An Ontology-based Approach for Building and Querying ICH Video Datasets

Sihem Belabbes, Yacine Izza, Nizar Mhadhbi, Tri-Thuc Vo, Karim Tabia, Salem Benferhat

An ontology-based approach for building and querying ICH video datasets

Sihem Belabbes[1], Yacine Izza[1], Nizar Mhadhbi[1], Tri-Thuc Vo[2], Karim Tabia[1], and Salem Benferhat[1]

[1]*CRIL Université d'Artois & CNRS, Lens, France*
[2]*CICT, Can Tho University, Vietnam*

Abstract

The diversity of Southeast Asian Intangible Cultural Heritage (ICH) is showcased in many art forms and notably in traditional dances. We focus on the preservation of Vietnamese ICH by building an ontology for *Tamia đwa buk* dances. We propose a completion of the ontology by semantically enriching traditional dance videos through manual annotation. Once annotated video datasets are built, we propose strategies for processing user queries. In particular, we address inconsistencies which emerge when the same video receives conflicting annotations from multiple sources. We also take into account different reliability levels of the sources in order to prioritize query answers.

1 INTRODUCTION

This study is conducted in the context of a European research project called AniAge[1], that focuses on the digital preservation of Southeast Asian Intangible Cultural Heritage (ICH), using high dimensional heterogeneous data-based animation techniques.

Indeed, Southeast Asia is known for its abundance of natural resources as well as for its rich ethnic and cultural diversity. Vietnam is no exception with its ancient history and its 54 ethnic groups [Luyên N. T., 2013]. Traditional dances play a crucial role in spiritual and cultural life [Sakaya, 2010]. They are one of the living art forms that the UNESCO[2] organization considers as ICH which should be preserved.

Ontologies play a key role in ICH digital preservation. In this paper, we build an OWL ontology for a Vietnamese traditional dance called *Tamia đwa buk* (a.k.a. Jar Dance), which is performed by female dancers holding jars. The ontology then serves as a backbone for semantic enrichment of Jar Dance videos, through a process of manual annotation. Indeed, domain experts annotate dance videos by providing descriptions of the cultural content conveyed in a Jar Dance. This is expressed by elements such as dancer postures and movements, costumes, jar position, as well as the symbolism portrayed in the dance.

In the field of knowledge representation and reasoning within AI, ontologies are formalised with languages based on Description Logics [Baader et al., 2007], such as DL-Lite [Calvanese et al., 2007], a lightweight fragment known for its expressive power and

[1]`http://www.cril.univ-artois.fr/aniage/`
[2]`https://ich.unesco.org/en/lists`

good computational properties. The ontology is translated into an axiomatic knowledge base (a terminological box, a.k.a. TBox), whereas annotations are translated into a set of ground facts (an assertional box, a.k.a. ABox). The TBox together with the ABox form a knowledge base (KB).

Recently, a tool has been developed for the manual annotation of traditional dance videos [Lagrue et al., 2019]. Domain experts may use such tool to annotate a video by decomposing it into segments (i.e., sets of frames). Several experts may annotate the same video, but they may not share the same opinion about dance elements showcased in some video segments. Thus the experts may disagree in their annotations. This may lead to inconsistencies (or conflicts) in the ABox with respect to the TBox, making the whole KB inconsistent. Standard query answering tools cannot be used in such a case as anything can be derived from an inconsistent KB. Furthermore, experts may assign confidence degrees to their annotations, reflecting various reliability levels of the information. This corresponds to defining a priority relation, namely a total preorder, over the assertions contained in the ABox. In this paper, we propose meaningful strategies for answering queries from inconsistent KBs that also take into account priorities to compute the answers.

After a brief overview of related work in Section 2, we present our ontology for Vietnamese Jar Dance in Section 3. We discuss query answering from fully reliable information in Section 4. We deal with conflicting and prioritized information and show how to rank-order query answers in Section 5, before concluding.

2 RELATED WORK

In [Saaw et al., 2012], an OWL ontology for movement in videos is built and an approach is proposed for the semantic annotation of movements using the Benesh Movement Notation. Ontology completion supports video retrieval through SPARQL queries. In [El Raheb and Ioannidis, 2011], a dance ontology is built in OWL-2 to represent and archive dance choreographies. SPARQL queries were applied for searching the steps and movements of dances within the ontology. In [Goienetxea Urkizu et al., 2012], a representation of folk song metadata is described using OWL. Based on using and extending the CIDOC CRM[3], a method is proposed to encode and structure metadata of folk song collections. In [El Raheb et al., 2016], OWL is used to represent ballet movements. A system is built to annotate dance videos based on concepts from a predefined ontology of ballet terminology and to search videos based on movement concepts. In [Bertini et al., 2008], OWL is used to support automatic semantic annotation and retrieval of video sequences. A Dance Video Semantic Model (DVSM) is proposed in [Ramadoss and Rajkumar, 2007]. It models dance video objects at various levels of detail which are determined by components of the accompanying song, and takes into account the meaning conveyed both by the song and movements.

The use of ontologies for modelling Vietnamese folk dances has been investigated in [Ma et al., 2018], with a focus on a popular dance in Vietnam known as Mõ folk dance. An ontology is proposed to define a taxonomy of dance movement classes and their relationships, for the traditional Vietnamese dances of type Mõ, taking into account the semantics of its art and its cultural anthropology. A subsequent study [Bourahla et al., 2019] extends the work in [Ma et al., 2018], by building a searchable knowledge base. This enables the search for non-elementary movements in Mõ dances. The ontology is augmented with classification rules, which are built with the OWL complementary language

[3]Conceptual Reference Model for cultural heritage documentation.

