VidOnt: a core reference ontology for reasoning over video scenes

Leslie F. Sikos

Centre for Knowledge and Interaction Technologies, Flinders University, Adelaide, Australia

ABSTRACT
The conceptualization of domains depicted in videos is a necessary, but not sufficient requirement for reasoning-based high-level scene interpretation, which requires the formal representation of the timeline structure, the moving regions of interest, and video production standards, facilities, and procedures as well. Multimedia ontologies, including the very few video ontologies, however, are not exhaustive in terms of concept coverage, redefine terms against Semantic Web best practices, are not aligned with standards, and do not define complex roles and role interdependencies. Because most multimedia ontologies implement only a minimal subset of the mathematical constructors of OWL and define a TBox and an ABox, but not an RBox, they do not support complex inferencing. This paper describes a formally grounded core reference ontology for video representation, which addresses many of these issues and limitations.

ARTICLE HISTORY
Received 28 April 2017
Accepted 5 February 2018

KEYWORDS
Video ontology; scene interpretation; video understanding; MPEG-7; spatiotemporal annotation; content-based video retrieval

INTRODUCTION TO REASONING OVER VIDEO SCENES

Description logics (DL), which are formal knowledge representation languages, have multimedia implementations since the 1990s (e.g., Meghini, Sebastiani, & Straccia, 1997). They are suitable for the expressive formalization of multimedia contents and the semantic refinement of video segmentation (Sikos, 2017b). DL-based knowledge representations that implement terms from OWL ontologies can serve as the basis for multimedia content analysis (Simou et al., 2006) and event detection (Town, 2006), differentiating between similar concepts via abductive reasoning (Möller & Neumann, 2008), interpreting hidden spatial information from the represented 3D space (Sikos, 2017c), constructing high-level media descriptors (Elleuch, Zarka, Ammar, & Alimi, 2011), and performing high-level video scene interpretation (Gómez-Romero, Patricio, García, & Molina, 2011) and content-based video retrieval (Sikos, 2018b). Ontology-based annotations are utilized in hypervideo applications and video sharing portals to narrow, if not bridge, the Semantic Gap in videos (Sikos, 2016b). However, high-level video semantics cannot be formalized in
ontologies that contain terminological and assertional axioms alone. More advanced formalisms, such as role boxes and rules, are also needed for this purpose (Sikos, 2016a), which enable video content understanding and improve the quality of structured annotations of concepts and predicates (Jaimes, Tseng, & Smith, 2003). The represented video concepts can be curated by natural language processing algorithms while preserving provenance data, thereby assisting to achieve knowledge base consistency (Dasiopoulou, Heinecke, Saathoff, & Strintzis, 2007).

Similar to ontologies of other domains, such as the 3D Modeling Ontology in the 3D modeling domain (Sikos, 2016c), many of the RDFS and OWL ontologies used for video annotations were translated from XSD-based industry standards. Some of the prime examples are MPEG-7, MPEG-21, and TV-Anytime. In contrast to ontologies of other knowledge domains, however, video ontologies need not only a large number of concepts in a taxonomical structure, but also concepts, roles, and rules for spatiotemporal annotation to be used for describing video events (Sikos, 2017a), which are characterized by complexity, multiple interpretations, and ambiguity (D’Odorico & Bennett, 2013). But even so, several formalisms are mentioned in the literature for describing video events, particularly those in constrained videos, such as medical videos, news videos, and sports videos (see, for example, Ballan, Bertini, Del Bimbo, & Serra, 2010). These have promising results in video classification, surveillance, medical training, real-time activity monitoring, automated subtitle generation, content-based video indexing, and intelligent video analytics.

**Rationale**

The quality of the conceptualization and the formal description of relations between concepts, predicates, and individuals determine the feasibility and efficiency of ontological reasoning (Hitzler, Krötzsch, & Rudolph, 2009). However, the correlations between video properties, which could be exploited in video annotation and classification, are not defined in any ontology. Moreover, despite the known benefits of multimedia reasoning in video classification, scene interpretation, and content-based video retrieval, most multimedia ontologies lack the constructors necessary for expressing axioms upon which complex inference tasks could be performed (Sikos & Powers, 2015).

