AN ONTOLOGY GUIDED FRAMEWORK FOR MULTIMEDIA CONTENT ANNOTATION

NISHA PAHAL

BHARTI SCHOOL OF TELECOMMUNICATION
TECHNOLOGY AND MANAGEMENT
INDIAN INSTITUTE OF TECHNOLOGY DELHI
MAY 2017

AN ONTOLOGY GUIDED FRAMEWORK FOR MULTIMEDIA CONTENT ANNOTATION

by
NISHA PAHAL

Bharti School of Telecommunication
Technology and Management

Submitted
in fulfilment of the requirements of the degree of DOCTOR OF PHILOSOPHY
to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

May, 2017

Certificate

This is to certify that the thesis titled "**An Ontology-guided Framework for Multimedia Content Annotation**" being submitted by **Ms. Nisha Pahal** to the Bharti School of Telecommunication Technology and Management, Indian Institute of Technology Delhi, for the award of the Degree of **Doctor of Philosophy**, is a record of bona-fide research work carried out by her under my guidance and supervision. In our opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted to any other university or institute for the award of the degree or diploma.

Dr. S. Chaudhury
Professor
Department of
Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

Dr. Brejesh Lall
Associate Professor
Department of
Electrical Engineering
Indian Institute of Technology Delhi
New Delhi - 110 016

Acknowledgements

I would like to express my profoundest gratitude to my supervisors **Prof. Santanu Chaudhury** and **Dr. Brejesh Lall** for their unwavering support, guidance and invaluable suggestions during my research.

I am sincerely grateful to Prof. Santanu Chaudhury for enlightening me through his wisdom, knowledge and commitment to the highest standards and inspired me to develop a research competence. Sir constantly motivated me to comply to and carry his resolute research aptitude further to build my research career. His patience, cutting edge knowledge and innovative approach were key motivations throughout my Ph.D. His crucial comments on structure and organization during thesis writing have greatly enhanced value of my research work and the form in which it finally came out. I am indebted to him for his supervision and constant help without which this dissertation would not have been possible.

I would also like to express my deep gratitude to Dr. Brejesh Lall for his supervision, thought provoking conversation and enthusiastic approach. I am thankful to him for sparing many hours of his busy schedule for reviewing my thesis. His advice on both research as well as on my career have been priceless. His insightful suggestions, detailed discussions and the focussed guidance that he has provided were indispensible for the completion of this research.

I owe a debt of gratitude to the members of SRC - Prof. Subratkar, Prof. Prem kalra and Dr. Sumantra Dutta Roy for sharing

their illuminating views on a number of issues related to research. I am grateful for the guidance, comments and advices given by them during the presentation.

Furthermore I would also like to acknowledge with much appreciation the crucial role of Department of Bharti School of Telecommunication Technology and Management and Electrical Engineering for giving me the opportunity to carry out my doctoral research and to attend conferences abroad by providing the financial support.

I am thankful to the administrative staff of the Multimedia Lab and have special regard for Mrs. Usha Bhola, Mrs. Deepti Wadhwa and Mrs. Manrika Malhotra who were kind enough to advise and help in their respective roles. I am thankful to Mrs Narendra Kaur in Electrical Engineering Department for assisting me in best possible ways. I thank Sushila ji and Amit Bhaiya for their help.

I would like to extend sincere regards to my family. Words cannot express how blessed I feel on receiving endless love, support from my parents and the sacrifices they have made to make me the person I am today.

At the end, I would also like to thank all of my friends - Sandeep, Deepti Goel, Deepti Jaiswal, Reema, Anupama Mam, Vandana Mam, Kavita Mam, Hassan Sir, Vishwanath, Geetanjali, Nivedita and all my labmates who supported me and persuaded me constantly to strive towards my goal and for helping me get through the difficult times.

Thank you all for your unshakable and determined encouragement and cooperation.

