, "confidence-degree" corresponds to the level of confidence of the user about the information he/she annotates. The user can annotate a video at different levels and each clip in one level can be divided into several clips at the next level.

5 CASE STUDY: VIETNAMESE DANCE ONTOLOGY

An important input of the annotation tool is a dance concept taxonomy in a dance ontology. Illustrating this part, we introduce a case study: Vietnamese folk dance ontology.

When building Vietnamese folk dance ontology, we have considered two main facts. Firstly, aiming to understand dances, we need

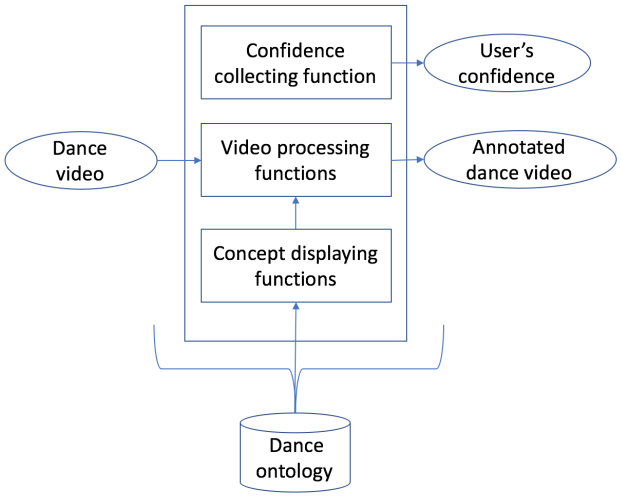

Figure 3: Annotation tool.

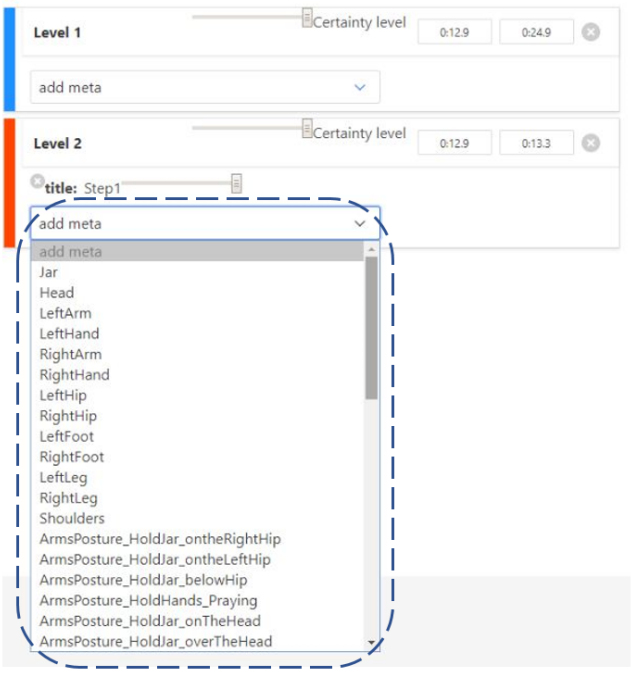

Figure 4: Concepts from the ontology

Table 1: Description for different levels.

