

Proposal of an Ontology for Home Movies

PROPUESTA DE ONTOLOGÍA PARA EL CINE DOMÉSTICO

Received on 03/05/2024 | Accepted on 09/07/2024 | Published on 15/01/2025 https://doi.org/10.62008/ixc/15/01Propue

 María Ángeles López Hernández | Universidad de Sevilla

 ☑ alhernan@us.es |
 Image: https://orcid.org/0000-0001-8562-1575

 Rubén Domínguez Delgado | Universidad de Sevilla
 Image: https://orcid.org/0000-0001-9885-2831

 Elena de la Cuadra de Colmenares | Universidad Complutense de Madrid
 Image: https://orcid.org/0000-0003-3057-2176

Abstract: A proposal for a static ontology is presented for the domain of home movies preserved in film archives, which optimizes the organization processes and the search of this type of collections. Given that there is no standard methodology for designing ontologies, for the construction of our proposal the one formulated by Mendonça and Soares was chosen, for its simplicity and clarity. After watching 300 home movies, a glossary of terms was obtained, from which a taxonomy was built, using a double hierarchical-relational approach, in order to be able to navigate in a more flexible way through these contents. Following the OWL scheme, the classes, subclasses, relationships between classes and axioms of our ontology were defined. Currently, this ontological proposal, the first one which is developed in this field, is in an experimental phase for verification and validation.

Keywords: Film Archiving; Artificial Intelligence; Expert Systems; Knowledge Representation; Ontologies; Home Movies.

Resumen: Se presenta una propuesta de ontología para el dominio del cine doméstico conservado en filmotecas, que optimice los procesos de organización y búsqueda de este tipo de documentos. Al no existir una metodología estándar de diseño de ontologías, se optó por la formulada por Mendonça y Soares, por su simplicidad y claridad. Tras el visionado de 300 películas domésticas se obtuvo un glosario de términos, a partir del cual se construyó una taxonomía, utilizando un doble enfoque jerárquico-relacional, a fin de poder navegar de forma más flexible por los contenidos. Siguiendo el esquema OWL se definieron clases, subclases, relaciones entre clases, y axiomas. Actualmente, esta propuesta ontológica, la primera que se desarrolla en este ámbito, se encuentra en fase experimental para su verificación y validación.

Palabras clave: documentación audiovisual; filmotecas; inteligencia artificial; sistemas expertos; ontologías; cine doméstico.



1. Introduction

The debate on the viability of implementing artificial intelligence in the field of librarianship is by no means new. Since the 1970s, only a few years after research on expert systems got underway in the United States, a number of authors (Smith, 1976; Borko, 1985; Micronet, 1986; Fernández Muñoz, 1988; Andón and Bermúdez, 1988; Suárez Martínez, 1988; Aluri and Riggs, 1990; Agustí, 1990; Aguado, 1990 and 1995; García Figuerola, 1990; Martín, 1994; Montoro, Montoya and Vargas, 1990; Morris, 1992; Ribes, 1994) were already discussing their development (the pioneering field of artificial intelligence) as the next evolutionary step in the quest for a more efficient search and retrieval method for documents housed in libraries and data management centres.

Even though some significant studies have been performed on the use of ontologies in librarianship, documentation, archival science and museums (Vickery, 1997; Currás, 2005; Quílez, 2011; Martín Suquía, 2012; Biagetti, 2016; Hidalgo, Senso, Leiva and Hípola, 2016; Pastor and Llanes, 2017), in the audiovisual field such studies are still exceedingly scarce (e.g. Isaac and Tronci, 2004; Caldera and Sánchez, 2008; López de Quintana, 2020; Pastor Sánchez *et al.*, 2020). Indeed, in the scientific literature review performed for this study it was impossible to encounter any publication offering insights into the use of ontologies in the film archive field and, specifically, that of home movies, an amateur film subclass that, in the words of Odín (2010:39) encompasses films (or videos) «made by a family member about people, events or objects linked, in one way or another, to the history of that family, and preferentially viewed by its members».

Boleslaw Matuszewski, regarded as the father of documentary films and reportage, was already referring to these films, forming part of our documentary heritage, in his pioneering initiative for creating film archives in 1898 (Domínguez-Delgado y López-Hernández, 2019), to which, according to the UNESCO (1980: 155), «access should be made available as far as possible». However, and despite the fact that since the beginning of the new millennium researchers have shown increasingly greater interest in moving images of this sort in such diverse fields as history, anthropology, art and cultural studies (López-López and Alcalde-Sánchez, 2024: 2), the level of access to them in film archives still leaves a lot to be desired. These obstacles should be overcome through technological development, as has already been achieved with other types of film documents and archives (Domínguez-Delgado and López-Hernández, 2016).



The recent progress in the implementation of artificial intelligence thus poses the question of whether it can also be applied to home movies.

It is not so much a question of whether it is possible to employ an expert system for home movie search and retrieval, because this obviously goes without saying, for one of the main uses of such systems is precisely data retrieval.

First and foremost, the intention is to determine the purpose, namely, how expert systems, as opposed to conventional databases, could improve the accessibility of film archives. Secondly, the aim is to identify the most appropriate knowledge representation model for organising film data on a computer, «so that it can be manipulated by the control systems of the knowledge base» (Hartnell, 1985:236).

1.1. Using an Expert System in Film Archives

It is essential to consider the purpose of an expert system to understand the use to which it could be put in film archives. As Génova (2016: 139) observes, it is ultimately the experts in these archives who should decide on whether or not employing such a system would be worthwhile, because «an algorithmic or computational machine cannot decide on what objectives should be pursued because, if so, it would cease to be a machine». In the case at hand, the aforementioned purpose seems clear enough: to enhance the performance of film dissemination, search and retrieval systems, thus resolving the shortcomings of conventional databases.

The disadvantage of the conventional databases most film archives currently employ is that their use of logs for knowledge representation places too many constraints on stored data, to the point that adding a new field or expanding an existing one can be extremely costly. In this connection, Lasala (1994: 28) claims that «the log is an ideal formalism for representing information but the least suitable for representing knowledge». As Rolston (1988: 113) notes,

Historically, software systems have taken what might be called the oracle approach: The user presents a problem, and the system metaphorically «goes away to work the problem» and eventually returns with a response that is to be accepted without explanation (Rolston, 1990: 111).

