An ontology web-application-based annotation tool for intangible culture heritage dance videos

ABSTRACT

Collecting dance videos and 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 modelling 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.

KEYWORDS

Video annotation, Ontologies, Knowledge Representation, conflict resolve.

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 [Kipp 2014] 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) ANVIL supports annotation on 3D motion capture data in BVH format.

In 2005, Olivier Aubert and Yannick Prié introduced Advene [Aubert and Prié 2005], a project which provide a model and 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 [Ramadoss and Rajkumar 2007], the paper presents the Dance Video Semantic Model (DVSM) to model dance video objects at multiple levels of detail. This taxonomy of multiple level is determined by the components of the accompanying song. The model is built thanks to the close combination of the song's meaning and movements. Besides, the author introduced Agent, used to express the spatial actions of dancers.

In [Raheb et al. 2016], 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 featured, 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.

The authors, in [Chaudhry et al. 2017], 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 not very complex 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 evaluation results showed that despite

the complex movements in dance, the proposed solution requires less human input effort and the proper accuracy for complex dance motion notes. [Ma-Thi et al. 2017] have applied an approach similar to the one in [Chaudhry et al. 2017] for Vietnamese folk dance.

In [Raheb et al. 2018], a web-based annotation tool have 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 WhoLo-DancE, through a tabular and a timeline view. This tool is used to collect annotation contents from dance experts. To accomplish the goal to collect 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 select which one is expressing a higher level of fluidity. Users can choose between five levels, or even skip the comparison if they decide the comparison is not meaningful or they do not feel confident enough to make a decision.

These tools have been well known and used widely, however, 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 that users could unify the terms and contents when annotating videos. Furthermore, we are also aiming 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 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 that, we designed a web, dance ontology-based annotation tool, whereas, 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 normal movements in daily life or sport movements. Due to the conventional rules of postures and movements, dance performs the duty of conveying informative content. Not only that, dance is influenced by history and culture. So, the available ontology

is a source of dance-related concepts that orients and guide users when they annotate dance videos. The annotation tool is a web-based one, which helps a large number of users approaching the tool simultaneously. In addition, the users are very flexible when using the tool. The users exploit the tool anywhere, any time, and on any computer without depending on the hardware and the computer's operating system.

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 confidence of users.

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 means of information transmission to the public in accordance with the goal of promoting dances. Secondly, the tool output can be exploited as updating data into the Tbox and Abox of the dance ontology which is the content and foundation structure for a query tool. Besides, the system output is a well-annotated data source used in machine learning as training to create automatic annotation tools.
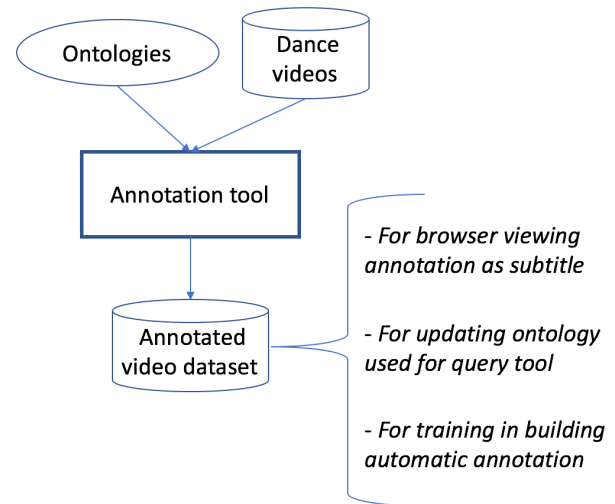

Figure 1: Our annotation system.