To address the reasoning limitations of previous multimedia ontologies, the Video Ontology (VidOnt) has been introduced, with the namespace http://purl.org/ontology/video/, which defines fundamental video production concepts, video characteristics, and spatiotemporal segments, and the relationship between them, aligned with standard concept definitions. This is the very first video ontology that exploits the entire range of mathematical constructors of OWL 2 DL, rather than a small subset of constructors available in OWL DL. Ontology-based reasoning is not feasible without expressive constructors (Sikos, 2017d), most of which are not implemented in multimedia ontologies other than the Video Ontology. For example, the Visual Descriptor Ontology, which was published as an ‘ontology for multimedia reasoning’ (Simou, Tzouvaras, Avrithis, Stamou, & Kollias, 2005), has in fact a rather limited reasoning potential, because it is based on the constructors of the basic description logic $ALC$.}

The video terms defined by multimedia vocabularies and ontologies are very much isolated, and their relationships are not defined formally. Based on the explicit statements...
about the vague terminology of video production and classification provided by previous multimedia ontologies, reasoning services cannot infer high-level semantics, only statements about consistency and satisfiability. None of the previous video ontologies was designed with formal grounding and high enough expressivity to capture complex relationships between concepts and roles. In fact, the video terminology of previous ontologies is limited to either low-level multimedia descriptors, most of which do not convey information about the video content, or basic classification schemata for Hollywood movies, which may indicate the nature of the video content, but not the actual meaning of the content. To address these issues, the Video Ontology

1. provides a comprehensive coverage of videography concepts by
   a) defining core concepts not defined in any other video or multimedia ontology,
   b) adding taxonomical structure to standard and de facto standard terminology (see Figure 1),
   c) defining concept relations across standard vocabularies and ontologies;
2. defines domains, ranges, and datatypes for video properties, which inherently add rich semantics about concept specificity, role applicability, and permissible values for roles, as opposed to vocabularies that declare terms with just some or none of these;
3. captures fundamental correlations and rules for reasoning over video metadata;
4. integrates the RCC8 Region Connection Calculus (Randell, Cui, & Cohn, 1992) and Allen’s relations (Allen, 1983) for spatiotemporal knowledge representation.

As a result, the Video Ontology is suitable for the knowledge representation and annotation of video characteristics as well as high-level semantics, upon which multimedia reasoning and Linked Open Data (LOD) interlinking can be performed. The Video Ontology was designed to act as a mediator between upper ontologies, domain ontologies, and application ontologies used for the semantic enrichment of videos, automated scene interpretation, and video understanding.

Formal grounding

OWL (2) DL ontologies are implementations of general DLs that are decidable fragments of first-order logic (FOL) (Sikos, 2015), as opposed to spatial, temporal, and fuzzy DLs, many of which implement formalisms beyond FOL and are not decidable.

The data model of the Video Ontology has been formalized in the very expressive but decidable SROIQ(D) description logic, thereby exploiting all constructors of OWL 2 DL, and can efficiently model concepts, roles, and individuals in the form of countably finite and pairwise disjoint sets of atomic concepts (NC), atomic roles (NR), and individual names (NI), denoted by Internationalized Resource Identifiers (IRIs), i.e., a string of Unicode characters of the form `scheme://[user:password@host[:port]]//path[?query][#fragment].`

Using the atomic concepts in compound statements, complex concept expressions can be formed. The SROIQ description logic supports a wide range of concept expression constructors, including concept assertion, conjunction, disjunction, complement, top concept, bottom concept, role restrictions (existential and universal restrictions), number restrictions (at-least and at-most restrictions), local reflexivity, and nominals, formally

$$C ::= NC \mid (C \sqcap D) \mid (C \sqcup D) \mid \neg C \mid T \mid \bot \mid \exists R.C \mid \forall R.C \mid \geq n R.C \mid \leq n R.C \mid \exists R.Self \mid \{NI\},$$

where C represents concepts, R represents roles, and n is a nonnegative integer.

In SROIQ, properties and relationships can be described not only with atomic roles, but also with inverse roles and the universal role.

Axioms

Using the previous sets, logical statements can be expressed as axioms. The description logic axioms of the Video Ontology have been implemented in OWL 2, based on the transparent translation between DL and OWL, as demonstrated in Table 1.

In contrast to previous video ontologies, the Video Ontology utilizes not only terminological and assertional axioms, but also role box axioms. As you will see, constructors not used in previously released multimedia ontologies, in particular those typical to role box axioms, are in fact very important, because they significantly extend the application potential of the ontology in video data integration, knowledge management, and ontology-based automated reasoning.
The concepts and roles of the Video Ontology have been defined in a hierarchy that incorporates de facto standard structured definitions. The Video Ontology captures the terminological knowledge of video authoring and video timeline structure by defining the relationships of classes and properties as subclass axioms and subproperty axioms, and class unions, intersections, and equivalence. For example, TBox axioms in the Video Ontology express that keyframes are frames and that movie scenes are equivalent to cinematic scenes (see Table 2).