Nisha Pahal

Abstract

This thesis proposes an ontology guided framework for annotating multimedia content. The annotation scheme is based upon a model of events and facilitates discovery and exploration of media content in a novel fashion. We examine the problem of multimedia content analysis to automatically detect the event related aspects from the large collection of multimedia data scattered over the web. The related conditions that enables an event to occur can be viewed as the context of an event. The context gets defined with respect to an event instance where the event context has its own structure and it involves spatial, temporal and agent/ entity related aspects. Our annotation framework aggregates contextually relevant documents from the web to assist users in different kinds of retrieval tasks. In this thesis we have presented:

1. Formalisation of a new Multimedia Web Ontology Language (E-MOWL) to handle events with media depictions. The temporal, spatial and entity aspects that are implicitly linked to an event are represented through this language to model temporal and geographical context of events. Apart from encoding the knowledge of concepts and relations in a specific domain, the language can be used to formally express the association between the various concepts and event related aspects. The formalism of the language enables reasoning with the uncertainty associated with the events and is organized in the form of Bayesian Network (BN). The observed evidences are mapped onto the BN and subsequently

the posterior probabilities are calculated to detect the presence of a higher-level event. Uncertainty reasoning allows to exploit the knowledge in the ontology for the purpose of drawing an inference.

2. The ontology-driven framework to design an unsupervised concept-driven multimedia tagging scheme that assigns contextual and expressive tags to a query video. The most important aspect of our endeavour is the automatic identification of context information from multimedia data. The context information associated with an event ties the audio-visual data with its temporal, spatial and entity related aspects. All these aspects when considered altogether provide the evidence and contribute towards recognizing an event from multimedia documents.

We have designed and implemented an ontology-guided multimedia tagging application that provides a comprehensive base for tagging multimedia content on the basis of context information. The document corpus has been formed automatically by utilizing the context information so as to assimilate semantically aligned documents from diverse web sources. This is subsequently used for identifying the conceptual evolution of event patterns temporally as well as spatially. To trace the conceptual evolution of events over time, Temporal Trend Analysis has been utilized. It makes use of statistical topic modeling and topic-to-concept mapping approaches to associate a conceptual tag to topics over time. To track the emergence of event patterns over space, Geographical Trend Analysis has been used. For this, geo-spatial knowledge has been exploited wherein, the geographical named entities are mapped onto the ontology concepts to identify the evolution of event patterns spatially. We have demonstrated the efficacy of our approach by constructing an ontology for the News Domain

and by offering various applications like tagging and trend analysis for thematic linking of documents dispersed over the web along with providing the temporal and geographical trends.

सार

इस थीसिस ने एनोटेटिंग के लिए एक आन्टोलॉजी मार्गदर्शित रूपरेखा का प्रस्ताव किया है मल्टीमीडिया सामग्री। एनोटेशन योजना पर आधारित है घटनाओं का एक मॉडल और की खोज और अन्वेषण की सुविधा एक उपन्यास फैशन में मीडिया सामग्री। हम इस समस्या की जांच करते हैं मल्टीमीडिया सामग्री विश्लेषण को स्वचालित रूप से ईवेंट का पता लगाता है मल्टीमीडिया डेटा के बड़े संग्रह से संबंधित पहलुओं को बिखरे हुए वेब पर। संबंधित स्थितियां जो एक ईवेंट को सक्षम करती हैं होने के लिए एक घटना के संदर्भ के रूप में देखा जा सकता है। प्रसंग एक घटना के उदाहरण के संबंध में, जहां कार्यक्रम आयोजित किया जाता है संदर्भ की अपनी संरचना है और इसमें स्थानिक, लौकिक शामिल है और एजेंट / संस्था संबंधित पहलुओं। हमारे एनोटेशन फ्रेमवर्क समुच्चय वेब से प्रासंगिक प्रासंगिक दस्तावेज़ों की मदद के लिए प्रयोक्ता के लिए पुनर्प्राप्ति कार्य के प्रकार। इस थीसिस में हमारे पास है प्रस्तुत किया:

- एक नई मल्टीमीडिया वेब ओटोलोजी भाषा का औपचारिकरण (ई-MOWL) मीडिया के चित्रणों के साथ घटनाओं को संभालने के लिए। अस्थायी, स्थानिक और संस्था के पहलुओं जो सम्पूर्ण रूप से जुड़े हुए हैं एक घटना के लिए मॉडल के लिए इस भाषा के माध्यम से प्रतिनिधित्व कर रहे हैं घटनाओं का अस्थायी और भौगोलिक संदर्भ। इसके अलावा एक विशेष डोमेन में अवधारणाओं और संबंधों के ज्ञान को एन्कोडिंग, भाषा को औपचारिक रूप से व्यक्त करने के लिए इस्तेमाल किया जा सकता है संबंधित विभिन्न अवधारणाओं और घटना के बीच संबंध पहलुओं। भाषा की औपचारिकता तर्क को सक्षम करती है घटनाओं से जुड़े अनिश्चितता के साथ और बायेसियन नेटवर्क (बी एन) के रूप में आयोजित किया जाता है। मनाया गया साक्ष्य बीएन पर मैप किए जाते हैं और बाद में पिछली संभावनाओं की उपस्थिति का पता लगाने के लिए गणना की जाती है एक उच्च स्तरीय घटना का। अनिश्चितता तर्क की अनुमति देता है के उद्देश्य के लिए ओण्टोलॉजी में ज्ञान का फायदा उठाने के लिए एक निष्कर्ष ड्राइंग।
- एक अनसॉर्ज्ड डिजाइन करने के लिए आन्टोलॉजी आधारित ढांचे अवधारणा-आधारित मल्टीमीडिया टैगिंग योजना जो प्रासंगिक प्रदान करती है और एक प्रश्न वीडियो में अभिव्यंजक टैग। सबसे अधिक हमारे प्रयास का महत्वपूर्ण पहलू है स्वचालित पहचान-मल्टीमीडिया डेटा से संदर्भ जानकारी का निर्माण। एक घटना से संबंधित संदर्भ जानकारी ऑडियो-विज़न से जुड़ी है अपने अस्थायी, स्थानिक और संस्था संबंधित पहलुओं के साथ डेटा। इन सभी पहलुओं को पूरी तरह से प्रदान करने पर विचार किया गया सबूत और एक घटना को पहचानने के लिए योगदान मल्टीमीडिया दस्तावेज़ों से।

हमने एक ओटोलॉजी-निर्देशित मल्टीमीडिया को डिज़ाइन और कार्यान्वित किया है टैगिंग अनुप्रयोग जो इसके लिए एक व्यापक आधार प्रदान करता है संदर्भ जानकारी के आधार पर मल्टीमीडिया सामग्री टैगिंग। दस्तावेज़ कॉरपस का इस्तेमाल करके स्वचालित रूप से गठन किया गया है संदर्भ संबंधी जानकारी ताकि अर्थात् रूप से गठबंधन को आत्मसात किया जा सके विभिन्न वेब स्रोतों से दस्तावेज़। यह बाद में उपयोग किया जाता है घटना पैटर्न के वैचारिक विकास को अस्थायी रूप से पहचानने के लिए साथ ही स्थानिक रूप से। वैचारिक विकास का पता लगाने के लिए समय के साथ की घटनाएं, टेम्पोरल ट्रेंड विश्लेषण का उपयोग किया गया है। यह सांख्यिकीय विषय मॉडलिंग और विषय-टू-अवधारणा मानचित्रण का उपयोग करता है समय के साथ विषयों पर एक अवधारणात्मक टैग को जोड़ने के लिए दृष्टिकोण। अंतरिक्ष के ऊपर घटना के पैटर्न के उद्भव को ट्रैक करने के लिए, भौगोलिक ट्रेंड विश्लेषण का इस्तेमाल किया गया है। इसके लिए, भू-स्थानिक ज्ञान जिसका फायदा उठाया गया है, भौगोलिक नामित संस्थाएं हैं के विकास की पहचान करने के लिए टेटोलॉजी अवधारणाओं पर मैप किया गया घटना पैटर्न स्थानिक। हम की क्षमता के

प्रदर्शन का प्रदर्शन किया है समाचार डोमेन के लिए एक ओटोलोजी का निर्माण करके हमारा दृष्टिकोण और टैगिंग और प्रवृत्ति विश्लेषण जैसे विभिन्न अनुप्रयोगों के द्वारा वेब पर छितरी गई दस्तावेजों के विषयगत लिंक के लिए अस्थायी और भौगोलिक रुझान प्रदान करने के साथ।

Contents

List of Figures

List of Tables

Glossary

OWL	Web Ontology Language
MOWL	Multimedia Web Ontology Language
VERL	Video Event Representation Language
SVM	Support Vector Machine
CPT	Conditional Probability Table
XML	Extensible Markup Language
HTML	Hyper Text Markup Language
BN	Bayesian Network
OM	Observation Model
E-MOWL	Multimedia Event Ontology Language
LDA	Latent Dirichlet Allocation
SIFT	Scale-Invariant Feature Transform
Corr-LDA	Correspondence LDA
NLP	Natural Language Processing
NER	Named Entity Recognition
hLDA	Hierachical Latent Dirichlet Allocation
CRP	Chinese Restaurant Process
GIS	Geographic Information System