Figure 1: Tamia đwa buk dancers with traditional costumes

SWRL (Semantic Web Rule Language), to entail movement phrases having complete meaning. Furthermore, in [Ma-Thi et al., 2017], an approach is presented for annotating automatically movement phrases, mainly dancer movements, in Vietnamese folk dance videos. In this approach, annotation of videos is performed in three steps: movement phrase detection, movement phrase classification and movement phrase annotation.

To the best of our knowledge, there have been no studies dealing with an ontology-based modelling of Tamia đwa buk dances. In [Belabbes et al., 2019], very early steps were taken into the investigation of query answering from annotated videos of traditional Malaysian dances. The originality of the present work compared to the existing literature lies in the provision of query answering strategies that take into account conflicts and various reliability levels of the information to rank-order query answers. This corresponds to situations where different experts may provide conflicting annotations for a given video and may attach various confidence degrees to their annotations.

3 AN ONTOLOGY FOR VIETNAMESE TRADITIONAL DANCE TAMIA DWA BUK

3.1 Overview of Vietnamese Traditional Dances

Vietnamese traditional dances, with their wide variety, are rooted in the country's rich natural, historic, cultural, regional and ethnic diversity. Most of Vietnamese traditional dances are preserved through transmission in local communities and documents are extremely rare. Cham people are one of the 54 ethnic groups of Vietnam [Luyên N. T., 2013], located in the south of the country. Their traditional dances play a crucial role in spiritual and cultural life [Sakaya, 2010]. Some popular dances of Cham people include: Tamia đwa buk (a.k.a. Jar Dance), Tamia tadik (dance with paper fan), Tamia (towel dance) and Tamia jwak apwei (fire dance) [Ngô V. D., 2002]. These traditional dances can be classified into four main categories, namely: daily activity dances, religious dances, dances with props and dances without props. This paper focuses on the dance type Tamia đwa buk, later referred to as Jar Dance, in which female dancers hold jars, as illustrated in Figure 1.

A session of Jar Dance usually lasts between 3 and 5 minutes, and is performed by 4 to 8 female dancers. Jar Dance movements include: dancers standing, sitting, leaning and carrying a jar on the head (without holding it with the hand). Traditional costumes include: long dress (áo dài Chăm), long skirt (váy), a belt called "talei ka-in" (dây thắt lưng ngang), a scarf on the right shoulder called "talei kabak", head scarf (khăn đội đầu), earrings (khuyên tai) and jewellery around the neck.

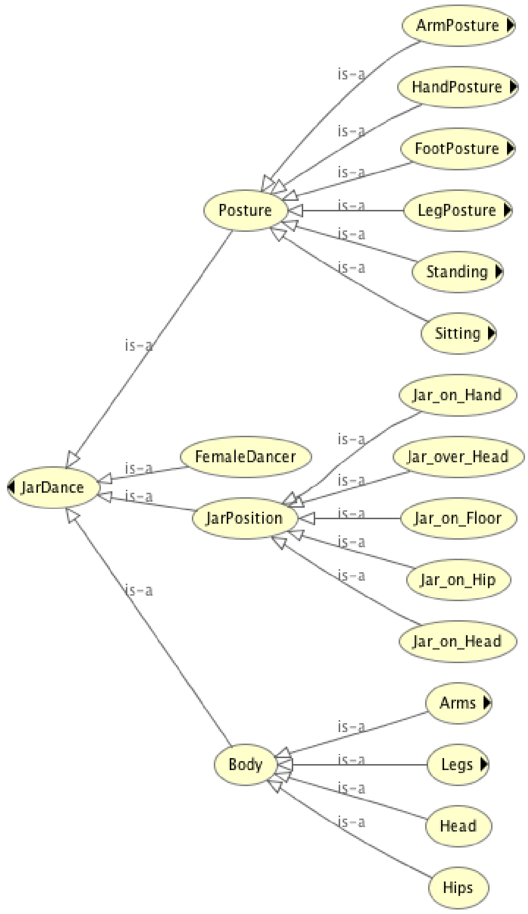

Figure 2: Overview of JARDANCE Ontology

3.2 The JARDANCE Ontology

We have built an ontology for Jar Dance using the free and open source tool Protégé[4]. The obtained OWL ontology, called JARDANCE.OWL, comprises 38 classes, 2 object properties and 54 logical axioms, making a total of 94 axioms. The corresponding OWL document JARDANCE.OWL is available for download from the link `http://www.cril.univ-artois.fr/aniage/ICAART19`.

Figure 2 gives an overview of the concepts in the ontology JARDANCE.OWL. It consists of a main class "JarDance", from which classes representing dance concepts are derived. There are three main concepts, namely: jar position, dancer body and dancer posture.

The jar's position in a Jar Dance has an important significance. Each position symbolises a daily life task performed by women in the community of Cham people. For instance, a dancer lifting the jar on her head refers to women carrying water. There are five jar positions as can be seen in Figure 2, namely:

- *Jar on hand:* dancer holds the jar in her hands.
- *Jar over head:* dancer carries the jar with both hands and lifts it up over the head.
- *Jar on floor:* jar is on the floor and dancer moves around it and makes a gesture like an activity of getting water.

[4]https://protege.stanford.edu/

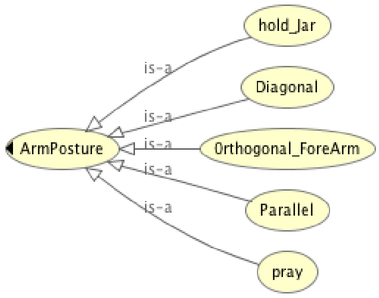

Figure 3: Arm Posture class and its sub-classes

- *Jar on hip:* left or right hand holds the jar on same side hip. The free hand moves following the music played by a traditional instrument.
- *Jar on head:* dancer carries the jar on the head without holding it with her hands.