1	Dance	Name of dance title, start-date, total duration, location, theater, musicians, song name, Basic unit, category, ethnics
2	Basic unit	Name of basic unit title, duration, dancer, movement phrase, orientations, dancer ID
3	Movement phrase	Name of movement phrase title, place, characters, characters' names, description, keywords, music theme
4	Dancers/ Groups of Dancers	Name of dancer/ Name of the group dancers, dancers' names, characters' names, event, groups, groups' names

```
00:00:00.000 --> 00:00:02.115, confedence: 1,
{
   "isInRelation": "A/B/C/D;"behind each other", confedence: 1,
   "isInShape": "A/B/C/D;"line", confedence: 1,
   "perform": "Dancers;Preparing", confedence: 0.8,
   "isInOrientation": "A;"Orient6", confedence: 1,
   "isInOrientation": "B;"Orient8", confedence: 1,
   "isInOrientation": "C;"Orient2", confedence: 1,
   "isInOrientation": "D2;"Orient4", confedence: 0.7,
   "isInOrientation": "D3;"Orient4", confedence: 1,
   "isInOrientation": "D1;"Orient2", confedence: 1,
   "hasPosition": "A; "right front" , confedence: 1,
   "hasPosition": "B; "right back" , confedence: 1,
   "hasPosition": "C; "left back" , confedence: 1,
   "hasPosition": "D2; "left front", confedence: 1,
   "hasPosition": "D3; "left front" , confedence: 1,
   "hasPosition": "D1; "center", confedence: 1,
   "Title": "Step1", confedence: 1
}

00:00:02.115 --> 00:00:10.080, confedence: 1,
{
   "perform": "D1;"FreeDancing", confedence: 0.9,
   "Title": "Step2", confedence: 1,
}

00:00:10.080 --> 00:00:14.500, confedence: 1,
{
   "isInRelation": "A;"behind each other", confedence: 1,
   "isInShape": "A;"line", confedence: 1,
   "perform": "D1/A;"MoInviting", confedence: 1,
   "perform": "D1;"Trans_05", confedence: 0.8,
   "perform": "A;"Trans_01", confedence: 1,
   "hasPosition": "center", confedence: 0.5,
   "Title": "Step3", confedence: 1,
}
```

Figure 5: An example of annotation in WebVTT format

to have background knowledge of regional culture and history. The background knowledge is an effective means to approach overall topics of a specific dance. Secondly, the dance imparts the message to the audience through movements of dancers; so we need to analyze the movements to understand the detailed messages of each dance.

For specific dances, among factors including dancers, dancers' movements, stage, music, singing (optional), costumes, and accompanying things (optional), the movements of the dancer are the most important one. When analyzing the movements we approached in two aspects. Firstly, the movements could be divided according to the human body structure; secondly, the movement could be divided by dance content. In the first aspect, there are two main

types of movements; the movements of the whole body in the stage space and the movements of body parts. The stage is divided into different areas where dancers move in a certain direction and trajectory. When moving, the dancers make movements of different parts of the body; this movement only includes rotation, being analyzed according to the structure and degrees of freedom of body joints. In the second aspect, a dance which is a story could be divided into different steps which are different scenes of the whole story. With the combination of the above movements and the dance story, we use two concepts related to dance: movement phase and basic unit [2]. "Each basic unit is defined as the smallest movement with a complete meaning" and is specifically described in the ontology. "Each movement phase is a simple body movement, which has a trajectory in the form of a line: an acre, a dot on a plane as a straight pathway, a curved". All the facts analyzed above have been included in the ontology.

- Concepts, attributes, descriptions:

After analyzing Vietnamese history and culture, and studying motions related to the specific movements, we have developed a taxonomy of concepts related to Vietnamese folk dance (figure 6), then we have defined attributes and have performed descriptions. This part of our proposal provides concepts and descriptions used in the dance video annotation part described in the former section.

The dances that are analyzed in detail, following the main topics category or the movement analysis, are related to the time element, so in the ontology we consider the concept of time. This concept encodes the content of dance and the movement of the dancer being changed by the time. The next concept group is about region and message. The reality of the dance description will be the attributes of this group of concepts. These concepts represent regional and content information of the dance. Last but not least are concepts of movements and their shapes. Motion analysis gives the detailed descriptive attributes of these concepts which encode the information about phase movements, primitives, poses and the information of dancer groups in the dance.

Relationships among the concepts make different types predicate groups. We have a predicate group which relates to performing different poses and movements. For instance, $StartWith(a, b)$ means PhaseMovement a starts with Pose b. The spatial relationships with the stage of dancers and dancer groups are in a location predicate group such as $Position1(a)$ which describes that Dancer/Dancer Group a is in the center of the stage. The third predicate group concerns dancer groups. For example, $MemberOf(a, b)$ means Dancer a is member of Group b. The fourth predicate group deals with the relationship between the content of folk dances and their regions. $BelongToRegion(a, b)$, for example, indicates that Dance a/BasicUnit a belongs to Region b.

Descriptions of the concepts, especially details of phrase movements and basic units, are added to the ontology as metadata. The metadata is really important in case we need to describe annotated data in detail. Some typical dances are included in the ontology.