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. 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.
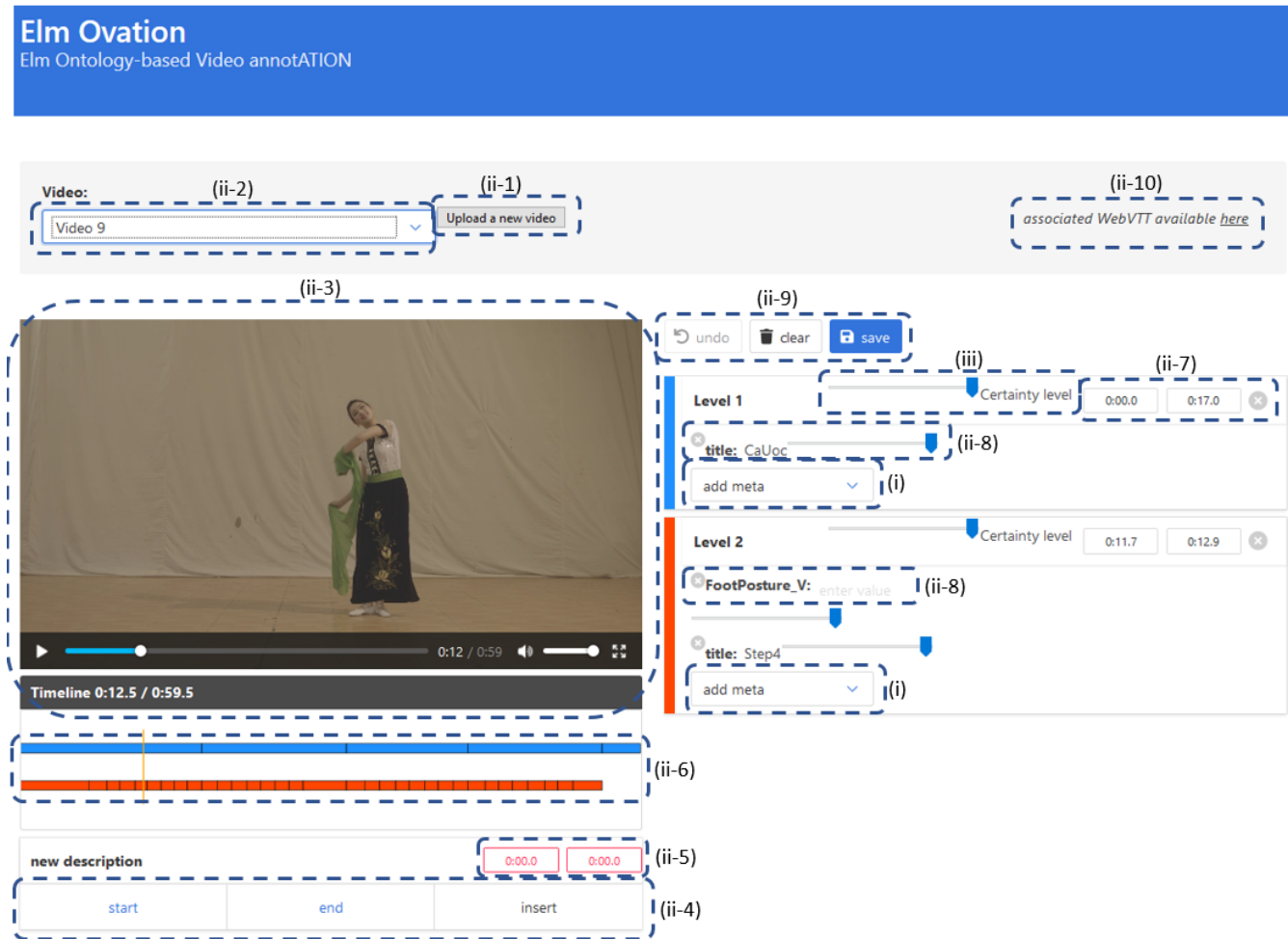
Figure 2: The interface of the annotation tool.

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 represent 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 confidence of users. The meaning and behavior of each part, corresponding to the elements on the tool's interface, will be described in detail below.