This problem, together with other also commonplace drawbacks when running searches on databases, underlines the fact that it is high time that film archives moved forward and implemented expert systems, which have many advantages, as evidenced by the literature review (Verdejo, 1987: 53; Lasala,



1994: 33; Rolston, 1988: 5 and 10; Pajares and Santos, 2005: 49; Fernández, 1987: 23; Proenza and Pérez, 2012: 51). For these authors, unlike conventional databases, such systems are reusable, adaptable and easy to maintain, on the one hand, and more efficient, easy to use and to interpret, uniform and user-friendly when performing information searches, on the other.

In light of the foregoing, designing the first ever ontology for representing knowledge in the field of home movies seems to be a relevant research objective. Furthermore, such an ontology would have a favourable impact on data organisation, search and retrieval processes in film management systems (Rosel, Senso and Leiva, 2016: 546) and, consequently, on users interested in such material, insofar as searches based on conceptual schemas obtain better results that are more consistent with the search criteria of those users, while helping them find what they are looking for in an easier and more automatic fashion.

1.2. Knowledge Representation

In this study, knowledge representation is understood as the area of artificial intelligence devoted to knowledge coding, manipulation and interpretation which has focused on the quest for highly structured data organisations for coding the necessary information on a domain. Such organisations are essential for a computer, under a suitable interpretation of content, to perform intelligently (Millán, Cortes and Del Moral, 1992: 28).

Throughout the history of artificial intelligence, different knowledge representation formalisms have been developed (Sierra and Sangüesa, 1992: 115). For Millán, Cortes and Del Moral (1992:28), there are basically three which have since become knowledge medium paradigms: logical schemas and semantic networks; production systems; and, lastly, frames and scripts, as declaratory and procedural schemas of representation.

Likewise, Pajares and Santos (2005: 50-52) have drawn attention to procedural, relational and hierarchical knowledge representation paradigms, the last two being familiar to researchers in the field of documentation.

As to the relational knowledge representation paradigm, there are relational databases storing vast amounts of information which have been designed to facilitate access to one or more specific pieces of information in an expeditious and efficient manner. For retrieving stored data, relational calculi, like, for example, the widely used Structured Query Language (SQL), are used as a data manipulation method.

With respect to the hierarchical knowledge representation paradigm, the clearest examples include thesauruses and languages with a combinatorial

structure. Documents sharing a series of characteristics can be naturally associated in classes or groups. This information processing technique allows to use reasoning algorithms for processing information at different levels of specificity. For instance, a concept like 'genre' can be specified as 'fiction genre' or 'documentary genre'. In further detail, it is possible to talk about 'nature documentaries' or 'travel documentaries'. Its automatic implementation usually involves the use of objective-oriented (OO) programming languages like C++ and Java, among others.

As observed by Simons (1987: 125), all these paradigms are fundamental for constructing a variety of expert systems with different aims and levels of complexity. According to Rolston (1990:32), this is the reason why there is no representation technique universally accepted as the 'best', but that in each case it is first necessary to assess the area in question before deciding on the most applicable schema.

As also indicated by Simons (1987: 113), given that different types of problems require different modes of reasoning and that each mode needs to be adequately represented, the first step should be to analyse the most appropriate knowledge representation method or methods, in terms of efficiency, for film documentaries and, in particular, home movies. Therefore, for Rolston (1988: 165) it is important to «select a knowledge representation model as soon as possible, even though it may not be the optimal (or final) representation. Such an early selection is important because the KE [knowledge engineering] must have some way to preserve the knowledge once it is acquired». Enquiring into the best method for representing knowledge is thus an essential first step for developing and ensuring the success of an expert system.

In the case at hand, given the nature of the domain in question, it was decided that the most appropriate theoretical-conceptual knowledge representation paradigm was the relational or, in other words, the declaratory schema of representation, for being, among other reasons, a paradigm with which documentologists and documentalists are familiar.

1.2.1. Semantic Networks

A semantic network is a model based on multiple non-linear associations between basic conceptual elements whose data can be electronically processed.

Semantic networks are used for defining the meaning of a concept in terms of its association with others. This pioneering object modelling technique was proposed by Quillian back in 1968, based on previous works of his published before the 1960s (Pajares and Santos, 2005: 69).



Frequently used in intelligent systems, there are many different types of semantic networks but all consist of nodes and arcs (Simons, 1987: 121), the former normally represented by circles or boxes, indicating objects, concepts or situations in a specific field, and the latter representing the links between nodes.

In this model, the knowledge base is a set of lattices which are modified by inserting or eliminating nodes or manipulating the links between them with arcs (Millán, Cortes and Del Moral, 1992: 41). The nodes and arcs can be labelled in natural language to specify the type of associations between nodes. The most common labels, according to Mylopoulos, Borgida, Jarke and Koubarakis (1992: 328 and ff.) are as follows:

- Generalisation. Relating an object or concept belonging to a more general class. Concepts are organised hierarchically, from the most general to the most specific. Label: 'Is a subset of'.
- Classification. Relating an object or concept to its class. Label: 'Member of'.
- Aggregation. Relating an object or concept to its components. Label: 'Part of'.

Problem solving methods based on this representation are search processes with restrictions associated with the types of arc in the network (Cuena, 1987: 13).

When referring to semantic networks, one of the most important concepts is that of «property inheritance» (Rolston,1988: 49), according to which any property that is true for a class of elements, should also be true for any example of the class. This concept makes semantic networks especially interesting for representing domains which can be structured like taxonomies, namely, they can be organised hierarchically, from the most general terms to the most specific, including related and associated ones.

In the network category, it is possible to find ontologies, understood by García Marco (2007: 543) as a field of research on artificial intelligence and, more specifically, on the branch relating to knowledge representation which governs the construction of expert systems.

1.2.2. Ontologies

The term 'ontology' comes from philosophy, the branch of metaphysics focusing on the study of the nature of existence, beings and their transcendental properties; in philosophy, therefore, an ontology is regarded as a systematic



explanation of existence. Deriving from its original meaning, the term 'ontology' is employed in the field of knowledge engineering, in a more applied and pragmatic way, as a synonym of a structured body of knowledge. It is this last meaning that is used here to establish the associations between ontologies and semantic networks, for in knowledge-based systems, what 'exists' is precisely what can be represented.

According to Gruber (1993: 199), the term 'ontology' was incorporated into the field of artificial intelligence for the purpose of describing computational models capable of supporting automatic reasoning and knowledge capture. It was precisely this researcher who put forward the most well-established declaratory definition of ontology – subsequently developed by other authors (Studer *et al.*, 1998: 186; Rosell *et al.*, 2016: 547) – describing it as an explicit specification of a conceptualisation.

It is thus assumed that an ontology is a description of concepts and relationships in a given domain, described in a language with a formal semantics shared and agreed upon by a community and readable and interpretable for a computer system.