Similar to other OWL ontologies, the terms of the Video Ontology can be deployed in fully featured knowledge representations in any RDF serialization, such as Turtle and RDF/XML, and as lightweight markup annotations in RDFa, JSON-LD, and HTML5 Microdata (Sikos, 2011).

The Video Ontology exploits that conditional statements can be expressed as disjunction (Whitehead & Russell, 1925), as the truth table of ‘IF p THEN q’ and ‘NOT p OR q’ are identical, i.e.,

\[
p \rightarrow q = (\sim p \lor q)
\]

This approach made it possible to formally define core video production knowledge without Semantic Web Rule Language (SWRL) rules, such as by stating the following:

- if a disc release is a Blu-Ray release, the video mode is either 720p, 1080i, or 1080p;
- if a disc release is an Ultra HD Blu-ray disc release, it uses the 2160p video mode;
- if the video mode of a video is 2160p, the frame height is 2160 pixels or less (depending on the aspect ratio of the video).

### TBox axioms

The concepts and roles of the Video Ontology have been defined in a hierarchy that incorporates de facto standard structured definitions. The Video Ontology captures the terminological knowledge of video authoring and video timeline structure by defining the relationships of classes and properties as subclass axioms and subproperty axioms, and class unions, intersections, and equivalence. For example, TBox axioms in the Video Ontology express that keyframes are frames and that movie scenes are equivalent to cinematic scenes (see Table 2).

Similar to other OWL ontologies, the terms of the Video Ontology can be deployed in fully featured knowledge representations in any RDF serialization, such as Turtle and RDF/XML, and as lightweight markup annotations in RDFa, JSON-LD, and HTML5 Microdata (Sikos, 2011).

The Video Ontology exploits that conditional statements can be expressed as disjunction (Whitehead & Russell, 1925), as the truth table of ‘IF p THEN q’ and ‘NOT p OR q’ are identical, i.e.,

\[
p \rightarrow q = (\sim p \lor q)
\]

This approach made it possible to formally define core video production knowledge without Semantic Web Rule Language (SWRL) rules, such as by stating the following:

- if a disc release is a Blu-Ray release, the video mode is either 720p, 1080i, or 1080p;
- if a disc release is an Ultra HD Blu-ray disc release, it uses the 2160p video mode;
- if the video mode of a video is 2160p, the frame height is 2160 pixels or less (depending on the aspect ratio of the video).

### ABox axioms

The Video Ontology defines individuals and their relationships as a finite collection of concept assertion axioms of the form $C(a)$, role assertion axioms of the form $R(a,b)$, negated role assertions of the form $\neg R(a,b)$, equality statements of the form $a \approx b$, and
inequality statements of the form $a \not\approx b$, where $a, b \in N_i$ individual names, $C \in C$ represents concept expressions, and $R \in R$ represents role expressions, as demonstrated in Table 3.

**RBox axioms**

Most multimedia ontologies described in the literature correspond to formalisms in which the knowledge base is defined as the combination of a TBox and an ABox (Sikos, 2017b). RBox axioms are supported in very expressive DLs only, such as $SROIQ$, to collect all statements related to roles and the interdependencies between roles, including a role hierarchy, a finite collection of generalized role assertion axioms of the form $R \subseteq S$, role equivalence axioms of the form $R \equiv S$, complex role inclusions of the form $R_1 \circ \ldots \circ R_n \subseteq S$, asymmetric role declarations of the form $\text{Asymmetric}(R)$, reflexive role declarations of the form $\text{Reflexive}(R)$, and disjoint role declarations of the form $\text{Disjoint}(R, S)$, where $R, R_i, a n d S$ represent roles, and transitivity axioms of the form $R^+ \subseteq R$, where $R^+$ is a set of transitive roles. These axioms are vital for video scene interpretation and video understanding, because they can capture rich semantics about the complex properties and relationships of video contents. Some examples are shown in Table 4.