Similarly to jar position, dancer postures have different symbolic meanings. They can be divided into six categories: arm posture, hand posture, leg posture, foot posture, standing posture and sitting posture. These are described below.

- *Arm posture*: it can be either with or without holding the jar. There are five arm postures, shown in Figure 3, as follows:
  - Hold jar: hold the jar either on the head, or over the head, or on the hip or on the hand.
  - *Diagonal*: hands aligned and tilted diagonally.
  - *Orthogonal forearm*: both hands are open and forearms are orthogonal to upper arms.
  - *Parallel*: both arms are parallel and straightforward, or parallel on the left/right side and point to the ground in a 45 degrees angle.
  - *Pray*: hold arms in front of chest like a prayer.
- *Hand Posture*: it has three sub-classes as shown in Figure 4, where Orthogonal-Straight Fingers means that index to baby finger are straight while thumb is orthogonal to the other four fingers. Notice that Hold Jar has two super-classes: Arm Posture and Hand Posture.
- *Leg Posture*: this depends on whether the dancer is in sitting posture or standing posture. Genuflect posture corresponds to sitting posture, while flexed and straight leg postures are performed within standing posture. (See Figure 5)
- *Foot Posture*: there are four sub-classes as shown in Figure 6, where Tiptoed Foot Straight Foot means one foot is tiptoed and the other is straight.
- *Standing posture*: it can be of two types. Buckled standing posture combines buckled leg postures (one buckled leg and one straight leg, or both buckled legs) and arm postures. Straight standing posture combines straight leg postures (tiptoed foot or straight foot) and arm postures.
- *Sitting posture*: it includes genuflection, rotated body and upright body. Genuflections on the floor are combined with arm movements.

In the rest of this paper, we discuss how an ontology can be used in answering queries from ICH digital content. We distinguish the case where information is consistent (i.e., conflict-free) and fully reliable, and the case where it is inconsistent and prioritized.

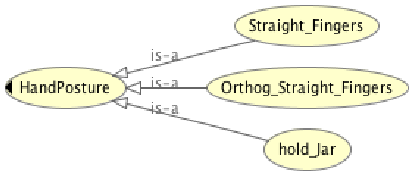

Figure 4: Hand Posture class and its sub-classes

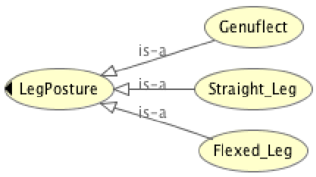

Figure 5: Leg Posture class and its sub-classes

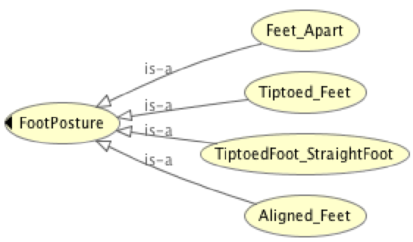

Figure 6: Foot Posture class and its sub-classes

4 QUERYING CONSISTENT ASSERTIONS

Ontology-Mediated Query Answering (OMQA) is an active research area that takes advantage of the semantic knowledge specified in an ontology to answer queries about data [Bienvenu and Ortiz, 2015]. Typically, querying is performed on a knowledge base (KB) which is composed of an ontology (a TBox) and a dataset of assertional facts (an ABox). Description Logic languages are widely used to encode, integrate and maintain ontologies [Baader et al., 2017]. In this paper, we formalise our ontology using the lightweight fragment DL-Lite [Calvanese et al., 2007], since it provides a good trade-off between expressive power and computational complexity. For lack of space, we do not provide the full transcription of our ontology in DL-Lite, but we mention some axioms in the examples when needed. We first provide a brief reminder on DL-Lite languages which are used for formalising lightweight ontologies, followed by a small illustrative example issued from our ontology, then a section on querying fully reliable assertions.

4.1 A Brief Reminder on DL-Lite

The basic notions of DL-Lite languages are as follows. Consider finite sets of concept names, role names and individual names. Let B (resp. C) denote a *basic* (resp. *complex*) *concept*. Let R (resp. E) denote a *basic* (resp. *complex*) *role*. An *inclusion axiom* on concepts (resp. on roles) is a statement of the form $B \sqsubseteq C$ (resp. $R \sqsubseteq E$). Inclusion axioms with $\neg$ in the right-hand side are called *negative inclusions*.

A *TBox* $\mathcal{T}$ is a finite set of inclusion axioms. An *ABox* $\mathcal{A}$ is a finite set of assertions, that is, concept or role names defined over individual names. A *knowledge base* (KB) is a tuple $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$. A KB is said to be *consistent* if it admits at least one model, it is *inconsistent* otherwise. A TBox $\mathcal{T}$ is *incoherent* if there is a concept name that is empty in every model of $\mathcal{T}$, it is *coherent* otherwise.

We refer the reader to the work of [Calvanese et al., 2007] for further details on DL-Lite.

4.2 Illustrative Example

In order to build a dataset for enriching our ontology, we consider a set of Jar Dance videos. Domain experts are asked to semantically annotate each video in terms of the concepts and roles defined in the ontology. Annotations are then translated into assertions of an ABox. We use the following running example to illustrate query answering from annotated videos.