The main components of an ontology are classes (e.g. domestic scenes), subclasses (e.g. family scenes), individuals (e.g. children) and the relationships established between such classes and subclasses. The meaning of a concept is defined by its relationships with other concepts. It is thus possible to infer knowledge by using the most adequate relationships, like, for example, 'Is a subset of', 'Member of', 'Part of', 'Has', 'Belongs to', 'Associated to' and so forth.

In the literature there are different classifications of ontologies according to the approach:

- 1. For Van Heijst, Schereiber and Wielinga (1997: 192), ontologies can be classified in terms of their conceptual structure as follows:
 - Terminological ontologies which specify the terms employed to represent knowledge in the universe of discourse. They are normally used to standardise vocabulary in a particular field.
 - Information ontologies which describe the storage structure of databases. They provide a framework for standard data storage.
 - Knowledge modelling ontologies which specify conceptualisations of knowledge. They contain a rich internal structure and tend to adapt to the particular use of the knowledge describing them.
- 2. Guarino (1998: 9-10) classifies ontologies similarly according to their dependence and association with a task:



- Upper ontologies. They describe more general concepts, such as space, time, matter and objects.
- Domain ontologies. They describe more specific concepts relating to a particular domain (e.g. documentary films).
- Task ontologies. They describe elements and relationships of tasks, activities or artifacts, including components, processes or functions.
- Application ontologies. They usually describe specific concepts that depend on both a specific domain and task.
- 3. In relation to the aspects of the real world which they attempt to model, Jurisica, Mylopoulos and Yu propose the following ontologies (2004: 384-392):
 - Static ontologies. They describe things that exist, their attributes and the associations between them. This classification assumes that the world is full of entities that possess a unique and immutable identity. They use terms like entity, attribute and association.
 - Dynamic ontologies. They describe aspects that can change in the world which they model using finite state machines and Petri nets, among other things. They use terms like process, state and state transition.
 - Intentional ontologies. They describe aspects that refer to the motivations, intentions, goals, beliefs, alternatives and choices of the agents involved. They use terms like aspect, objective, support and agent.
 - Social ontologies. They describe things relating to social aspects, organisational structures, networks and interdependencies. They use terms like actor, position, role, authority and commitment.

For their part, Aranda and Ruiz (2005) propose a more technical specification of each one of these ontological typologies or classifications.

In accordance with these classifications, for this study it was decided to design a static ontology for modelling knowledge for a particular domain that allowed to organise and define a set of concepts in a specific knowledge area (Granados and Rojas, 2011: 102), in this case, home movies.

Lastly, the principal objectives of the ontology proposed here are as follows: (a) to standardise the vocabulary applicable to home movies; (b) to create an association network between concepts that makes the domain of home movies more precise; (c) to facilitate the sharing of knowledge pertaining to the management and search for domestic scenes; (d) to develop a model for



expanding and transforming it in different contexts, facilitating interoperability between existing systems.

The interoperability of ontologies is possible thanks to the use of conceptual schemas or models, including the Resource Description Framework (RDF) (Brickley, 2023), whose semantic extension (RDF Schema) can be used directly for describing an ontology. Objects, classes and properties can be signified. Predefined properties can be used to model relationship instances and subclasses, as well as domain and attribute range restrictions. Another such schema is Web Ontology Language (hereinafter OWL), built on RDF and coded in XML (McGuinness and Van Harmelen, 2004). Likewise, the Simple Knowledge Organisation System (SKOS) is an RDF application that, among other things, allows to identify concepts through uniform resource identifiers (URIs), to label them in several languages, to document them with different types of notes, to interrelate them by means of informal hierarchal structures or associative networks, and to aggregate them to conceptual schemas (Miles and Bechhofer, 2009).

The RDF model provides a standard syntax for developing ontologies, as well as a standard set of modelling primitives as relationship instances and subclasses, allowing to represent information by means of direct lattices in which the vertices have an established meaning and constitute triples. The structure of the RDF triple (subject, resource – predicate, a property – object, a value or literal) allows, with limitations, to enunciate statements on any resource, as stated by Barber *et al.* (2018: 16).

According to Pastor and Llanes (2017:298), with RDF Schema and OWL it is possible to define schemas of ontologies with classes, properties and relationships for providing more precise and specialised descriptions. Ontologies possess a huge capacity for semantic interoperability, thus allowing for complex descriptions of objects and the logical relationships between them.

2. Research Objective and Methodology

The main research objective here is to compile information on recognisable common patterns in home movies for the purpose of designing an ontology for knowledge representation of this specific domain.

It warrants noting that what is involved is a descriptive and exploratory study that does not offer any empirical evidence supporting a previously defined hypothesis.

There is no correct methodology for designing ontologies. Indeed, Senso, Leiva and Domínguez (2011: 335) observe, «There is no sole way of con-



structing an ontology, but the final product should be consistent with its origins. [...] Although it is true that each ontology corresponds to a different way of considering or understanding the state of a specific knowledge, it follows that concepts are just that and it is always possible to employ only particular elements as required.»

Naturally, as new ontologies are designed, new development techniques emerge, which identify the steps that should be followed to construct them. Some authors (e.g. Barber *et al.*, 2018: 23-27; Guzmán, López and Durley, 2012: 135 and ff.; Velásquez, Puentes and Guzman, 2011:214; Hernández and Saiz, 2007: 103 and ff.) agree that the most representative methods include those put forward by Uschold and King, Grüninger and Fox, Lenat and Guha, Noy and McGuinness and Swartout, Patil, Knight and Russ, while also highlighting others, such as KAKTUS, METHONTOLOGY and CommonKADS.

After reviewing those methodologies, it was decided to select Mendonça and Soares' (2017: 49-50) to build the ontology not only because their method includes the main stages of the aforementioned models, but also because of the simplicity, clarity and objectivity of the established steps, which can be summarised as follows:

- 1. Specifying the ontology. The design goal of this ontology was to construct a knowledge representation model of the home movies housed in film archives, with the aim of streamlining the search and retrieval processes of this type of film document.
- 2. Knowledge gain. A combination of two knowledge elicitation or acquisition models were employed: reading manuals and scientific papers on the subject and viewing a large corpus of domestic scenes (n = 300) with a view to identifying the patterns characteristic of this type of film document. It is important to clarify that reference is made here to domestic scenes and not to home movies/videos because different scenes, which are ultimately those that serve to build the ontology, can appear in the same medium. As Gómez Segarra (2008: 85) observes, in those documentary films in which there is no specific purpose other than exhibiting or presenting different situations, it is essential to give each scene its own entity.
- 3. Conceptualising the ontology. In light of the information collected, the domain concepts, subsequently included as classes of the ontology, were identified and analysed. Additionally, knowledge was structured in a class hierarchy, as a graphic conceptual model, to represent the cate-



gories and subcategories (the basis for developing the ontology), as well as the associations between the concepts registered.