**Increasing expressivity**

While $SROIQ^{(D)}$, the description logic behind OWL 2 DL is very expressive, it is not adequate for spatiotemporal annotation and the formal description of video events. Although there are spatial and temporal DLs that are suitable for describing spatial relations between regions of interest (ROI), timepoints, and time intervals, most of them utilize constructors that cannot be implemented in OWL 2 DL, and many of them are not decidable (Sikos, 2018a). Therefore, they were not implemented in the Video Ontology. Instead, ontologies such as the OWL-Time Ontology,5 the SWRL Temporal Ontology,6 and the 4D-Fluent Ontology (Welty & Fikes, 2006), and—even though standard OWL 2 DL was

| DL Syntax | Turtle Syntax |
|-----------|---------------|
| MovieCharacter(ZORRO) | video:Zorro a video:MovieCharacter . |
| portrayedBy(ZORRO, GUYWILLIAMS) | video:Zorro video:portrayedBy video:GuyWilliams . |
| ZORRO $\approx$ DONDIEGODELAVEGA | video:Zorro owl:sameAs video:DonDiegoDeLaVega . |
| ZORRO $\not\approx$ CARLOSMARTINEZ | video:Zorro owl:differentFrom video:CarlosMartinez . |

### Table 3. Asserting and describing individuals with ABox axioms.

| DL Syntax | Turtle Syntax |
|-----------|---------------|
| covers $\subseteq$ topologicallyRelated | video:covers rdfs:subPropertyOf video:topologicallyRelated . |
| Disjoint(protagonistOf, supportingActorOf) | video:protagonistOf owl:propertyDisjointWith video:supportingActorOf . |
| basedOn $\circ$ basedOn $\subseteq$ basedOn | video:basedOn a owl:TransitiveProperty . |
preferred—languages, such as tOWL (Milea, Frasincar, & Kaymak, 2012) and SOWL (Batsakis & Petrakis, 2011) have been analysed to find the best possible balance between expressivity and reasoning complexity. You see, the DL expressivity of the SWRL Temporal Ontology is just $\mathcal{ALC}^{(D)}$, and the expressivity of OWL-Time is $\mathcal{SROIN}^{(D)}$, so they are both decidable. Using the experiences from these ontologies, spatial and temporal relations have been defined in the Video Ontology with standards alignment using general description logic axioms without breaking decidability.

However, the formal description of video events needs more than just spatiotemporal segmentation, namely, rule-based mechanisms. While some OWL 2 axioms correspond to rules, such as class inclusion and property inclusion, some classes can be decomposed as rules, and property chain axioms provide rule-like axioms, there are statements that cannot be expressed in OWL 2. For example, a rule head with two variables cannot be represented as a subclass axiom, and a rule body that contains a class expression cannot be described as a subproperty axiom. To add additional expressivity to OWL 2 DL, ontologies can be extended with rule-based formalisms, such as SWRL rules. Such rules, however, might break decidability. For this reason, the Video Ontology employs DL-safe rules (Motik, Sattler, & Studer, 2005) wherever possible by restricting rule variables to asserted individuals. In the Video Ontology, SWRL rules are only used to formalize complex video events, which are suitable for, among other things, recognizing violent scenes for automated age rating and finding a particular type of video event in different movies to provide customized movie recommendations.

Evaluation

The integrity of the Video Ontology was confirmed using industry-leading reasoners, namely HermiT$^7$ and FaCT++,$^8$ to ensure that the ontology is meaningful, correct, minimally redundant, and richly axiomatized. The semantics of the ontology concepts and roles have been evaluated by comparing 100 video scenes from well-established video datasets, including the 2017 TRECVID Multimedia Event Detection dataset,$^9$ feature films, including movie clips from Hollywood$^2,$$^{10}$ randomly selected YouTube videos,$^{11}$ and camera recordings. The scenes have been described formally using the well-established TV-Anytime Content Ontology,$^{12}$ the Core Ontology for Multimedia (COMM),$^{13}$ the EBU Core Ontology,$^{14}$ the Large Scale Concept Ontology for Multimedia (LSCOM),$^{15}$ the Linked Movie Database (LMD),$^{16}$ the Ontology for Media Resources 1.0,$^{17}$ Schema.org,$^{18}$ as well as the Video Ontology. In order to be application- and domain-independent, the comparison considered automatically extractable low-level descriptors, manually encoded high-level semantics,$^{19}$ administrative metadata, and spatiotemporal annotation.

The low-level video descriptions were aimed at capturing both visual (dominant color of scenes, camera motion) and audio features (loudness, harmonic-noise ratio, silence), as well as automatically detectable video file characteristics (video frame rate, audio sampling rate). The high-level descriptions covered the semantics of 50 concepts, such as man, horse, stone, tree, building, car, etc., and action-describing roles such as ‘is looking at,’ ‘is talking to,’ and ‘is holding.’ The administrative metadata included concepts such as director, country, writer, and scene, and roles such as ‘based on,’ ‘written by,’ ‘filmed at,’ ‘directed by,’ and ‘narrated by.’ The spatiotemporal annotation considered temporal segmentation, ROI, duration, and scene transitions.
To compare the performance of the aforementioned ontologies, the ratio of captured and intended semantics has been calculated as follows. The list of all concepts and roles have been enumerated per category, and the number of represented concepts and roles have been divided by the total number of concepts and roles (see Table 5).

