

Description Logics in Multimedia Reasoning

Leslie F. Sikos

Description Logics in Multimedia Reasoning

 Springer

Leslie F. Sikos
Centre for Knowledge & Interaction Technologies
Flinders University
Adelaide
Australia

ISBN 978-3-319-54065-8 ISBN 978-3-319-54066-5 (eBook)
DOI 10.1007/978-3-319-54066-5

Library of Congress Control Number: 2017934914

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The immense and constantly growing number of videos urges efficient automated processing mechanisms for multimedia contents, which is a real challenge due to the huge semantic gap between what computers can automatically interpret from audio and video signals and what humans can comprehend based on cognition, knowledge, and experience. Low-level features, which correspond to local and global characteristics of audio and video signals, and low-level feature aggregates and statistics, such as various histograms based on low-level features, can be represented by low-level feature descriptors. These automatically extractable descriptors, such as dominant color and motion trajectory, are suitable for a limited range of applications only (e.g., machine learning-based classification) and are not connected directly to sophisticated human-interpretable concepts, such as concepts depicted in a video, which can be described using high-level descriptors only.

To narrow the semantic gap, feature extraction and analysis can be complemented by machine-interpretable background knowledge formalized with description logics (DL) and implemented in ontology languages, in particular the Web Ontology Language (OWL).

While many knowledge representations employ general-purpose DL, multimedia descriptions, such as the description of moving objects in videos, may utilize spatial, temporal, and fuzzy DL as well. This enables the representation of topological relationships between spatial objects, the representation of the process state and state sequences, quantitative and qualitative spatial reasoning, handling imprecision and uncertainty, and selecting the most probable interpretation of a scene.

The depicted concepts and their relationships are usually described in the Resource Description Framework (RDF), which can express machine-readable statements in the form of subject-predicate-object triples (RDF triples), e.g., scene-depicts-person. The formal definition of the depicted concepts and relationships are derived from controlled vocabularies, ontologies, common sense knowledge bases, and Linked Open Data (LOD) datasets. The regions of interest (ROIs) can be annotated using spatial DL formalisms and media fragment identifiers. The temporal annotation of actions and video events can be performed using temporal

DL and rule-based mechanisms. The fusion of these annotations, including descriptors of different modalities, is suitable even for the machine-interpretable spatio-temporal annotation of complex video scenes.

Based on these structured annotations, various inference tasks can be performed to enable the automated interpretation of images, 3D models, audio contents, and video scenes. For example, video frame interpretation can be performed via abductive reasoning and video event recognition via reasoning over temporal DL axioms. The structured annotation of multimedia contents can be efficiently queried manually and programmatically using the very powerful query language SPARQL, although high-level concept mapping usually requires human supervision and judgment, and automatic annotation options need more research. Research results for high-level concept mapping in constrained videos, such as medical, news, and sport videos, are already promising. The structured description of multimedia contents and the semantic enrichment of multimedia metadata can be applied in video understanding, content-based video indexing and retrieval, automated subtitle generation, clinical decision support, and automated music and movie recommendation engines.

High-level concept mapping relies on formal concept definitions provided by RDFS (RDF Schema) vocabularies and OWL ontologies. However, it is a common and bad practice to create OWL ontologies using a tree structure of concept hierarchies visualized by the graphical user interface of ontology editors, and in particular that of Protégé, without formal grounding in DL.

Without logical underpinning, the computational properties of ontologies remain unclear, and reasoning over RDF statements leveraging formal definitions of ontologies that lack formal grounding might be not decidable. Beyond decidability, a crucial design principle in ontology engineering is to establish favorable trade-offs between expressivity and scalability, and when needed maximize expressivity. However, the higher the DL expressivity, the higher the reasoning complexity. Since the best balance between language expressivity and reasoning complexity depends on the intended application, a variety of DL have been developed, supporting different sets of mathematical constructors.

The level of semantic representation and logical formalization, together with the knowledge base size, presence or absence of instances, and the capabilities of the reasoner, determines what kind of reasoning is feasible. The reasoning tasks utilize different sets of reasoning rules that provide semantics for RDF and RDFS vocabularies, RDFS datatypes, and OWL ontologies, to infer new statements, check knowledge base consistency, determine concept satisfiability, calculate the subsumption hierarchy, and perform instance checking.