4. Representing the ontology. The domain knowledge, previously addressed only at a conceptual level, was then examined at a formal ontological level. In the ontology, the microworld of home movies is described in terms of a lattice – whose root is 'Home movies' – in which nodes and arcs are labelled. Based on their utility, the basic types of node used here included 'concepts' and 'events'. As for the arcs, represented as links between the nodes, following Mylopoulos *et al.* (1992: 328 and ff.), the three most frequent labels were employed: generalisation, classification and aggregation.

A fifth step in Mendonça and Soares' (2017: 49-50) methodology was to evaluate the proposed ontology, requiring to this end both its validation (the extent to which it adapted to the home movie domain) and its verification (an analysis of the adequacy of its construction). Because of space restraints, however, this issue will be addressed in a future study.

Lastly, it is important to note that stratified random sampling was used to select the home movies to be viewed (Mayntz, Holm and Hübner, 1993: 102). In this method, stratification is performed on the basis of an attribute that plays a central role in the research context, in this case, home movies.

The sample included 300 domestic scenes that served to identify patterns, from which two important principles of comparability and classification also derived. When viewing the scenes, generalisations were gradually encountered, interrelating different pieces (Taylor and Bogdan, 2002: 164). In line with De Vega (1989: 348), although all the scenes belong to the same root class of home movies, it is possible to build an ontology by viewing variations on the same theme. In this case, the viewing of 300 domestic scenes was sufficient to obtain an acceptable number of classes and subclasses.

The home movies viewed belong to the following collections: the 'Mi vida' project run by the Film Archive of Andalusia; the home movie archive project run by the MOCA museum; and the 'Memorias Celuloides' project. These collections can be viewed on video platforms like YouTube and Vimeo or, in such an event, on their official websites (i.e. the collection of the Moca Museum¹).

¹ Moca Museum web page available at https://museomoca.com/nuevo-archivo-de-cine-domestico.



3. An Ontology for the Domain of Home Movies

Before describing the ontology, it should be clarified that, in the framework of this study, and as specified in the Spanish Language Dictionary of the Real Academia Española (RAE), the terms 'classes' and 'categories' are employed here as synonyms, which is often the case in the field of expert systems.

To develop an ontology it is essential to define the classes comprising the domain in question, to organise those classes in a taxonomic hierarchy, to define the properties of each class and to indicate the restriction of their values to the properties to create individuals (instances). A class is a set of objects with common properties. The subclasses, which inherit their properties from the class, allow to draw distinctions between properties that can only be associated with a specific subset of the class. Properties are the attributes linked to particular classes/subclasses (Barber *et al.*, 2018:16).

Following the recommendations of Caldera and Sánchez (2008: 89), from the outset it was clear that the proposed ontology had to be straightforward, coherent, adaptable and specific to the domain under study, avoiding any possible ambiguity of the concepts, the imprecise construction of categories and subcategories and the definition of very general relationships. Only in this way is it possible to build an ontology that is easy to use in film archives and, moreover, easy to manage dynamically, with the possibility of adding new concepts or nodes without affecting its basic structure.

Following Ruíz and Ispizua (1989: 197), other basic rules which were complied with when designing the ontology were as follows: (a) that the categories should be mutually exclusive; (b) that they should possess descriptive capacity and should be sufficiently significant; (c) that they should be precise, namely, that the members of a class or subclass should be cognitively differentiable from those of other classes or subclasses, so that analysts should have no doubt about which of them should be included in a specific scene; and (d) that they should be replicable, viz. that two documentalists should be able to include scenes of the same family in the same categories, rather than in different ones. In fact, the semantic validity of an ontology is accepted when several people give the same meaning to the same scene.

Another aspect that should be borne in mind is that it is impossible to represent the real world, or a part of it, in full detail. To reproduce a phenomenon or part of the world, called 'domain' or 'microworld', it is necessary to concentrate or limit the number of concepts relevant enough for creating an abstraction of the phenomenon (Barchini and Álvarez, 2011).



Nor does an ontology necessarily have to include a complete description of all the relevant aspects of a domain. As McGuinnes (cited in García Marco, 2007: 543) contends, ontologies should possess a limited but extendible controlled vocabulary. In other words, it is not a question of encompassing all the available knowledge on each class in question, which, according to Noy and McGuinness (2001: 19), is not recommendable. Only the knowledge that is really indispensable for its proper functioning should be represented, for on the contrary it would be less efficient in that unnecessary information would be shared and the inferential processes would be slower. This implies that, broadly speaking, the ontology should be 'selective', that is, partial and speculative. Of course, any ontology should start with something, signifying that some 'complexities' should be simplified or ignored, in order focus on the essential aspects.

Furthermore, it is important to realise that, unlike conventional databases, ontologies allow for the insertion of new nodes, in addition to other operations such as the connection of concepts, plus the elimination, reduction or simplification of classes and subclasses. This cycle forms part of a process called 'knowledge refinement'.

Lastly, it is worth noting that, as De Vega (1989: 339) states, categories have a diffuse structure, namely, they are not closed or hermetic, for which reason scenes do not have to possess all the attributes assigned to their superclass to fall into it. In fact, these can be 'prototypical' or 'peripheral' scenes in the same category which closely resemble other categories, even though they possess some different attributes. These are the typical scenes that are defined as 'difficult to classify'. One such example would be scenes exclusively of landscapes of places visited which could fall into the 'Trips' or 'Outings' category, within the 'Family scenes' class.

That clarified, after viewing the randomly selected sample of home movies (n = 300), the results indicate that this film genre has a series of distinctive patterns that make it clearly distinguishable from other types of non-fiction film documents. Consequently, it can be claimed that home movies require – as a mode of knowledge representation – a specific ontology. Thanks to the patterns detected, it was possible to determine the principal concepts and the relationships between them, within the specific domain of home movies, the different types of scenes identified being the most noteworthy result.

After creating a glossary of terms based on a controlled vocabulary, a taxonomy of concepts for the ontology was developed. There are several approaches to developing a taxonomy of classes (flat, hierarchal, relational and



faceted taxonomy). In the case at hand, in an initial stage a top-down hierarchical process was followed, starting with the definition of the most general concepts in the domain, followed by that of the more specific ones. In a second stage, transversal relationships between the classes were inserted (relational taxonomy).