During the evaluation, for each concept and role which could be captured using a particular ontology, the number of successfully captured terms has been increased by 1. Only accurate semantics were considered, making the presence of a term in an ontology a necessary but not sufficient requirement. For example, COMM incorrectly defines several MPEG-7 properties as classes, such as `comm:dominant-color-descriptor` and `comm:edge-histogram-descriptor`, which prevents it from capturing the corresponding property values, because classes can have subclasses or individuals only, but not datatype values. In contrast, the Video Ontology integrates correct MPEG-7 property definitions.

While there are several multimedia ontologies that can be used for describing administrative metadata, there are huge differences in terms of precision, specificity, and the availability of constraint definitions. The evaluation clearly indicates that currently the Video Ontology is the only purposefully designed video ontology for spatiotemporal annotation, and is aligned with the de facto standard SWRL Temporal Ontology. As a result, none of the other ontologies could compete with the Video Ontology when annotating moving ROI and video positions, as reflected in Table 5.

To demonstrate the spatiotemporal annotation capabilities of the Video Ontology, one of the video scenes of the evaluation is briefly presented here, which is a video scene of a medical procedure, namely, Nd:YAG laser vitreolysis. This procedure is used to treat bothersome vitreous opacities, which negatively affect the quality of life (Shah & Heier, 2017). The selected scene was described using spatiotemporal segmentation, which involved the annotation of the region of interest, i.e., an eye floater, as a moving region, and the description of the medical video scene and the concept depicted in the scene. By using standard Media Fragment URIs, the spatiotemporal segmentation was performed by specifying the positions in Normal Play Time format (as defined in RFC 2326), and the ROI was represented by its minimum bounding box, as shown in Figure 2.

This scene was described in Turtle as follows:

| Table 5. Comparison of ontologies for video scene representation. |
|---------------------------------------------------------------|
| TV-Anytime | COMM | EBU Core | LSCOM | LMD | Ontology for Media Resources | Schema.org | VidOnt |
| High-level scene semantics | | | | | |
| 6.25% | 14.06% | 31.25% | 21.88% | 13.87% | 18.75% | 12.51% | 68.75% |
| Low-level feature description | Not captured | 16.72% | 14.29% | Not captured | Not captured | 4.97% | Not captured | 25.14% |
| Administrative metadata | | | | | |
| 6.25% | Not captured | 12.53% | Not captured | 37.64% | 15.62% | 59.38% | 93.75% |
| Spatiotemporal annotation | Not captured | 55.66% | 44.16% | Not captured | Not captured | 65.61% | 50.28% | 92.03% |
| Total | 3.13% | 21.61% | 25.56% | 21.88% | 25.76% | 26.24% | 30.54% | 69.92% |
Among the eight ontologies in the comparison, only the Video Ontology could identify the video from which this scene was derived, not to mention the relationship between the

Figure 2. Spatial annotation of the region of interest depicting a Weiss ring floater by the top left corner coordinates and dimensions of its minimum bounding box. Video by James H. Johnson. https://www.youtube.com/watch?v=hZibbnjiGuY.
video and its spatiotemporal segment that depicts the ROI. Note that highly specialized domain knowledge have been captured using other ontologies as usual, although these were used the same way for all the ontologies in the evaluation, therefore they did not affect the outcome of the comparison. The formal definition of the medical term used in this video scene description was retrieved from SNOMED CT and DBpedia. To maximize interoperability, terms of the MPEG-7 Ontology\textsuperscript{23} and the SWRL Temporal Ontology were reused for spatiotemporal segmentation, and XSD for datatype declarations. Based on the formal descriptions, new statements have been inferred using the RDFS reasoning rules (Hayes, 2014), the Ter Horst reasoning rules (Ter Horst, 2005), and the OWL reasoning rules (Calvanese et al., 2012). Using the formalism outlined above, the YAG laser vitreolysis scene has been identified, and the complex video event successfully recognized via reasoning.