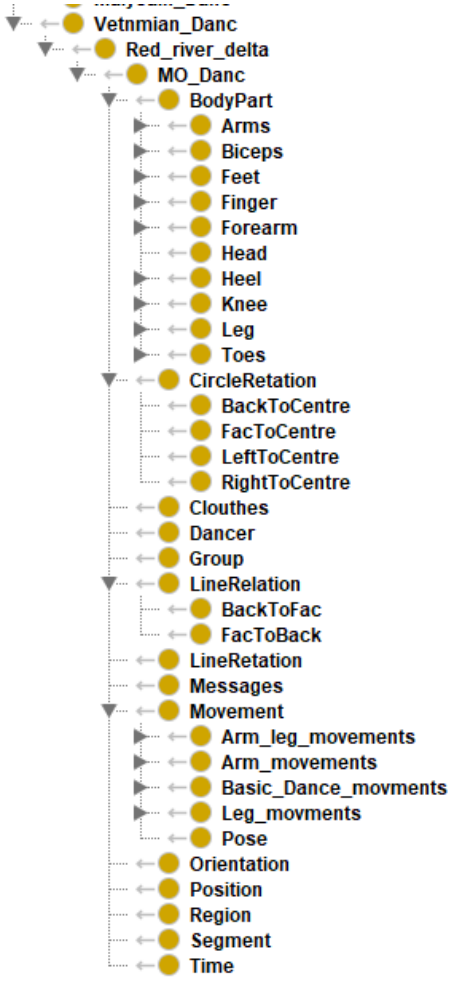

Figure 6: Taxonomy of Vietnamese dance concepts .

6 DISCUSSION AND CONCLUSION

In summary, we built a system in which the annotation tool is well connected to the available dance ontology to query dance concepts in order to reduce the conflict concepts between different users. The system allows users to annotate videos related to dancers' movement with different levels of detail. Besides, the system supports handling conflicts between different users in query part by collecting users' confidence. Currently the system has been applied to annotations for videos with different dance contents in the project of preserving and promoting intangible culture. In the future, the system is expected to be deployed in museums or art teaching schools to support museum staffs as well as art teachers in annotating their materials. In addition to using the ontology for annotating video content, we also aim to use the annotated content itself, described in these annotations, to update the metadata for ABox and TBox of the ontology.

REFERENCES

[1] Olivier Aubert and Yannick Prié. 2005. Advene: Active Reading through Hypervideo. In *Proceedings of ACM Hypertext'05*. 235–244.

[2] Ma Thi Chau and Nguyen Thanh Thuy. 2018. A labanatation based ontology for representing Vietnamese folk dances. In *In Proceedings of International Conference on Digital Arts, Media and Technology (ICDAMT)*.

[3] Huma Chaudhry, Karim Tabia, Shafry Abdul Rahim, and Salem BenFerhat. 2017. Automatic annotation of traditional dance data using motion features. 254–258. https://doi.org/10.1109/ICDAMT.2017.7904972

[4] M. Kipp. 2014. ANVIL: A Universal Video Research Tool.. In *In: J. Durand, U. Gut, G. Kristofferson (Eds.) Handbook of Corpus Phonology, Oxford University Press, Chapter 21,*. 420–436.

[5] Chau Ma-Thi, Karim Tabia, Sylvain Lagrue, Ha Le-Thanh, Duy Bui-The, and Thuy Nguyen-Thanh. 2017. Annotating Movement Phrases in Vietnamese Folk Dance Videos. In *Advances in Artificial Intelligence: From Theory to Practice*, Salem Benferhat, Karim Tabia, and Moonis Ali (Eds.). Springer International Publishing, 3–11.

[6] Katerina El Raheb, Aristotelis Kasomoulis, Akrivi Katifori, Marianna Rezkalla, and Yannis Ioannidis. 2018. A Web-based system for annotation of dance multimodal recordings by dance practitioners and experts. In *In Proceedings of the 5th International Conference on Movement and Computing*.

[7] Katerina El Raheb, Nicolas Papapetrou, Akrivi Katifori, and Yannis E. Ioannidis. 2016. BalOnSe: Ballet Ontology for Annotating and Searching Video performances. In *MOCO*.

[8] Balakrishnan Ramadoss and Kannan Rajkumar. 2007. Modeling and annotating the expressive semantics of dance videos. In *International Journal "Information Technologies and Knowledge" Vol.1.*