Example 1. *Assume a video v_1 annotated by three experts, E_1, E_2 and E_3. For the sake of simplicity, we focus only on annotations associated with one video segment, denoted by $[f_1, f_2]$, where f_1 represents its first frame and f_2 its last frame. Let frames f_1 and f_2 be those of Figure 7 and Figure 8, respectively.*
Annotations of experts E_1, E_2 and E_3 for segment $[f_1, f_2]$ are listed in Tables 1 to 3.
For instance, annotations in Table 1 depict: dancers sitting (i.e., $[f_1, f_2]$ is an instance of the concept Sitting*), holding a jar on the head (i.e., $[f_1, f_2]$ is an instance of the concept* Jar_on_Head*), and having a posture of straight bust and orthogonal forearms (i.e., $[f_1, f_2]$ is an instance of concepts* Straight_Bust *and* Orthogonal_FA*).* □

The tables also contain confidence degrees attached to annotations. Values α_1, α_2, β_1 and γ_1 are positive numbers defined over a totally ordered uncertainty scale, where "1" is the highest value and "0" is the lowest value. As we shall see later, assertional

Figure 7: Frame f_1 of a Jar Dance video

Figure 8: Frame f_2 of a Jar Dance video

Expert E_1	
Annotation (assertion)	Confidence degree
Sitting$([f_1, f_2])$	1
Jar_on_Head$([f_1, f_2])$	1
Straight_Bust$([f_1, f_2])$	α_1
Orthogonal_FA$([f_1, f_2$	α_2

Table 1: Expert E_1's annotations of segment $[f_1, f_2]$

Expert E_2	
Annotation (assertion)	Confidence degree
Sitting$([f_1, f_2])$	1
Jar_on_Head$([f_1, f_2])$	1
Genuflect$([f_1, f_2])$	1
Pray$([f_1, f_2])$	β_1

Table 2: Expert E_2's annotations of segment $[f_1, f_2]$

facts with confidence degrees can actually be represented by a totally pre-ordered (or prioritized) ABox.

For lack of space, we do not provide the full transcription of the ontology (TBox) in DL-Lite and mention only some axioms when needed.

Next, we discuss query answering in the case where the ABox is consistent w.r.t. the TBox (i.e., all experts agree in their annotations), and the ABox is non-prioritized or flat

Expert E_3	
Annotation (assertion)	Confidence degree
$Sitting([f_1, f_2])$	1
$Jar\_on\_Head([f_1, f_2])$	1
$Tilted\_Bust([f_1, f_2])$	γ_1

Table 3: Expert E_3's annotations of segment $[f_1, f_2]$

(i.e., experts are fully confident in their annotations).

4.3 Non-prioritized Consistent ABoxes

Let us assume that all experts agree in their annotations of any given video and that they are fully confident about their own annotations. Namely, we assume values α_1, α_2, β_1, γ_1 (given in Tables 1 to 3) are all equal to 1. This means that the ABox is consistent w.r.t. the TBox, and that no priorities are assigned to assertions (i.e., the ABox is flat). Then query answering (QA) simply amounts to using any standard QA tool (for instance, a DL-Lite QA tool when the TBox and ABox are expressed in DL-Lite).

In this case, the input of the QA tool consists of a set of annotated videos (an ABox), an ontology (a TBox) and a conjunctive query $q(\vec{x})$. The output of the QA tool is a set of answers X. Note that when $\vec{x}$ is empty, then $q(.)$ is a boolean query and its answer is either yes or no.

Example 1 (continued)**.** *Consider the TBox $\mathcal{T}$ contains at least two axioms: $\mathcal{T} \supset \{Straight\_Bust \sqsubseteq Standing,\ Standing \sqsubseteq \neg Sitting\}$. In other words, concept* Straight_Bust *is a type of* Standing*, and concepts* Standing *and* Sitting *are disjoint.*
Consider annotations of video v_1 by expert E_1 and ignore confidence degrees. Thus:

$$\mathcal{A}_1 = \{Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]),\\ Straight\_Bust([f_1, f_2]),\ Orthogonal\_FA([f_1, f_2])\}$$

Consider the boolean query $q(.) \leftarrow \exists x, Standing(x)$. ABox $\mathcal{A}_1$ contains assertion $Straight\_Bust([f_1, f_2])$, but it also contains assertion $Sitting([f_1, f_2]$. According to $\mathcal{T}$, Standing and sitting are disjoint, then the query's answer is no. □

One may want to rank-order videos instead of answers. This can be achieved by first collecting the answers associated with each video, and then rank-ordering the sets of answers, as follows:

Definition 1. *Let v_1, v_2 be two videos. Then v_1 is presented to the user before v_2, denoted $v_1 > v_2$, iff $|X(v_1)| > |X(v_2)|$, where $|X(v_i)|, i = 1, 2$, is the size of the set of answers obtained from v_i and the ontology.*

Example 1 (continued)**.** *Consider a video v_2 with a segment $[f_3, f_4]$, annotated by some expert. Let the corresponding ABox be:*
$\mathcal{A}_2 = \{Standing([f_3, f_4]),\ Jar\_on\_Floor([f_3, f_4])\}$.
Consider now a query $q(x) \leftarrow \exists x, Jar\_on\_Head(x)$, asking for all individuals holding a jar on the head.
Querying $\mathcal{A}_1$ of video v_1 returns $X(v_1) = \{[f_1, f_2]\}$, while querying $\mathcal{A}_2$ of video v_2 returns $X(v_2) = \emptyset$. Since $|X(v_1)| > |X(v_2)|$, then video v_1 is ranked first and shown to the user because its component better fits the query. □

When different experts are asked to annotate the same video, their annotations may potentially be conflicting since the experts may disagree about some elements depicted in video segments. This issue is addressed in the next section.