The formal knowledge representation of, and reasoning over, depicted objects and events can be used in high-level scene interpretation. Application areas include, but are not limited to, classification, video surveillance, intelligent video analysis, moving object detection and tracking, human intention detection, real-time activity monitoring, next-generation multimedia indexing, and content-based multimedia retrieval.

This book introduces researchers to multimedia semantics by providing an in-depth review of state-of-the-art standards, technologies, ontologies, and software tools. It draws attention to the importance of formal grounding in the knowledge representation of multimedia objects, and the potential of multimedia reasoning in intelligent multimedia applications. It presents both theoretical discussions and multimedia ontology engineering best practices. In this book, the reader familiar with mathematical logic, Internet, and multimedia fundamentals can learn to develop formally grounded multimedia ontologies and map concept definitions to high-level descriptors. The core reasoning tasks, reasoning algorithms, and industry-leading reasoners are also presented, and scene interpretation via reasoning demonstrated. This book aims to illustrate how to use DL-based formalisms to their full potential in the creation, indexing, and reuse of multimedia semantics.

Adelaide, Australia

Leslie F. Sikos

Contents

1	Multimedia Semantics	1
1.1	Rationale	1
1.2	Feature Extraction and Feature Statistics for Classification	3
1.3	Machine Learning for Multimedia Understanding	4
1.4	Object Detection and Recognition	5
1.5	Spatiotemporal Data Extraction for Video Event Recognition	5
1.6	Conceptualization of Multimedia Contents	6
1.7	Concept Mapping	8
1.8	Implementation Potential: From Search Engines to Hypervideo Applications	10
1.9	Summary	10
2	Knowledge Representation with Semantic Web Standards	11
2.1	The Semantic Web	11
2.2	Unstructured, Semistructured, and Structured Data	12
2.3	RDF	14
2.4	RDFS	19
2.5	OWL	20
2.5.1	OWL Variants	21
2.5.2	Modeling with OWL	22
2.5.3	Serialization	30
2.6	Simple Knowledge Organization System	33
2.7	Rule Languages	35
2.7.1	Semantic Web Rule Language	35
2.7.2	Rule Interchange Format	39
2.8	Structured Data Deployment	39
2.8.1	Linked Open Data Datasets	39
2.8.2	Graph Databases: Triplestores and Quadstores	44
2.8.3	Lightweight Annotations	45
2.9	Summary	49

3	The Semantic Gap	51
3.1	Low-Level Descriptors	51
3.1.1	Common Visual Descriptors	52
3.1.2	Common Audio Descriptors	53
3.1.3	Common Spatiotemporal Feature Descriptors, Feature Aggregates, and Feature Statistics	54
3.2	The Discrepancy Between Low-Level Features and High-Level Semantics	55
3.3	Semantic Enrichment of Multimedia Resources	57
3.3.1	Semantic Enrichment of Images	58
3.3.2	Structured 3D Model Annotation	59
3.3.3	Semantic Enrichment of Audio and Video	60
3.4	Summary	66
4	Description Logics: Formal Foundation for Web Ontology Engineering	67
4.1	Description Logics	67
4.1.1	Nomenclature	68
4.1.2	Annotation and Naming Conventions	69
4.1.3	Interpretation	71
4.1.4	DL Constructor Syntax and Semantics	73
4.1.5	DL Axiom Syntax and Semantics	75
4.1.6	TBox, ABox, and RBox	78
4.1.7	Relation to Other Logics	81
4.2	Description Logic Families	83
4.2.1	\mathcal{ALC} and the Basic Description Logics	83
4.2.2	The \mathcal{EL} Family of Description Logics	84
4.2.3	The DL-Lite Family of Description Logics	84
4.2.4	Frame-Based Description Logics (\mathcal{FL})	86
4.2.5	The \mathcal{SH} Family of Description Logics	86
4.2.6	Spatial Description Logics	87
4.2.7	Temporal Description Logics	91
4.2.8	Spatiotemporal Description Logics	102
4.2.9	Fuzzy Description Logics	103
4.3	Extending DL Expressivity	105
4.4	Formal Representation of Images	106
4.5	Formal Representation of 3D Models and Scenes	108
4.6	Formal Representation of Audio	109
4.7	Formal Representation of Video Scenes	112
4.7.1	Spatial Annotation	113
4.7.2	Temporal Annotation	113
4.7.3	Spatiotemporal Annotation	116
4.8	Summary	120