Given the nature of the ontology described here, the conceptual model chosen for its development was OWL, a language forming part of the W3C's Semantic Web technology stack and used in the field of artificial intelligence to represent ontologies. As is common knowledge, ontologies defined using OWL have classes, subclasses, relationships between classes and axioms. For a clearer understanding, the work stages are summarised below, with examples of the decisions made.

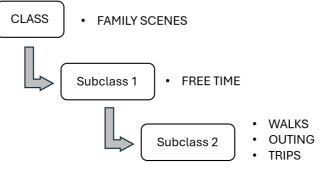
3.1. First Stage: Hierarchical Taxonomy

Following the OWL schema, the main classes were defined, from which more specific classes, called 'subclasses', belonging to a second level, were derived. As Sánchez and Gil (2007: 558) observe, the classes and subclasses are interrelated by a subsumption mechanism, which implies that given a class C with a subclass C1, if m is a member of C1, it also belongs to C. In other words, applying this to the ontology in question, if the class 'Domestic scenes' (C) has a subclass 'Family scenes' (C1), which in turn has another subclass called 'Outings' (C2), it follows that 'Outings' is a type of 'Domestic scene'. This implies that all the attributes inherent to C, are also intrinsic to C1, C2 and so forth, which is commonly defined as an inheritance mechanism.

The level of granularity evinces the level of specificity with which a question can be formulated when using the ontology. In this proposal, granularity affects different levels of subcategories, depending on the concept described, as can be observed in the examples shown in Figures 1 and 2.

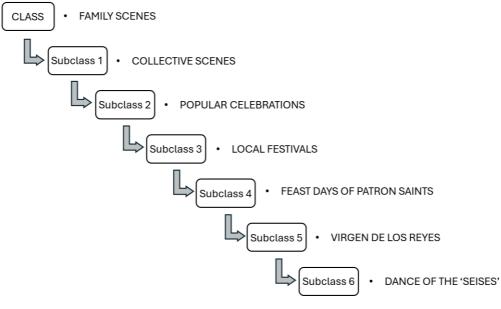


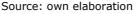




Source: own elaboration

Figure 2. Level of granularity of the proposed ontology. Example 2







3.2. Second Stage: Relationships Between Classes and Subclasses

As already noted, after completing the hierarchical process (hierarchical taxonomy), in the second stage the transversal relationships between classes and subclasses (relational taxonomy) were added, linking the categories not only to their superclasses and subclasses. They could also be associated with categories belonging to other branches of the hierarchical tree with which they might correlate in some way or another. As a result, users can browse and explore content not only vertically in the taxonomy but also transversally and, therefore, in a more flexible manner. This can be seen in the following fragment of the proposed ontology (Figure 3).

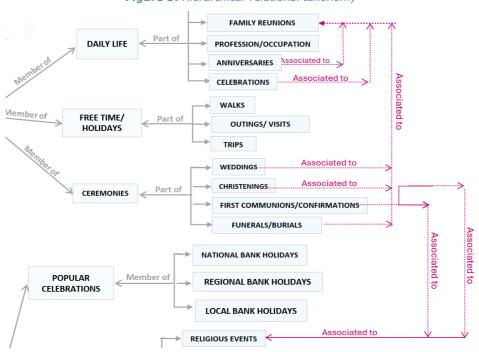


Figure 3. Hierarchical-relational taxonomy

Source: own elaboration

3.3. Description of the Proposed Ontology

As it has been impossible to include all the subcategories of the scenes identified in the viewings owing to space restraints, only a snapshot of the ontology is shown in Figure 1.



The ontology is presented as a model in lattice form (Figure 1), whose root class is 'Home movies'. Nodes like 'Film data', 'Technical data' and 'Content data' represent classes that form its foundations.

The classes have several subclasses which in turn determine other subordinate relationships. The classes and subclasses are defined independently but always under the principle of declaring their conceptual similarity. Each subclass inherits all the properties and operations of its superclass, while also possessing its own (Temesio, 2020: 79).

The colour of the nodes represents the different concepts: green = film data; yellow = technical data; and blue = content data.

The arcs are labelled in natural language to specify how the nodes interrelate with one another, which is essential for avoiding the ambiguity of the terms and concepts. In the proposed ontology, the most frequently used labels are as follows: 'Is a subset of'; 'Member of'; and 'Part of'. Likewise, the label 'Associated to' has been used to symbolise the associations between concepts.

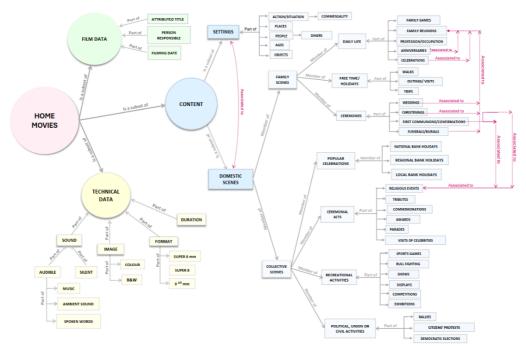


Figure 4. Proposal for an anthology

Source: own elaboration



4. Conclusions

Home movies are becoming increasingly more complex and relevant. However, the concepts used in many home movie archives are defined loosely and in an unstructured way with an ambiguous vocabulary. There have been few initiatives to date aimed at structuring this body of knowledge (Nogales and Suárez, 2010: 94-99).

The main novel contribution of our study has been to provide a framework for future debate on how to use artificial intelligence for resolving problems relating to the documentary retrieval of home movies. It has also been our intention to take an initial approach to the issue that serves to help others to develop ontologies. We have drawn from the premise that, as Noy and McGuinness (2001) would say, there is no one and only ontology for any domain. The design of ontologies is a creative process and no two ontologies designed by different people are alike.

The intention of the ontology we have proposed here is not to perform an in-depth exploration of the domain, namely, not all the terms and concepts inherent to the domain of applied knowledge are employed. This manually designed ontology is based on knowledge gleaned from the different sources consulted and can be enriched with new common knowledge using natural language processing tools or computational learning.

For the benefit of future developers, we believe it is convenient to note the following design decisions:

- a) The ontology does not contain all the information on the domain: we have considered it to be superfluous to include more than is strictly necessary for its application.
- b) The ontology does not contain all the possible subcategories of the classes. Of course, we have not included the subcategories of the more specific levels that a concept can have.
- c) We have not included all the possible relationships between all the terms in our system.