5 QUERYING CONFLICTING ASSERTIONS

In this section, we assume that any given video may receive conflicting annotations obtained from multiple experts. Hence, there may be conflicts between assertions of the ABox w.r.t. the TBox, resulting in an inconsistent KB. Moreover, we assume that experts can express confidence in their own annotations. This can be captured by applying a priority relation over ABox assertions.

Here, we go beyond standard OMQA and propose strategies for querying inconsistent KBs when the ABox is prioritized. We take the reasonable assumption stating that the TBox is stable, coherent and reliable, and distinguish two cases as follows:

1. The ABox contains conflicting assertions and all experts are fully confident in their annotations.
2. The ABox contains conflicting assertions and experts may assign confidence degrees to their annotations.

5.1 Non-prioritized Conflicting ABoxes

Let us take a closer look at the case where experts may disagree with one another in their annotations of a given video, but they are fully confident about their own annotations. Here we still consider that the confidence degrees α_1, α_2, β_1 and γ_1 used in Tables 1, 2 and 3 are all equal to 1. Hence the KB is inconsistent and the ABox is flat. In this case, one may not use standard QA tools because every tuple would be returned as a query answer from the inconsistent KB.

In what follows, we explain how to circumvent this situation and introduce some definitions. We first define an assertional conflict.

Definition 2. *Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be a flat ABox and $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an inconsistent KB. A sub-base $\mathcal{C} \subseteq \mathcal{A}$ is an assertional conflict of $\mathcal{K}$ iff $\langle \mathcal{T}, \mathcal{C} \rangle$ is inconsistent and $\forall g \in \mathcal{C}$, $\langle \mathcal{T}, \mathcal{C} \setminus \{g\} \rangle$ is consistent.*

Example 2. *Consider annotations on video v_1 by experts E_1 to E_3 from Tables 1 to 3, without confidence degrees. The resulting ABox, denoted by $\mathcal{A}$, is:*

$$\begin{aligned}\mathcal{A} = \{&Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]),\\ &Straight\_Bust([f_1, f_2]),\ Genuflect([f_1, f_2]),\\ &Pray([f_1, f_2]),\ Orthogonal\_FA([f_1, f_2]),\\ &Tilted\_Bust([f_1, f_2])\}\end{aligned}$$

The TBox $\mathcal{T}$ also contains the following two axioms: $\mathcal{T} \supset \{Straight\_Bust \sqsubseteq \neg\, Tilted\_Bust$, $Pray \sqsubseteq \neg Orthogonal\_FA\}$. The first negative axiom means that Straight_Bust *and* Tilted_Bust *are disjoint concepts. The second negative axiom means that* Pray *and* Orthogonal_FA *are disjoint concepts.*

This implies that $\{Straight\_Bust([f_1, f_2]), Tilted\_Bust([f_1, f_2])\}$ and $\{Pray([f_1, f_2])$, $Orthogonal\_FA([f_1, f_2])\}$ are two assertional conflicts of ABox $\mathcal{A}$. □

A pivotal notion when dealing with an inconsistent KB is that of a repair [Lembo et al., 2010]. Formally:

Definition 3. *Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be a flat ABox and $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an inconsistent KB. A sub-base $\mathcal{R} \subseteq \mathcal{A}$ is a repair iff $\langle \mathcal{T}, \mathcal{R} \rangle$ is consistent, and $\forall \mathcal{R}' \subseteq \mathcal{A}$: $\mathcal{R} \subsetneq \mathcal{R}'$, $\langle \mathcal{T}, \mathcal{R}' \rangle$ is inconsistent.*
Furthermore if $\langle \mathcal{T}, \mathcal{A} \rangle$ is consistent, then there exists only one repair $\mathcal{R} = \mathcal{A}$.

It follows that a repair is a maximal subset of $\mathcal{A}$ that is consistent w.r.t. $\mathcal{T}$. Usually, there are several repairs for any given ABox, as shown in this example:

Example 2 (continued). *Assertions involved in a conflict may not appear in the same repair. Thus the repairs of $\mathcal{A}$ are:*

$$
\begin{aligned}
\mathcal{R}_1 = \{ & Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]), \\
& Straight\_Bust([f_1, f_2]),\ Genuflect([f_1, f_2]), \\
& Pray([f_1, f_2])\} \\
\mathcal{R}_2 = \{ & Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]), \\
& Straight\_Bust([f_1, f_2]),\ Genuflect([f_1, f_2]), \\
& Orthogonal\_FA([f_1, f_2])\} \\
\mathcal{R}_3 = \{ & Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]), \\
& Genuflect([f_1, f_2]),\ Pray([f_1, f_2]), \\
& Tilted\_Bust([f_1, f_2])\} \\
\mathcal{R}_4 = \{ & Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]), \\
& Genuflect([f_1, f_2]),\ Orthogonal\_FA([f_1, f_2]), \\
& Tilted\_Bust([f_1, f_2])\}
\end{aligned}
$$

□

For answering queries from several repairs of an ABox, one needs to define some strategy for selecting repairs. Let us define the notion of a cardinality-preferred repair.

Definition 4. *Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ be a flat ABox and $\mathcal{K} = \langle \mathcal{T}, \mathcal{A} \rangle$ be an inconsistent KB. Let $\mathcal{R}_1, \mathcal{R}_2$ be two repairs of $\mathcal{A}$. Then $\mathcal{R}_1$ is cardinality-preferred to $\mathcal{R}_2$ iff $|\mathcal{R}_1| > |\mathcal{R}_2|$.*

Note the difference between Definitions 3 and 4. In the former, a repair is a maximal set in terms of consistency, namely, adding any new assertion makes the set inconsistent. In the latter, a repair is cardinality-preferred to another one in terms of set cardinality, namely, it is a maximal consistent set containing a larger number of assertions.