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 could 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 users clicks on the menu) is showed in the section with a dashed line in Figure 4. Attributes and descriptions of the concepts will be
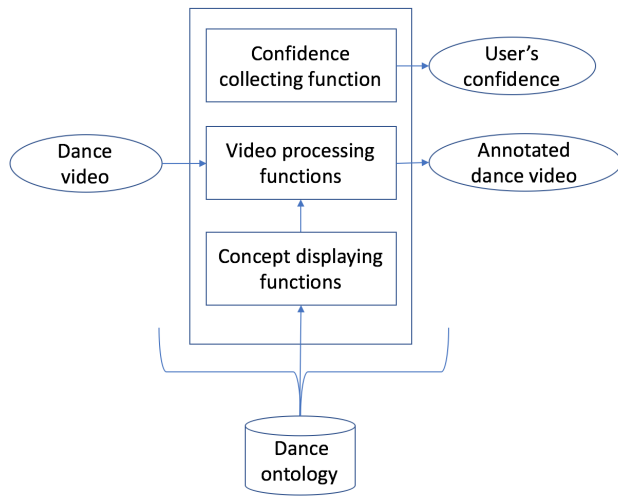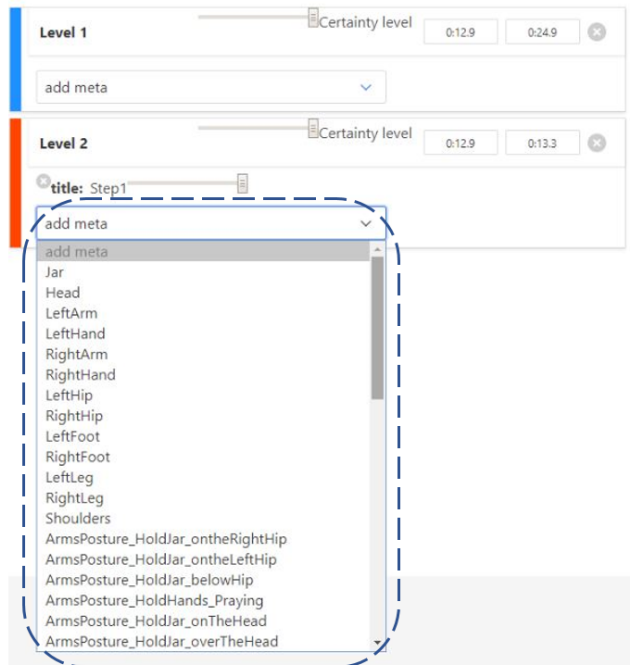
Figure 3: Annotation tool.

Figure 4: Concepts from the ontology

displayed when videos 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.

Video processing functions

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

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

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

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 tool's interface; 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 complex process, the system allows users to re-visit 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 by manipulating in the section (ii-10).

Confidence collecting function

In the third part, TDAT collects information about users' confidence when perform annotation. In the second 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 resolving conflicts appearing in annotated content from different users for one video. The tool allows many users to annotate a same video, so conflicts may appear in annotated content from different users for one video because of their different backgrounds, knowledge, and perspectives. User 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 including: start frame (start time), end frame (end time), concepts from the input ontology and their corresponding accompanying values, the level of confidence of the user about the information he/she annotates. The user can annotate the video at different levels, 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 to have background knowledge of regional culture and history. The background knowledge is an effective means to approach overall topics of a specific dance.

Figure 5: Chằm thau trống đồng (left), Xòe khăn (middle), Hầu đồng (right).

Secondly, the dance imparts the message to the audience through movements of dancers; so we analyze the movements to understand the detailed messages of each dance.