We have designed our ontology in such a way as to make it easy to update and expand with categories or nodes referring to any class that could be detected in the domestic scenes we have analysed. Considering that we work in a field which, as with home movies, is highly dynamic and changeable, this is a fundamental aspect.



We anticipate that the aspects of our study relating to both the observations made and the fact that the ontology can be expanded will prove to be useful for future developments in this regard. Until now, this framework has satisfied our needs. Nonetheless, further research is required to demonstrate the viability of our ontology.

Having said that, we are convinced that our ontology for home movies, despite being the first of its kind in this field and its experimental nature, will provide film archives with a very useful tool for representing information (conceptual schemas), knowledge management and film search and retrieval, in addition to serving as a cooperative system for those institutions housing home movies.

Nevertheless, we are also aware of the obstacles that film archives must overcome when implementing it in the mid-term. Such drawbacks include its cost, for which reason it would be necessary to request (public and private) funding, and the legal restrictions on the public dissemination of this type of domestic footage. This signifies that it would be necessary to convince custodians of the importance of disseminating their footage or, alternatively, to amend regulations or laws so that those private collections that wish to remain in the hands of public institutions should allow all the interested parties access to them.

4.1. Future Lines of Research

The fact that there are no other ones applicable to home movies makes our ontology harder to assess. Therefore, the next step towards consolidating this proposal would be to evaluate it (something we have already started to do), both validating (the ontology's adaptation to the domain of home movies) and verifying it (analysing the ontology with an eye to modifying its construction, if need be).

After its evaluation, the ontology's formal content should be exported to an editor to facilitate the analysis of home movies, to which end it is essential to have an expert system that is reuseable, flexible and easy to maintain.

Lastly, it is important to note that the efficiency of information retrieval can only be measured and evaluated in terms of the pertinence and relevance of the search results obtained by users. It should be recalled that it is the users who know, to a greater or lesser extent, what they are looking for and, therefore, resort to the knowledge base of a computer system to find it. However, if users do not know how to specify their request for information, the expert system should continuously offer them guidelines and instructions until it can



give them the best possible answer. As Lasala (1994: 8) states, «There should be interactive communication between users and systems so that the former can request solutions from the latter, while the latter request data from the former to encounter them».

Likewise, as it is an iterative process (i.e. to fine tune and supplement our proposal), it would be convenient to interview specialists in this domain. Furthermore, as Rolston (1988: 6) indicates, the users of an expert system can also act as testers to verify its validity, detecting failings or glitches.

Lastly, despite the fact that our proposal is still in its infancy, we can confirm that the explorations performed hitherto generally seem to validate and verify the ontology's adaptation to the domain of home movies. In this regard, the representativeness of the domain, as well as the adequacy of both the classes and subclasses specified in the original ontology presented here, has already been verified. For instance, we have reviewed redundant, conflictive or contradictory concepts, non-referenced classes, properties, axioms and so forth.

The next step will be to run reasoning texts for verifying the consistency of the ontology with respect to the relationships established between the concepts, on the one hand, and determining the knowledge that it is capable of inferring, on the other, which we will achieve by employing the semantic reasoners Pellet and FaCT ++, both integrated into the open-source ontology editor Protégé (https://protege.stanford.edu/).



Ethics and transparency

Conflict of Interest

There is no conflict of interest.

Funding

This research has been funded by the R&D Project "Home movies in Spain: preservation, dissemination and appropriation" (2021-2024) (PID2020-115424RB-I00), funded by the Ministry of Science, Innovation and Universities of the Government of Spain.

Author's Contributions

Contribution	Author 1	Author 2	Author 3	Author 4
Conceptualization	Х			
Data curation	Х	Х	Х	
Formal analysis			Х	
Funding acquisition		Х		
Research	Х	Х	Х	
Methodology	Х			
Project administration	Х			
Resources	Х	Х	Х	
Software	Х	Х		
Supervision		X	Х	
Validation		Х	Х	
Visualization	Х	Х	Х	
Writing – original draft	Х			
Writing – review and editing		Х	Х	

4.2. Data Availability Statement

(for data-driven articles)

There is no possibility of accessing the data.

References

- AGUADO, P. M. (1990). Las herramientas inteligentes: una ayuda al usuario. *III Jornadas españolas de Documentación Automatizada*, 450-460.
- AGUADO, P. M. (1995). Los sistemas expertos y la recuperación documental: ejemplos de aplicación, *Scire*, 1(2), 21-32.
- ALURI, R. & RIGGS, D. E. (1990). Expert Systems in libraries. Ablex.
- ANDÓN, V. & BERMÚDEZ, F. (1988): Interfases de lenguaje natural, sistemas expertos y bases de datos. ¿Es posible hablar con el sistema? *Actas del II Congreso Iberoamericano de Informática y Documentación*, 211-225.
- ARANDA, G. & RUIZ, F. (2005). Clasificación y ejemplos del uso de ontologías en Ingeniería del Software. *Workshop en Ingeniería del Software y Bases de Datos WISBD, XI Congreso Argentino de Ciencias de la Computación*. 12 páginas.
- BARBER, E.; PISANO, S.; ROMAGNOLI, S.; DE PEDRO, G.; GREGUI, C.; BLANCO, N. & MOS-TACCIO, M. (2018). Metodologías para el diseño de ontologías Web. *Información, cultura y sociedad, 39*, 13-36.
- BARCHINI A. & ÁLVAREZ-HERRERA M. (2011). Dimensiones e indicadores de la calidad de una ontología. *Revista Avances en Sistemas e Informática*, 7(1), 29-38.
- BEDNAR, K. & WINKLER, T. (2020). Ontologies and knowledge graphs: a new way to represent and communicate values in technology design. *Proceedings of the ETHICOMP 2020*, 203-206.
- BIAGETTI, M. T. (2016) Un modello ontologico per l'integrazione delle informazioni del patrimonio culturale. CIDOC-CRM. JLIS.it: Italian Journal of Library, Archives and Information Science. *Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione, 7*(3), 43-77. https://doi.org/10.4403/jlis.it-11930
- BORKO, H. (1985). Artificial intelligence and expert systems research and their possible impact on information science. *Education for Information. 3*(2), 103-114.
- BRICKLEY, D. (editor) (2023). Lenguaje de descripción de vocabulario RDF: Esquema RDF 1.1. *Recomendación del W3C, 25 de febrero de 2014*. Última versión publicada, consultable en: http://www.w3.org/TR/rdf11schema/
- CALDERA SERRANO, J. & SÁNCHEZ JIMÉNEZ, R. (2008). Ontología para el control y recuperación de información onomástica en televisión. *El Profesional de la Información*, *17*(1), 86-91. https://doi.org/10.3145/epi.2008.ene.10
- CUENA, J. (coord.) (1987). *Inteligencia artificial: Sistemas expertos*, Alianza Editorial.



