

# Theory and Applications of Natural Language Processing

Series Editors:

Graeme Hirst (Textbooks)

Eduard Hovy (Edited volumes)

Mark Johnson (Monographs)

## Aims and Scope

The field of Natural Language Processing (NLP) has expanded explosively over the past decade: growing bodies of available data, novel fields of applications, emerging areas and new connections to neighboring fields have all led to increasing output and to diversification of research.

“Theory and Applications of Natural Language Processing” is a series of volumes dedicated to selected topics in NLP and Language Technology. It focuses on the most recent advances in all areas of the computational modeling and processing of speech and text across languages and domains. Due to the rapid pace of development, the diversity of approaches and application scenarios are scattered in an ever-growing mass of conference proceedings, making entry into the field difficult for both students and potential users. Volumes in the series facilitate this first step and can be used as a teaching aid, advanced-level information resource or a point of reference.

The series encourages the submission of research monographs, contributed volumes and surveys, lecture notes and textbooks covering research frontiers on all relevant topics, offering a platform for the rapid publication of cutting-edge research as well as for comprehensive monographs that cover the full range of research on specific problem areas.

The topics include applications of NLP techniques to gain insights into the use and functioning of language, as well as the use of language technology in applications that enable communication, knowledge management and discovery such as natural language generation, information retrieval, question-answering, machine translation, localization and related fields.

The books are available in printed and electronic (e-book) form:

- \* Downloadable on your PC, e-reader or iPad
- \* Enhanced by Electronic Supplementary Material, such as algorithms, demonstrations, software, images and videos
- \* Available online within an extensive network of academic and corporate R&D libraries worldwide
- \* Never out of print thanks to innovative print-on-demand services
- \* Competitively priced print editions for eBook customers thanks to MyCopy service  
<http://www.springer.com/librarians/e-content/mycopy>

For other titles published in this series, go to  
[www.springer.com/series/8899](http://www.springer.com/series/8899)

Alessandro Oltramari • Piek Vossen  
Lu Qin • Eduard Hovy  
Editors

# New Trends of Research in Ontologies and Lexical Resources

Ideas, Projects, Systems



Springer

*Editors*

Alessandro Oltramari  
Psychology Department  
Carnegie Mellon University  
Pittsburgh, PA  
USA

Lu Qin  
Department of Computing  
Hong Kong Polytechnic University  
Hong Kong

Piek Vossen  
Faculty of Arts, Language Cognition  
and Communication (LCC)  
Vrije University Amsterdam  
Amsterdam  
Netherlands

Eduard Hovy  
Language Technologies Institute  
Carnegie Mellon University  
Pittsburgh, PA  
USA

ISSN 2192-032X

ISBN 978-3-642-31781-1

DOI 10.1007/978-3-642-31782-8

Springer Heidelberg New York Dordrecht London

ISSN 2192-0338 (electronic)

ISBN 978-3-642-31782-8 (eBook)

Library of Congress Control Number: 2012954237

© Springer-Verlag Berlin Heidelberg 2013

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

*In Memory of Emanuele Pianta*



# Contents

<b>1</b>	<b>Introduction</b> .....	<b>1</b>
	Alessandro Oltramari, Piek Vossen, Lu Qin, and Eduard Hovy	
<b>Part I Achieving the Interoperability of Linguistic Resources in the Semantic Web</b>		
<b>2</b>	<b>Towards Open Data for Linguistics: Linguistic Linked Data</b> .....	<b>7</b>
	Christian Chiarcos, John McCrae, Philipp Cimiano, and Christiane Fellbaum	
2.1	Motivation and Overview .....	8
2.2	Modelling Linguistic Resources as Linked Data .....	10
2.2.1	Modelling Lexical-Semantic Resources: WordNet .....	12
2.2.2	Modelling Annotated Corpora: MASC .....	14
2.3	Benefits of Linked Data for Linguistics .....	16
2.3.1	Structural Interoperability .....	17
2.3.2	Linking and Federation .....	18
2.3.3	Conceptual Interoperability .....	19
2.3.4	Ecosystem .....	20
2.3.5	Dynamic Import .....	20
2.4	Community Efforts Towards Lexical Linked Data .....	21
2.4.1	The Open Linguistics Working Group .....	21
2.4.2	W3C Ontology-Lexica Community Group .....	22
2.5	Summary .....	23
	References .....	24
<b>3</b>	<b>Establishing Interoperability Between Linguistic and Terminological Ontologies</b> .....	<b>27</b>
	Wim Peters	
3.1	Introduction .....	27
3.2	Linguistic Knowledge .....	29
3.3	Networking Linguistic Ontologies .....	31