Let $CR(\mathcal{A}) = \{\mathcal{R}_1 \subseteq \mathcal{A}$, $\mathcal{R}_1$ is a repair s.t. $\nexists \mathcal{R}_2 \subseteq \mathcal{A} : \mathcal{R}_2$ is a repair and $|\mathcal{R}_2| > |\mathcal{R}_1|\}$ denote the set of cardinality-preferred repairs of an ABox $\mathcal{A}$, i.e., those with the largest number of assertions. We then define a query answer as follows:

Definition 5. *Let $\mathcal{A}$ be an ABox and $CR(\mathcal{A})$ be the set of cardinality-preferred repairs. Given a query $q(\vec{x})$, an answer $\vec{x}$ is valid iff it can be derived from every repair $\mathcal{R} \in CR(\mathcal{A})$.*

Example 2 (continued). *All repairs of $\mathcal{A}$ contain five assertions. Hence, the set of cardinality-preferred repairs of $\mathcal{A}$ is $CR(\mathcal{A}) = \{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4\}$.*
Consider the query $q(.) \leftarrow \exists x, Jar\_on\_Head(x)$. Since all repairs of $\mathcal{A}$ contain $Jar\_on\_Head([f_1, f_2])$, then the answer is yes.
Consider now the query $q(.) \leftarrow \exists x, Pray(x)$. Then the answer is no, since there is no individual d such that $Pray(d)$ can be derived from repairs $\mathcal{R}_2$ and $\mathcal{R}_4$. □

Now the question is how to rank-order videos. Let us define $X_{CR}(v)$ as the set of answers based on $CR(\mathcal{A})$, where the ABox $\mathcal{A}$ encodes annotations of video v. Thus we are also able to compare videos by applying a variant of Definition 1 in which $X(v_i)$ is replaced with $X_{CR}(v_i)$. Formally:

Definition 6. *Let v_1, v_2 be two videos. Then $v_1 > v_2$ iff $|X_{CR}(v_1)| > |X_{CR}(v_2)|$, $i = 1, 2$.*

Example 3. *Consider a video v_3 with a segment $[f_5, f_6]$, of which annotations produce ABox $\mathcal{A}_3$:*

$$\mathcal{A}_3 = \{Standing([f_5, f_6]),\ Jar\_on\_Floor([f_5, f_6]),\ Straight\_Bust([f_5, f_6]),\ Tilted\_Bust([f_5, f_6])\}$$

There are two repairs, $\mathcal{R}'_1$ and $\mathcal{R}'_2$, for $\mathcal{A}_3$. Namely:

$$\mathcal{R}'_1 = \{Standing([f_5, f_6]),\ Jar\_on\_Floor([f_5, f_6]),\ Straight\_Bust([f_5, f_6])\}$$
$$\mathcal{R}'_2 = \{Standing([f_5, f_6]),\ Jar\_on\_Floor([f_5, f_6]),\ Tilted\_Bust([f_5, f_6])\}.$$

Consider the query $q(x) \leftarrow \exists x,\ Jar\_on\_Head(x)$. The answer is yes with ABox $\mathcal{A}_1$ of video v_1 (from Example 1), while it is no with ABox $\mathcal{A}_3$ of video v_3. Then video v_1 is ranked first and presented to the user. □

5.2 Prioritized Conflicting ABoxes

Let us assume that the experts disagree with one another in their annotations and that they are not fully confident about their own annotations, so some annotations are deemed as more reliable than others. Hence the KB is inconsistent and the ABox is no longer flat. In fact, the introduction of confidence degrees requires adapting the notions of ABox, repairs and query answers.

As stated previously, the assessment of confidence degrees is done on a totally ordered uncertainty scale (a qualitative uncertainty scale). Hence, for any given video v, the corresponding ABox $\mathcal{A}$ is partitioned into strata like so: $\mathcal{A} = (\mathcal{S}_1, \ldots, \mathcal{S}_n)$, where $\mathcal{S}_1$ (resp. $\mathcal{S}_n$) contains the most (resp. least) reliable assertions. Assertions of the same stratum have the same confidence degree. In this case, Definition 3 of a repair still holds.

Example 4. *Consider annotations of experts E_1, E_2 and E_3 of video v_1 (given in Example 1), together with their confidence degrees, where:*

$1 > \alpha_1 > \beta_1 > \alpha_2 = \gamma_1$.

Here, a total preorder applied to ABox assertions produces a partitioning of $\mathcal{A}$ into four strata. Namely:

$$\mathcal{S}_1 = \{Sitting([f_1, f_2]),\ Jar\_on\_Head([f_1, f_2]),\ Genuflect([f_1, f_2])\}$$
$$\mathcal{S}_2 = \{Straight\_Bust([f_1, f_2])\}$$
$$\mathcal{S}_3 = \{Pray([f_1, f_2])\}$$
$$\mathcal{S}_4 = \{Orthogonal\_FA([f_1, f_2]),\ Tilted\_Bust([f_1, f_2])\}$$

The repairs of $\mathcal{A}$ remain the same as in Example 2, namely: $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4$. □

When the ABox is stratified, the definition of a preferred repair needs to be adapted in order to take into account priorities. We introduce the notion of a PC-preferred repair (where PC stands for priorities and cardinality), first proposed in the context of propositional logic [Benferhat et al., 1993].

Definition 7. *Let* $\mathcal{A} = (\mathcal{S}_1, \ldots, \mathcal{S}_n)$ *be a prioritized ABox and* $\mathcal{R}_1, \mathcal{R}_2$ *be two repairs of* $\mathcal{A}$*. Then* $\mathcal{R}_1$ *is PC-preferred to* $\mathcal{R}_2$ *iff* $\exists i, 1 \leq i \leq n$, $|\mathcal{R}_1 \cap \mathcal{S}_i| > |\mathcal{R}_2 \cap \mathcal{S}_i|$ *and* $\forall j, 1 \leq j < i$, $|\mathcal{R}_1 \cap \mathcal{S}_j| = |\mathcal{R}_2 \cap \mathcal{S}_j|$.