- 177
- CUEVAS ÁLVAREZ, E. (2010). La casa abierta. El cine doméstico y sus reciclajes contemporáneos. Ayuntamiento de Madrid
- CURRAS, E. (2005). Ontologías, taxonomías y tesauros: manual de construcción y uso. Trea
- DE VEGA, M. (1989). Introducción a la psicología cognitiva. Alianza Editorial.
- DOMÍNGUEZ-DELGADO, R. & LÓPEZ-HERNÁNDEZ, M. A. (2016). The retrieval of moving images at Spanish film archives: The oversight of content analysis. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-4. https://doi.org/10.1002/pra2.2016.14505301140
- DOMÍNGUEZ-DELGADO, R. & LÓPEZ-HERNÁNDEZ, M. A. (2019). In Memoriam Boleslaw Matuszewski. The Origin of Film Librarianship. *Proceedings of the Association for Information Science and Technology*, 56(1), 636-638. https://doi.org/10.1002/pra2.116
- FERNÁNDEZ MUÑOZ, M. T. (1988): La informática en los centros de documentación. *Actas del II Congreso Iberoamericano de Informática y Documentación*, 119-130.
- FERNÁNDEZ, G. (1987). Panorama de los sistemas expertos. En J. CUENA, (coord.). *Inteligencia artificial: Sistemas expertos*. (pp. 23-52). Alianza Editorial.
- GARCÍA FIGUEROLA, C. ET AL. (1990). La catalogación retrospectiva de catálogos mediante un sistema experto. *III Jornadas españolas de Documentación Automatizada*, 784-795.
- GARCÍA MARCO, F. J. (2007). Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. *El profesional de la información*, *16*(6), 541-550. https://doi.org/10.3145/epi.2007.nov.01
- GÉNOVA FUSTER, G. (2016). Máquinas computacionales y conciencia artificial. *Naturaleza y Libertad, 7,* 123-143.
 - https://doi.org/10.24310/NATyLIB.2016.v0i7.6337
- GRANADOS PEMBERTY, E. & ROJAS PINEDA, E. Soluciones organizacionales a partir de ontologías. *Revista Avances en Sistemas e Informática*, 8(1), 101-112.
- GRUBER T. A. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*. 5(2),199-220. https://doi.org/10.1006/knac.1993.1008
- GUARINO, N. (Editor) (1998). Formal Ontology in Information Systems. *Proceedings of FOIS'98, Trento, Italy, 6*(8), 3-15.
- GUZMÁN LUNA, J. A.; LÓPEZ BONILLA, M. & DURLEY TORRES, I. (2012). Metodologías y métodos para la construcción de ontologías. *Scientia et Technica, 50*, 133-140.
- HARTNELL, T. (1985). Inteligencia Artificial: conceptos y programas. Anaya.



- HERNÁNDEZ RAMÍREZ, H. & SAIZ NOEDA, M. (2007). Ontologías mixtas para la representación conceptual de objetos de aprendizaje. *Sociedad Española para el Procesamiento del Lenguaje Natural, 38*, 99-106.
- HIDALGO-DELGADO, Y.; SENSO, J. A.; LEIVA-MEDEROS, A. & HÍPOLA, P. (2016). Gestión de fondos de archivos con datos enlazados y consultas federadas. *Revista española de documentación científica*, 39(3), 1-18. https://doi.org/10.3989/redc.2016.3.1299
- ISAAC, A. & TRONCY, R. (2004). Designing and Using an Audio-Visual Description Core Ontology. *Proceedings of the EKAW*04 Workshop on Core Ontologies in Ontology Engineering.* https://tinyurl.com/2vmm2ccz
- JURISICA, I.; MYLOPOULOS, J. & YU, E. (2004). Ontologies for Knowledge Management: An Information Systems Perspective. *Knowledge and Information Systems*, 6, 380-401.
- LASALA CALLEJA, P. (1994). Introducción a la inteligencia artificial y los sistemas expertos. Prensas Universitarias.
- LÓPEZ-LÓPEZ, J. DE D. & ALCALDE-SÁNCHEZ, I. (2024). Los desplazamientos del cine doméstico en España: el archivo, la práctica artística y la investigación. *Arte, individuo y sociedad, 36*(2), 381-393. https://dx.doi.org/10.5209/aris.91631
- LÓPEZ DE QUINTANA SÁENZ, E. (2020). La documentación en Atresmedia: tecnología y recursos humanos al servicio de la producción de contenidos, *Clip de SEDIC: Revista de la Sociedad Española de Documentación e Información Científica, 81*, 55-64.
- MARTÍN SUQUÍA, R. (2012). El modelo conceptual de las Normas Españolas de Descripción Archivística (CNEDA) y su potencial de evolución hacia el diseño de ontologías archivísticas y la incorporación a la web semántica, *Boletín de la ANABAD, 62*(3), 860-927.
- MARTÍN VEGA, A. (1994). Las redes de neuronas artificiales en la recuperación de información. Algunas fuentes para su estudio. *IV Jornadas españolas de Documentación Automatizada*, 403-410.
- MCGUINNESS, D. L. & VAN HARMELEN, F. (editores) (2004) OWL Web Ontology Language Overview. *W3C Recommendation 10 February 2004*. Latest published version: https://www.w3.org/TR/owl-features/
- MAYNTZ, R., HOLM, K. & HÜBNER, P. (1993). *Introducción a los métodos de la sociología empírica*. Alianza Editorial.
- MENDONÇA, F. M. & SOARES, A. L. (2017). Construindo ontologias com a metodologia ontoforinfoscience: uma abordagem detalhada das atividades do desenvolvimento ontológico. *Ciência da Informação*, 46(1), 43-59. https://tinyurl.com/mr2v2b8a



MICRONET (1986). Chicle: Sistema de recuperación integral de información. *II Jornadas españolas de Documentación Automatizada (Málaga)*, 545-558.