- 5 Multimedia Ontology Engineering** 121
 - 5.1 Introduction to Ontology Engineering 121
 - 5.1.1 Specification 122
 - 5.1.2 Knowledge Acquisition 123
 - 5.1.3 Conceptualization 123
 - 5.1.4 Assessment of Potential Term Reuse 123
 - 5.1.5 Enumerating the Terms of the Knowledge Domain . . . 124
 - 5.1.6 Building the Concept Hierarchy 124
 - 5.1.7 Defining Roles 125
 - 5.1.8 Adding Individuals 126
 - 5.1.9 Creating a Ruleset 126
 - 5.1.10 Evaluation 126
 - 5.1.11 Documentation 127
 - 5.1.12 Maintenance 127
 - 5.2 Ontology Engineering Tools 128
 - 5.2.1 Ontology Editors 128
 - 5.2.2 Ontology Analysis Tools 134
 - 5.3 The Evolution of Multimedia Ontology Engineering 134
 - 5.3.1 Semistructured Vocabularies 135
 - 5.3.2 Structured Ontologies Mapped from Semistructured
Vocabularies 136
 - 5.3.3 Structured Multimedia Ontologies 140
 - 5.4 Ontology-Based Multimedia Annotation Tools 143
 - 5.4.1 Structured Image Annotation Tools 143
 - 5.4.2 Structured Audio Annotation Tools 144
 - 5.4.3 Structured Video Annotation Tools 145
 - 5.4.4 Structured 3D Model Annotation Tools 148
 - 5.5 Summary 149
- 6 Ontology-Based Multimedia Reasoning** 151
 - 6.1 Rationale 151
 - 6.2 Core Reasoning Tasks 152
 - 6.3 Reasoning Rules 155
 - 6.3.1 RDFS Reasoning Rules 155
 - 6.3.2 Ter Horst Reasoning Rules 156
 - 6.3.3 OWL 2 Reasoning Rules 158
 - 6.4 DL Reasoning Algorithms 168
 - 6.4.1 Tableau-Based Consistency Checking 169
 - 6.4.2 Automata 171
 - 6.4.3 Resolution 171
 - 6.5 Reasoning Complexity 171
 - 6.6 DL-Based Reasoners 173
 - 6.6.1 HermiT 174
 - 6.6.2 Pellet 175
 - 6.6.3 FaCT++ 176

- 6.6.4 Racer 178
- 6.7 Image Interpretation 178
 - 6.7.1 Image Interpretation as Abduction 181
 - 6.7.2 Image Interpretation Using Fuzzy DL Axioms 183
- 6.8 Video Scene Interpretation 184
 - 6.8.1 Video Event Recognition via Reasoning over
Temporal DL Axioms 184
- 6.9 Distributed and Federated Reasoning 185
- 6.10 Summary 186

- References 187**

- Index 201**

About the Author



Leslie F. Sikos, Ph.D., is a researcher at Flinders University, Australia, specializing in knowledge representation of multimedia resources, multimedia ontology engineering, and automated video scene interpretation via spatiotemporal reasoning and information fusion. He has worked in both academia and the industry, thereby acquiring hands-on skills in Semantic Web technologies, image processing, video authoring, CGI, and 3D modeling. Dr. Sikos is an experienced author with a strong textbook publishing background, who regularly works with the world's leading publishing houses. He has developed two of the most expressive multimedia ontologies to date, which are formally grounded in description logics and

complemented by rulesets. Dr. Sikos is a recognized expert in the standardization of next-generation video indexing techniques that leverage semantic annotation at different levels of granularity, including entire videos, video shots, video segments, video frames, and regions of interest of video frames, and mapping depicted concepts to Linked Data. Inspired by the creation and exploitation of rich LOD datasets, Dr. Sikos actively contributes to the development of open standards and open data repositories. For more information, visit <https://www.lesliesikos.com>.