Similarly to the case discussed in Section 5.1, let $PCR(\mathcal{A})$ denote the set of PC-preferred repairs.

Example 4 (continued)**.** *Recall that the TBox contains these two axioms:* $\mathcal{T} \supset \{Straight\_Bust \sqsubseteq \neg\, Tilted\_Bust,\ Pray \sqsubseteq \neg Orthogonal\_FA\}$*. Since* $Straight\_Bust([f_1, f_2])$ *is in* $\mathcal{S}_2$*, it is strictly preferred to* $Tilted\_Bust([f_1, f_2])$ *which is in* $\mathcal{S}_4$*. Similarly, since* $Pray([f_1, f_2])$ *is in* $\mathcal{S}_3$*, it is strictly preferred to* $Orthogonal\_FA([f_1, f_2])$ *which is in* $\mathcal{S}_4$*. Then* $\mathcal{A}$ *has a single PC-preferred repair, namely :*

$$\begin{aligned} PCR(\mathcal{A}) = \{\{&Sitting([f_1, f_2]),\ Genuflect([f_1, f_2]),\\ &Jar\_on\_Head([f_1, f_2]),\ Pray([f_1, f_2]),\\ &Straight\_Bust([f_1, f_2])\}\}. \end{aligned}$$

□

Given a query $q(\vec{x})$, an answer $\vec{x}$ is PC-valid (or PC-consequence) if it can be derived from every repair $\mathcal{R} \in PCR(\mathcal{A})$. Furthermore, let $X_{PCR}(v)$ be the set of answers based on $PCR(\mathcal{A})$, where the ABox $\mathcal{A}$ encodes the annotated video v. Now, one may assign to each answer $\vec{x}$ a priority degree, denoted $\alpha_{\vec{x}}$, as the first rank (the most important rank) from which $\vec{x}$ is derived. More precisely, assume that $\vec{x}$ is a PC-consequence. Then a priority $\alpha_{\vec{x}}$ associated with $\vec{x}$ is $\alpha_{\vec{x}} = i$ obtained as follows:

i) $\vec{x}$ is a PC-valid answer of $PCR(\mathcal{S}_1 \cup \cdots \cup \mathcal{S}_i)$, and

ii) $\forall j > i, \vec{x}$ is not a PC-valid answer of $PCR(\mathcal{S}_j \cup \cdots \cup \mathcal{S}_n)$.

Thanks to the priorities associated with answers, $X_{PCR}(v)$ can be split into: $X^1_{PCR}(v) \cup \cdots \cup X^n_{PCR}(v)$, where $X^i_{PCR}(v)$ are answers obtained with priority i. Thus, we are able to compare videos like so:

Definition 8. *Let* v_1, v_2 *be two videos. Then* $v_1 > v_2$ *iff* $\exists i, 1 \leq i \leq n$, $|X^i_{PCR}(v_1)| > |X^i_{PCR}(v_2)|$ *and* $\forall j, 1 \leq j < i$, $|X^j_{PCR}(v_1)| = |X^j_{PCR}(v_2)|$.

Namely, answers with the highest priority degree are the most preferred ones. Note that Definition 8 extends Definition 6 when all answers have the same priority level.

Example 5. *Consider a video* v_4 *(similar to* v_1*) of which the corresponding ABox* $\mathcal{A}_4$ *is stratified into:*

$$\begin{aligned} \mathcal{S}'_1 &= \{Sitting([f_7, f_8]),\ Jar\_on\_Head([f_7, f_8]),\\ &\quad Genuflect([f_7, f_8])\}\\ \mathcal{S}'_2 &= \{Straight\_Bust([f_7, f_8]),\ Pray([f_7, f_8])\}\\ \mathcal{S}'_3 &= \{Orthogonal\_FA([f_7, f_8]),\ Tilted\_Bust([f_7,\\ &\quad f_8])\} \end{aligned}$$

One can check that $\mathcal{A}_4$ admits a single PC-preferred repair, similarly to $\mathcal{A}$ in Example 4. Thus:

$$\begin{aligned} PCR(\mathcal{A}_4) = \{\{ & Sitting([f_7, f_8]),\ Genuflect([f_7, f_8]), \\ & Jar\_on\_Head([f_7, f_8]),\ Pray([f_7, f_8]), \\ & Straight\_Bust([f_7, f_8])\}\} \end{aligned}$$

Consider the query $q(x) = \forall x.Pray(x)$. Clearly, $\vec{x} = \{[f_1, f_2]\}$ and $\vec{x}' = \{[f_7, f_8]\}$ are both answers from $\mathcal{A}$ and $\mathcal{A}_4$, respectively. However, $[f_1, f_2]$ is obtained with rank 3 (i.e., from stratum $\mathcal{S}_3$ of Example 4), while $[f_7, f_8]$ is obtained with rank 2 (i.e., from stratum $\mathcal{S}'_2$). Using Definition 8, video v_4 is preferred to video v_1, hence it is presented first to the user. □

6 CONCLUSION

In this paper, we first proposed an ontology, that can be represented in Description Logic languages, to capture the cultural knowledge conveyed by Vietnamese traditional dances *Tamia đwa buk*. We then proposed to enrich our ontology by manually annotating dance videos. A tool for manual annotation of videos, based on ontologies, has been developed [Lagrue et al., 2019] and used in this study. Lastly, we proposed strategies for querying ontologies in the presence of conflicting and prioritized data.