**Conclusions and future work**

The majority of upper and domain ontologies introduced for representing multimedia contents in the last decade and a half, with or without MPEG-7 alignment, lack complex role inclusion axioms and rule-based definitions of common video events, and are limited to highly specific terminological and assertional knowledge. For this reason, most of these formalisms are actually controlled vocabularies or taxonomies only, and not fully featured ontologies, hence they are not suitable for advanced multimedia reasoning. To address the above issues, concepts, roles, individuals, and relationships of professional video production, video timeline structure, and common video events have been formally modeled using \textit{SROIQ}\textsuperscript{(7)}, one of the most expressive decidable DLs, and implemented in OWL 2. The resulting ontology, the Video Ontology (VidOnt), features a vocabulary that has been aligned with standards such as MPEG-7, EBU CCDM, and Dublin Core, and defines concepts in a novel taxonomical structure. This helps the implementation of de facto standard ontologies for video scene interpretation, and provides core terms not defined by any other multimedia ontology. To push the expressivity boundaries of its logical underpinning, the Video Ontology employs rule-based formalisms, while taking into account the importance of keeping reasoning complexity to a minimum.

The Video Ontology described in this paper is the very first ontology that provides a comprehensive set of machine-interpretable concept definitions and axioms for mainstream and cutting-edge video technologies and standards, such as the 4K UHD video mode and the H.265/MPEG-H HEVC codec. This enables the machine-interpretability of the correlations between video mode and scanning, standard disc releases and video resolution, aspect ratio, maximum frame height, etc. The technical properties defined in the Video Ontology are suitable for describing the services of video streaming aggregators, video sharing portals, and broadcasting, such as arbitrary video streams and video files, e.g., videos on Netflix, YouTube, and Vimeo, as well as TV programs.

The efficiency and implementation potential of the presented ontology have been evaluated, and the comparison with other multimedia ontologies gave promising results for capturing low-level video features, high-level scene semantics, technical characteristics of video files, administrative metadata, and spatiotemporal annotation.

Beyond straightforward implementations in video scene representation, indexing, and retrieval, the Video Ontology can also be used to improve existing classifications by
identifying movie data incorrectly entered in a video database, and prevent incorrect data entries in the future. In fact, using the definitions of the ontology may potentially be used to add a dropdown list of all the possible options in video databases for entering video characteristics. In addition, the ontology is suitable for querying video descriptions and performing high-level video scene interpretation via reasoning.

The optimization and extension of the implemented video event formalism and the definition of further concepts and video event types have been scheduled as a future work.

Notes
1. Web Ontology Language.
2. http://3dontology.org.
3. http://videoontology.org.
4. 2160p videos with an aspect ratio wider than 1.78:1, such as 4K Blu-ray releases with an aspect ratio of 2.40:1, are usually called 2160p videos as well, despite that their frame height is actually smaller than 2,160 pixels.
5. http://www.w3.org/2006/time.
6. http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl.
7. http://www.hermit-reasoner.com.
8. http://owl.man.ac.uk/factplusplus/.
9. https://www.nist.gov/itl/iald/mig/med-2017-evaluation.
10. http://www.dl.ens.fr/~laptev/actions/hollywood2/.
11. https://www.youtube.com.
12. http://rhizomik.net/ontologies/2005/03/TVAnytimeContent.owl.
13. http://multimedia.semanticweb.org/COMM/visual.owl.
14. https://www.ebu.ch/metadata/ontologies/ebucore/20150122/ebucore_2015_01_22.rdf.
15. http://lscom.linkeddata.es/LSCOM.owl.
16. http://data.linkedmdb.org/all/.
17. http://www.w3.org/ns/ma-ont.rdf.
18. http://schema.org/.
19. LOD concept definitions have not been used in the comparison, because they can be used to complement any ontology.
20. Neodymium-doped yttrium aluminum garnet (Nd:Y₃Al₅O₁₂).
21. https://www.w3.org/TR/media-frags/.
22. https://www.ietf.org/rfc/rfc2326.txt.
23. http://mpeg7.org.

Disclosure statement
No potential conflict of interest was reported by the author.

Notes on contributor
Leslie F. Sikos, Ph.D., is an internationally recognized expert in multimedia knowledge management and ontology-based automated reasoning. He has worked in both academia and the enterprise, and acquired hands-on skills in the implementation of Semantic Web standards and multimedia semantics. Dr Sikos is an invited author, reviewer, editor, and speaker, and the author of the monograph ‘Description Logics in Multimedia Reasoning.’ He has developed a novel spatiotemporal formalism for video indexing and two of the most expressive multimedia ontologies with MPEG-7 and X3D alignment. Inspired by the creation and exploitation of rich LOD datasets, Dr Sikos actively contributes to the development of open standards and open data repositories.
ORCID

Leslie F. Sikos  http://orcid.org/0000-0003-3368-2215

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26, 832–843. doi:10.1145/82358434