- MILES, A. & BECHHOFER, S. (editores) (2009). SKOS Simple Knowledge Organization System Reference. *W3C Recommendation de 18 August 2009*. Latest published version: http://www.w3.org/TR/skos-reference
- MILLÁN, J.; CORTÉS, U. & DEL MORAL, A. (1992). El arte de la representación. En A. DEL MORAL BUENO, (coord.). *Nuevas tendencias en Inteligencia Artificial* (pp. 25-75). Universidad de Deusto.
- MONTORO AGUILERA, I.; MONTOYA RUIZ, Y. & VARGAS QUESADA, B. (1990). Una visión hacia el futuro: los sistemas expertos. *Cuadernos de la Asociación Nacional de Diplomados de Biblioteconomía y Documentación, 3/4*, 51-62.
- MORRIS, A. (1992). The application of expert systems in libraries and information centres. KG Saur Verlag.
- MYLOPOULOS, J.; BORGIDA, A.; JARKE, M. & KOUBARAKIS, M. (1990). Telos: representing knowledge about information systems. *ACM: Transactions on Information Systems*, 8(4), 325-362.
- NOY, N. F. & MCGUINNESS, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. *Stanford Knowledge Systems Laboratory, 32*, 1-25.
- NOGALES CÁRDENAS, P. & SUÁREZ FERNÁNDEZ, J. C. (2010). Evolución histórica y temática del cine doméstico español. En E. CUEVAS ÁLVAREZ *La casa abierta, El cine doméstico y sus reciclajes contemporáneos.* (pp. 89-117) Ayuntamiento de Madrid
- ODÍN, R. (2010), El cine doméstico en la institución familiar. En E. CUEVAS ÁL-VAREZ. *La casa abierta, el cine doméstico y sus reciclajes contemporáneos* (pp. 39-60). Ayuntamiento de Madrid.
- PAJARES MARTINSANZ, G. & SANTOS PEÑAS, M. (2005). Inteligencia Artificial e Ingeniería del Conocimiento. RA-MA.
- PASTOR-SÁNCHEZ, J. A. & LLANES-PADRÓN, D. (2017). Records in context: el camino de los archivos hacia la interoperabilidad semántica. *Anuario ThinkEPI*, *11*, 297-304. https://doi.org/10.3145/thinkepi.2017.56
- PASTOR SÁNCHEZ, J. A.; SAORÍN, T.; BAZÁN, V.; ESCRIBANO, M. & BAÑOS MORENO, M. J. (2020). Audio-visual Semantics. Propuesta de una ontología para la descripción de secuencias audiovisuales. *Actas del IV Congreso ISKO España-Portugal 2019*, 337-347.
- PROENZA A.; YUNIEL, Y. & PÉREZ SOSA, A. (2012). OntoCatMedia: Ontología para la búsqueda y clasificación automática de medias audiovisuales, *Ciencias de la Información*, 43(3), 49-54.



- QUÍLEZ MATA, J. (2011). Aproximació a les ontologies. Definició i construcció. Aplicacions en el camp de l'arxivística. *Lligall: revista catalana d'Arxivística, 32*, 105-156.
- REAL ACADEMIA ESPAÑOLA: Diccionario de la lengua española, 23.ª ed., [versión 23.7 en línea]. https://dle.rae.es [Consultado el 11 de octubre de 2024].
- RIBES LLOPES, I. (1994). Formación de usuarios: dar un pez o enseñar a pescar, *IV Jornadas españolas de Documentación Automatizada*, 601-609.
- ROLSTON, DAVID W. (1990). Principios de inteligencia artificial y sistemas expertos. McGraw-Hill.
- ROSELL, Y., SENSO, J. A. & LEIVA, A. (2016). Diseño de una ontología para la gestión de datos heterogéneos en universidades: marco metodológico. *Revista Cubana de Información en Ciencias de la Salud, 27*(4), 545-567.
- RUIZ OLABUÉNAGA, J. I. & ISPIZUA, M. A. (1989). La descodificación de la vida cotidiana. Métodos de investigación cualitativa. Universidad de Deusto.
- SÁNCHEZ JIMÉNEZ, R. & GIL URDICIAÍN, B. (2007). Lenguajes documentales y ontologías. *El profesional de la información*, 16(6), 551-560. https://doi.org/10.3145/epi.2007.nov.02
- SENSO, J. A.; LEIVA MEDEROS, A. A. & DOMÍNGUEZ VELASCO, S. E. (2011). Modelo para la evaluación de ontologías. Aplicación en Onto-Satcol. *Revista Española de Documentación Científica*, 34(3), 334-356. https://doi.org/10.3989/redc.2011.3.788
- SIERRA, C. & SANGÜESA, R. (1992). Herramientas de desarrollo de sistemas expertos (p. 111-128). En A. DEL MORAL BUENO. *Nuevas tendencias en Inteligencia Artificial*. Universidad de Deusto.
- SIMONS, G. L. (1987). *Introducción a la Inteligencia Artificial*. Ediciones Díaz de Santos.
- SMITH, L. C. (1976). Artificial intelligence in information retrieval systems. (1976). *Information Processing & Management.* 12(3),189-222.
- STUDER R.; BENJAMINS R. & FENSEL D. (1998). Knowledge Engineering: Principles and Methods. *Data and Knowledge Engineering*, *25*(1y2), 161-197.
- SUÁREZ MARTÍNEZ, D. F. (1988). Responsabilidad de los productores de información automatizada frente a los usuarios. *Actas del II Congreso Iberoamericano de Informática y Documentación*, 621-627.
- TAYLOR, S.J. & BOGDAN, R. (2002). Introducción a los métodos cualitativos de investigación. Paidós.
- TEMESIO VIZOSO, S. (2020). Registros en contexto (RiC): modelo conceptual archivístico. *Informatio 25*(2), 62-91.



- UNESCO (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) (1980). Recomendaciones para la Salvaguardia y Conservación de Imágenes en Movimiento. En *Actas de la Conferencia General. 21ª Reunión. 23 de septiembre – 28 de octubre de 1980. Volumen 1. Resoluciones*, 167-172. UNESCO.
- VAN HEIJST, G.; SCHEREIBER, A.T. & WIELINGA, B. J. (1997). Using Explicit Ontologies in KBS Development. *International Journal of Human and Computer Studies*, 45, 183-292. https://doi.org/10.1006/ijhc.1996.0090
- VELÁSQUEZ PÉREZ, T.; PUENTES VELÁSQUEZ, A. M. & GUZMÁN LUNA, J. A. (2011). Ontologías: una técnica de representación de conocimiento. *Revista Avances en Sistemas e Informática, 8*(2), 211-216.
- VERDEJO, F. (1987). Sistemas basados en reglas de producción y programación lógica (pp. 53-68). En J. CUENA, (coord.). *Inteligencia artificial: Sistemas expertos*, Alianza Editorial.
- VICKERY, B. C. (1997). Ontologies. *Journal of Information Science*, 23(4), 277-286.