Enriching ontologies by annotating videos is a crucial task for query answering. This task may appear to be daunting for large datasets. In this work, we took the first steps towards developing a comprehensive tool for automatic annotations of videos and for querying their content using machine learning. The idea is to build a training set from a set of manually annotated videos, then to use machine learning techniques to train machine learning models to annotate videos automatically.

ACKNOWLEDGEMENTS

Work supported by H2020-MSCA-RISE European project: AniAge (High Dimensional Heterogeneous Data-based Animation Techniques for Southeast Asian Intangible Cultural Heritage Digital Content).

References

[Baader et al., 2007] Baader, F., Calvanese, D., Mcguinness, D., Nardi, D., and Patel-Schneider, P. (2007). *The Description Logic Handbook: Theory, Implementation, and Applications*.

[Baader et al., 2017] Baader, F., Horrocks, I., Lutz, C., and Sattler, U. (2017). *An Introduction to Description Logic*. Cambridge University Press.

[Belabbes et al., 2019] Belabbes, S., Tan, C.-W., Vo, T.-T., Izza, Y., Tabia, K., and Benferhat, S. (2019). Query-answering from traditional dance videos: Case study of zapin dances. In *Proceedings of the 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 1630–1634. IEEE.

[Benferhat et al., 1993] Benferhat, S., Cayrol, C., Dubois, D., Lang, J., and Prade, H. (1993). Inconsistency management and prioritized syntax-based entailment. In *International Joint Conference on Artificial Intelligence*, pages 640–647. Morgan Kaufmann.

[Bertini et al., 2008] Bertini, M., Del Bimbo, A., and Serra, G. (2008). Learning ontology rules for semantic video annotation. In *Proceedings of the 2nd ACM Workshop on Multimedia Semantics*, pages 1–8, Vancouver, Canada. ACM.

[Bienvenu and Ortiz, 2015] Bienvenu, M. and Ortiz, M. (2015). Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web. Web Logic Rules - 11th International Summer School, Berlin, Germany, Tutorial Lectures*, pages 218–307.

[Bourahla et al., 2019] Bourahla, M., Telli, A., Benferhat, S., and Chau, M. T. (2019). Classifying non-elementary movements in vietnamese mõ dances. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion - 10th International Conference, DHM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part I*, pages 128–139.

[Calvanese et al., 2007] Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., and Rosati, R. (2007). Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning*, 39(3):385–429.

[El Raheb and Ioannidis, 2011] El Raheb, K. and Ioannidis, Y. (2011). A labanotation based ontology for representing dance movement. In *International Gesture Workshop*, pages 106–117. Springer.

[El Raheb et al., 2016] El Raheb, K., Papapetrou, N., Katifori, V., and Ioannidis, Y. (2016). BalOnSe: ballet ontology for annotating and searching video performances. In *Proceedings of the 3rd International Symposium on Movement and Computing*, pages 1–8. ACM.

[Goienetxea Urkizu et al., 2012] Goienetxea Urkizu, I., Arrieta, I., Bagüés, J., Cuesta, A., Leiñena, P., and Conklin, D. (2012). Ontologies for representation of folk song metadata. Technical Report EHU-KZAA-TR-2012-01, University of the Basque Country.

[Lagrue et al., 2019] Lagrue, S., Chetcuti-Sperandio, N., Delorme, F., Ma-Thi, C., Ngo-Thi, D., Tabia, K., and Benferhat, S. (2019). An ontology web application-based annotation tool for intangible culture heritage dance videos. In *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents*, SUMAC '19, pages 75–81. ACM.

[Lembo et al., 2010] Lembo, D., Lenzerini, M., Rosati, R., Ruzzi, M., and Savo, D. F. (2010). Inconsistency-tolerant semantics for description logics. In *International Conference on Web Reasoning and Rule Systems*, volume 6333 of *LNCS*, pages 103–117.

[Luyên N. T., 2013] Luyên N. T. (2013). *Giáo trình Trang phục các dân tộc Việt Nam*. Nhà xuất bản Đại học Quốc gia TP.HCM.

[Ma et al., 2018] Ma, T., Benferhat, S., Bouraoui, Z., Tabia, K., Do, T., and Nguyen, H. (2018). An ontology-based modelling of vietnamese traditional dances (S). In *The 30th International Conference on Software Engineering and Knowledge Engineering, Redwood City, USA, July 1-3, 2018*, pages 64–67.

[Ma-Thi et al., 2017] Ma-Thi, C., Tabia, K., Lagrue, S., Le-Thanh, H., Bui-The, D., and Nguyen-Thanh, T. (2017). Annotating movement phrases in vietnamese folk dance videos. In *Advances in Artificial Intelligence: From Theory to Practice*, pages 3–11. Springer.

[Ngô V. D., 2002] Ngô V. D. (2002). *Văn hóa Cổ Chămpa*. Nhà xuất bản văn hóa dân tộc Publishers.

[Ramadoss and Rajkumar, 2007] Ramadoss, B. and Rajkumar, K. (2007). Modeling and annotating the expressive semantics of dance videos. In *International Journal "Information Technologies and Knowledge" Vol.1*.

[Saaw et al., 2012] Saaw, S., Beul, D. D., Said, M., and Manneback, P. (2012). An ontology for video human movement representation based on benesh notation. In *International Conference on Multimedia Computing and Systems*, pages 77–82. IEEE.

[Sakaya, 2010] Sakaya (2010). *Văn hóa Chăm: Nghiên cứu và Phê bình, Tập I*. Nhà xuất bản Phụ nữ Publishers.