Ballan, L., Bertini, M., Del Bimbo, A., & Serra, G. (2010). Semantic annotation of soccer videos by visual instance clustering and spatial/temporal reasoning in ontologies. *Multimedia Tools and Applications*, 48, 313–337. doi:10.1007/s11042-009-0342-4

Batsakis, S., & Petrakis, E. G. M. (2011). SOWL: A framework for handling spatio-temporal information in OWL 2.0. In N. Bassiliades, G. Governatori, & A. Paschke (Eds.), *Rule-based reasoning, programming, and applications* (pp. 242–249). Heidelberg: Springer. doi:10.1007/978-3-642-22546-8_19

Calvanese, D., Carroll, J., De Giacomo, G., Hendler, J., Herman, I., Parsia, B., … Schneider, M. (2012). Reasoning in OWL 2 RL and RDF graphs using rules. In B. Motik, B. C. Grau, I. Horrocks, Z. Wu, A. Fokoue, & C. Lutz (Eds.), *OWL 2 Web ontology language profiles* (2nd ed.). Retrieved from https://www.w3.org/TR/owl2-profiles/#Reasoning_in_OWL_2_RL_and_RDF_Graphs_using_Rules

Dasiopoulou, S., Heinecke, J., Saathoff, C., & Strintzis, M. G. (2007). Multimedia reasoning with natural language support. Paper presented at First IEEE international conference on semantic computing (pp. 413–420). doi:10.1109/ICSC.2007.28

D’Odorico, T., & Bennett, B. (2013, May). Automated reasoning on vague concepts using formal ontologies, with an application to event detection on video data. Paper presented at the 11th International Symposium on Logical Formalizations of Commonsense Reasoning, Ayia Napa.

Elleuch, N., Zarka, M., Ammar, A. B., & Alimi, A. M. (2011). A fuzzy ontology-based framework for reasoning in visual video content analysis and indexing. Paper presented at the 11th International workshop on multimedia data mining, Article No. 1. doi:10.1145/2237827.2237828

Gómez-Romero, J., Patricio, M. A., García, J., & Molina, J. M. (2011). Ontology-based context representation and reasoning for object tracking and scene interpretation in video. *Expert Systems with Applications*, 38, 7494–7510. doi:10.1016/j.eswa.2010.12.118

Hayes, P. (2014). RDFS entailment rules. Retrieved from https://www.w3.org/TR/2004/REC-rdf-mt-20040210/#RDFSRules

Hitzler, P., Krötzsch, M., & Rudolph, S. (2009). *Foundations of semantic web technologies*. Boca Raton, FL: CRC Press.

Horst, T. J. H. (2005). Completeness, decidability and complexity of entailment for RDF schema and a semantic extension involving the OWL vocabulary. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3, 79–115. doi:10.1016/j.wsweb.2005.06.001

Jaimes, A., Tseng, B. L., & Smith, J. R. (2003). Modal keywords, ontologies, and reasoning for video understanding. In E. M. Bakker, M. S. Lew, T. S. Huang, N. Sebe, & X. S. Zhou (Eds.), *Image and video retrieval* (pp. 248–259). Heidelberg: Springer. doi:10.1007/3-540-45113-7_25

Meghini, C., Sebastiani, F., & Straccia, U. (1997). Reasoning about form and content of multimedia objects. In A. Hauptmann & M. Witbrock (Eds.), *Intelligent integration and use of text, image, video, and audio corpora*. Papers from the 1997 AAAI Spring Symposium (pp. 89–94). Menlo Park: AAAI Press.

Milea, V., Frasincar, F., & Kaymak, U. (2012). tOWL: A temporal web ontology language. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 42, 268–281. doi:10.1109/TSMCB.2011.2162582

Möller, R., & Neumann, B. (2008). Ontology-based reasoning techniques for multimedia interpretation and retrieval. In Y. Kompatsiaris & P. Hobson (Eds.), *Semantic multimedia and ontologies* (pp. 55–98). London: Springer. doi:10.1007/978-1-84800-076-6_3

Motik, B., Sattler, U., & Studer, R. (2005). Query answering for OWL-DL with rules. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3, 41–60. doi:10.1016/j.wsweb.2005.05.001

Randell, D. A., Cui, Z., & Cohn, A. G. (1992). A spatial logic based on regions and connections. In B. Nebel, C. Rich, & W. Swartout (Eds.), *Principles of knowledge representation and reasoning* (pp. 165–176). San Mateo, CA: Morgan Kaufmann.
Shah, C. P., & Heier, J. S. (2017). YAG laser vitreolysis vs sham YAG vitreolysis for symptomatic vitreous floaters: A randomized clinical trial. *JAMA Ophthalmology, 135*, 918–923. doi:10.1001/jamaophthalmol.2017.2388

