Information Sciences xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A multimedia ontology model based on linguistic properties and audio-visual features

Antonio M. Rinaldi *

DIETI – Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università di Napoli Federico II, 80125 Via Claudio, 21, Napoli, Italy
IKNOS-LAB – Intelligent and Knowledge Systems – LUPT, Università di Napoli Federico II, 80134 Via Toledo, 402, Napoli, Italy

ARTICLE INFO

Article history:
Received 21 September 2011
Received in revised form 17 January 2014
Accepted 9 February 2014
Available online xxxx

Keywords:
Multimedia ontology
OWL
WordNet
Semantic network
P2P

ABSTRACT

The exponential growth of informative contents needs intelligent information systems able to use data to create information. To aim this goal, these systems should have formal models to represent knowledge. In this way complex data can be managed and used to perform new tasks and implement innovative functionalities. This article describes a general and formal ontology model to represent knowledge using multimedia data and linguistic properties to bridge the gap between the target semantic classes and the available low-level multimedia descriptors. This model has been implemented in a system to edit, manage and share ontology in the WEB. The system provides a graphical interface to add multimedia objects by means of user interaction. The multimedia features are automatically extracted using algorithms based on MPEG-7 descriptors.

1. Introduction

Representing knowledge is one of the most important tasks in the information age. The exponential growth of informative contents in several contexts (e.g. internet, enterprise intranet, mobile devices and so on) needs intelligent information systems able to use data to create information.

In the semantic web context, multimedia contents have to be semantically described in order to be discovered and exploited by services, agents and applications. However, bridging the gap between the target semantic classes and the available low-level multimedia descriptors is an unsolved problem. Hence it is crucial to select an appropriate set of multimedia descriptors and to combine the low-level descriptors, so that the results obtained with individual descriptors are improved together with high level concepts annotation. In this perspective, a significant progress has been made on automatic segmentation or structuring of multimedia content and the recognition of low-level features. However, the generation of multimedia content descriptions is highly problematic due to the number and complexity of data and the subjectivity of human-generated descriptions.

New techniques have been developed to solve those problems. Some of them are based on ontologies to delete or at least smooth conceptual or terminological messes and to have a common view of the same information.

This article describes a general and formal ontology model to represent knowledge using multimodal "*signs*" defined as "something that stands for something, to someone in some capacity" [17]. Generally speaking, they include words, images, gestures, scents, tastes, textures, sounds essentially all of the ways in which information can be communicated as a message

* Tel.: +39 0817683911; fax: +39 0817683817.
E-mail address: antoniomaria.rinaldi@unina.it

http://dx.doi.org/10.1016/j.ins.2014.02.017

by any sentient, reasoning mind to another. The proposed approach is based on the use of linguistic properties in order to relate "*signs*" (text and audio-visual features) and "*signifiers*" (concepts). This model has been implemented in a system to edit, manage and share multimedia ontologies in the WEB.

The paper is organized as follow: related works about several aspects of ours are presented and discussed in Section 2; in Section 3 is defined and described the proposed model using OWL; the system architecture and its main functionalities are shown in Section 4; a case study is presented in Section 5 and conclusions and future works are in Section 6.

2. Related works

The management of multimedia objects have had contributions from several fields in the scientific community as artificial intelligence, computational vision, pattern recognition and other disciplines which now define a robust theoretical background. In this perspective, our article gives a systematic introduction to theoretical approaches for knowledge representation using formal models and languages with a discussion on the use of high and low features to interpret and analyze multimedia objects. Moreover several applications based on multimedia ontologies to mine and retrieve multimedia information have been presented and discussed.

2.1. Knowledge modeling and representation

Our approach starts from the *modeling view* of knowledge acquisition [15], where the modeling activity must establish a correspondence between a knowledge base and two separate subsystems: the agent's behavior (i.e., the problem-solving expertise) and its own environment (the problem domain) (see also [22,49,24]). This vision is in contrast with the *transfer view*, wherein a knowledge base is a repository of knowledge extracted from one expert's mind. Using the modeling view approach, knowledge is much more related to the classical notion of truth as correspondence to the real world, and it is less dependent on the particular way an intelligent agent pursues its goals.

Although knowledge representation is a basic step in the whole process of knowledge engineering, a part of the AI research community seems to have been much more interested in the nature of reasoning than in the nature of "real world" representation. This tendency has been especially evident among the disciples called logicist approach: in their well-known textbook on AI, Genesereth and Nilsson [23] explicitly state the "essential ontological promiscuity of AI" and devote just a couple of pages to the issue of conceptual modeling. They admit it is still a serious open problem. The issues of representation are also addressed in the same way in [47].

The dichotomy between reasoning and representation is comparable with the philosophical distinction between epistemology and ontology, and this distinction allows us to better understand our research aim and the proposed approach.

Epistemology can be defined as "the field of philosophy which deals with the nature and sources of knowledge" [43]. According to the usual logicistic interpretation, knowledge consists of propositions, whose formal structure is the source of new knowledge. The inferential aspect seems to be essential to epistemology (at least in the sense that this term assumes in AI): the study of the "nature" of knowledge is limited to its superficial meaning (i.e., the form), since it is mainly motivated by the study of the inference process.

Ontology, on the other hand, can be seen as the study of the organization and the nature of the world independent of the form of our knowledge about it.

2.2. Languages

A basic step in the knowledge engineering process is the use of "tools" to represent knowledge, both for inferring and organizing it. From this point of view, one of the most important advances in the knowledge representation (KR) applications is derived from proposing [40], studying [60,9,10] and developing [11,21,8] languages based on the specification of objects (concepts) and the relationships among them. The main features of all KR languages are the following: (i) *object-orientedness*, for which all the information about a specific concept is stored in the concept itself (in contrast, for example, to rule-based systems; (ii) *generalization/specialization* are basic aspects of the human cognition process [40], the KR languages have mechanisms to cluster concepts into hierarchies where higher-level concepts represent more general attributes than the lower-level ones, which inherit the general concept attributes but are more specific, presenting additional features of their own; (iii) *reasoning* is the capability to infer the existence of information not explicitly declared by the existence of a given statement; (iv) *classification* in which given an abstract description of a concept, there are mechanisms to determine whether a concept can have this description; this feature is a special form of reasoning.

Object orientation and generalization/specialization help human users in understanding the represented knowledge; reasoning and classification guide an automatic system in building a knowledge representation, as the system knows what it is going to represent.

The proposed approach arises from the above considerations and it is also suggested by the work of Guarino [25]. When a KR formalism is constrained in such a way that its intended models are made explicit, it can be classified as belonging to the ontological level [25] introduced in the distinctions proposed in [10], where KR languages are classified according to the kinds of primitives offered to the user.

Please cite this article in press as: A.M. Rinaldi, A multimedia ontology model based on linguistic properties and audio-visual features, Inform. Sci. (2014), http://dx.doi.org/10.1016/j.ins.2014.02.017