Vietnam is also known as an agricultural country with wet rice cultivation associated with two large deltas of the Red River in the north and the Mekong River in the south. Moreover, the Vietnam terrain has many rivers, lakes, and sea, being convenient for seafood aquaculture and fishing. As a result, there are dances that convey daily river activities of people such as cultivating, fishing, and sailing. Another aspect not to be missed is that Vietnam is a country with 54 ethnic groups, including diverse festivals, beliefs, costumes and languages. Those are also showed in dances. With such historical and cultural factors, traditional Vietnamese dances are categorized into three main topics: daily life activities, festival activities and spirituality [6], [7]. Figure 2 shows images of 3 typical dances: "Chằm thau trống đồng" (Mường ethnic) is a dance in which Mường people honour bronze drum. "Xòe khăn"(Thái ethnic) usually appears in the festival of Thái people."Hầu đồng" (Việt ethnic) is spirituality dance of Việt people that is the mean of communicating with the gods through the maternal grandparents. It is believed that the gods can incarnate the soul into the body of the maternal grandparents in the spirit of sublimation, ecstatic in order to propagate, eliminate evil spirits, heal, and bless for people.

For specific dances, among factors including dancers, dancers' movements, stage, music, singing (optional), costumes, and accompanying things (optional), the movements of the dancer is 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 [Chau and Thuy 2018]. "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
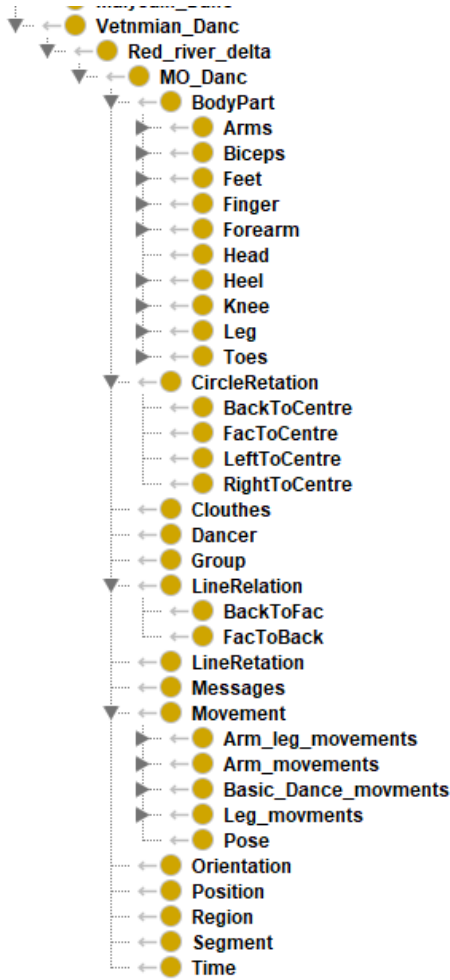
Figure 6: Taxonomy of Vietnamese dance concepts .

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 of Vietnamese history and culture, studying motions related to the specific movements, we have developed taxonomy of concepts related to Vietnamese folk dance (figure 6), then we have defined attributes and have performed descriptions. This proposal's part provides concepts and descriptions for the dance video annotation part in the next 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 time concept. 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 are 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 instant, $StartWith(a, b)$ means PhaseMovement a starts with Pose b. The spatial relationship with the stage of dancers and dancer groups are in a location predicate group such as, $Position1(a)$ describes that Dancer /Dancer Group a is in center of the state. 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 its region. $BelongToRegion(a, b)$, for example, indicates that Dance a/BasicUnit a belongs to Region b.

Descriptions of the concepts, especially detail 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. For instance, in "Múa mõ dưới trăng" dance (Việt ethnic), there are some typical basic units such as, *MInviting*, *MFootDragging*, *MPositionExchanging* etc... "Start with *legpose*1, arms away 3 cm in front of torso, bicep and forearm are orthogonal . Then, *repeat*: right/left legs straight ahead with heel touching the ground, right/left arms also straight ahead along the orientation of the leg with opened hand. Body can rotate any orient, backward or forward". is the description of *MInviting*. Some typical dances such as "Chằm thau trống đồng", "Xòe khăn" , "Hầu đồng" ,"Múa mõ dưới trăng" are included in the ontology.

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 expanded into 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.

ACKNOWLEDGMENTS

REFERENCES

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

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).

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

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.

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.

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.

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

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