Sikos, L. F. (2011). Advanced (X)HTML5 metadata and semantics for Web 3.0 videos. *DESIDOC Journal of Library and Information Technology, 31*, 247–252. doi:10.14429/djlit.31.4.1105

Sikos, L. F. (2015). Mastering structured data on the semantic web: From HTML5 microdata to linked open data. New York, NY: Apress. doi:10.1007/978-1-4842-1049-9

Sikos, L. F. (2016a). A novel approach to multimedia ontology engineering for automated reasoning over audiovisual LOD datasets. In N. T. Nguyen, B. Trawiński, H. Fujita, & T.-P. Hong (Eds.), *Intelligent information and database systems* (pp. 3–12). Heidelberg: Springer. doi:10.1007/978-3-662-49381-6_1

Sikos, L. F. (2016b). RDF-powered semantic video annotation tools with concept mapping to linked data for next-generation video indexing: A comprehensive review. *Multimedia Tools and Applications, 76*, 14437–14460. doi:10.1007/s11042-016-3705-7

Sikos, L. F. (2016c). Rich semantics for interactive 3D models of cultural artifacts. In E. Garoufallou, I. Subirats, A. Stellato, & J. Greenberg (Eds.), *Metadata and semantics research* (pp. 169–180). Cham: Springer. doi:10.1007/978-3-319-49157-8_14

Sikos, L. F. (2017a). 3D model indexing in videos for content-based retrieval via X3D-based semantic enrichment and automated reasoning. Paper presented at 22nd International conference on 3D web technology, Article No. 19. doi:10.1145/3055624.3075943

Sikos, L. F. (2017b). *Description logics in multimedia reasoning*. Cham: Springer. doi:10.1007/978-3-319-54066-5

Sikos, L. F. (2017c). A novel ontology for 3D semantics: Ontology-based 3D model indexing and content-based video retrieval applied to the medical domain. *International Journal of Metadata, Semantics and Ontologies, 12*, 59–70. doi:10.1504/IJMSO.2017.10008658

Sikos, L. F. (2017d). *Utilizing multimedia ontologies in video scene interpretation via information fusion and automated reasoning*. Federated conference on computer science and information systems (pp. 91–98). doi:10.15439/2017F66

Sikos, L. F. (2018a). Spatiotemporal reasoning for complex video event recognition in content-based video retrieval. In A. Hassanien, K. Shaalan, T. Gaber, & M. Tolba (Eds.), *Advances in intelligent systems and computing* (pp. 704–713). Cham: Springer. doi:10.1007/978-3-319-64861-3_66

Sikos, L. F. (2018b). Ontology-based structured video annotation for content-based video retrieval via spatiotemporal reasoning. In H. Kwaśnicka & L. C. Jain (Eds.), *Bridging the semantic gap in image and video analysis*. Cham: Springer. doi:10.1007/978-3-319-73891-8_6

Sikos, L. F., & Powers, D. M. W. (2015). Knowledge-driven video information retrieval with LOD: From semi-structured to structured video metadata. Paper presented at Eighth workshop on exploiting semantic annotations in information retrieval (pp. 35–37). doi:10.1145/2810133.2810141

Simou, N., Saathoff, C., Dasiopoulou, S., Spyrou, E., Voisine, N., Tzouvaras, V., … Staab, S. (2006). An ontology infrastructure for multimedia reasoning. In L. Atzori, D. D. Giusto, R. L. Riccardo, & F. Pereira (Eds.), *Visual content processing and representation* (pp. 51–60). Heidelberg: Springer. doi:10.1007/11738695_8

Simou, N., Tzouvaras, V., Avrithis, Y., Stamou, G., & Kollias, S. (2005, April). *A visual descriptor ontology for multimedia reasoning*. Paper presented at the 6th International workshop on image analysis for multimedia interactive services, Montreux.

Town, C. (2006). Ontological inference for image and video analysis. *Machine Vision and Applications, 17*, 94–115. doi:10.1007/s00138-006-0017-3

Welty, C., & Fikes, R. (2006). A reusable ontology for fluents in OWL. In B. Bennett & C. Fellbaum (Eds.), *Formal ontology in information systems*. Frontiers in Artificial Intelligence and Applications (Vol. 150, pp. 226–236). Amsterdam: IOS Press.

Whitehead, A. N., & Russell, B. (1925). *Principia mathematica* (2nd ed.). Cambridge: Cambridge University Press.