3.4	Related Work .....	33
3.5	LingNet .....	34
3.5.1	The LingNet Model.....	34
3.5.2	LingNet Implementation .....	36
3.6	Discussion .....	38
3.7	Conclusion and Future Work .....	40
	References.....	41
<b>4</b>	<b>On the Role of Senses in the Ontology-Lexicon .....</b>	<b>43</b>
	Philipp Cimiano, John McCrae, Paul Buitelaar, and Elena Montiel-Ponsoda	
4.1	Introduction.....	43
4.2	Senses: Universal or Context-Specific? .....	46
4.3	Senses in the Ontology-Lexicon Interface .....	48
4.3.1	Senses as Reification .....	49
4.3.2	Sense as Subset of Uses .....	50
4.3.3	Sense as a Subconcept.....	50
4.3.4	The Three Facets .....	52
4.4	Systematic Polysemy in the Ontology-Lexicon Interface .....	52
4.5	Senses in the Ontology-Lexicon Model Lemon .....	55
4.5.1	Sense Properties .....	56
4.5.2	Contexts and Conditions .....	57
4.5.3	Sense Relations .....	59
4.6	Conclusions.....	60
	References.....	61

## Part II Event Analysis from Text and Multimedia

<b>5</b>	<b>KYOTO: A Knowledge-Rich Approach to the Interoperable Mining of Events from Text .....</b>	<b>65</b>
	Piek Vossen, Eneko Agirre, German Rigau, and Aitor Soroa	
5.1	Introduction.....	65
5.2	Packaging of Events .....	66
5.3	KYOTO Overview .....	69
5.4	Ontological and Lexical Background Knowledge .....	72
5.4.1	Ontology .....	73
5.4.2	Wordnet to Ontology Mappings .....	74
5.5	Off-Line Reasoning and Ontological Tagging .....	76
5.6	Event Extraction .....	77
5.7	Experimental Results.....	80
5.7.1	In-Depth Evaluation .....	80
5.7.2	Large Scale Evaluation .....	83
5.7.3	Transferring to Another Language.....	87
5.8	Conclusion .....	88
	References.....	89



<b>6</b>	<b>Anchoring Background Knowledge to Rich Multimedia</b>	
	<b>Contexts in the KNOWLEDGESTORE</b> .....	91
	R. Cattoni, F. Corcoglioniti, C. Girardi, B. Magnini, L. Serafini, and R. Zanolì	
6.1	Introduction .....	92
6.2	State of the Art .....	94
6.3	The KNOWLEDGESTORE Approach .....	96
6.3.1	Representation Layers .....	96
6.3.2	Content Processing .....	99
6.4	System Implementation .....	100
6.4.1	KNOWLEDGESTORE Core .....	100
6.4.2	Resource Preprocessing .....	101
6.4.3	Mention Extraction .....	102
6.4.4	Coreference Resolution .....	102
6.4.5	Mention–Entity Linking .....	104
6.4.6	Entity Creation and Enrichment .....	105
6.5	Experiments and Results .....	105
6.5.1	KNOWLEDGESTORE Population .....	106
6.5.2	Entity-Based Search .....	107
6.5.3	Contextualized Semantic Enrichment .....	108
6.6	Conclusions and Future Work .....	110
	References .....	111
<b>7</b>	<b>Lexical Mediation for Ontology-Based Annotation of Multimedia</b> ...	113
	Mario Cataldi, Rossana Damiano, Vincenzo Lombardo, and Antonio Pizzo	
7.1	Introduction .....	113
7.2	Related Work .....	115
7.3	Case Study: Annotating Stories in Video .....	117
7.4	Accessing Large Scale Commonsense Knowledge Through a Lexical Interface .....	121
7.4.1	The Architecture of CADMOS .....	121
7.4.2	The Meaning Negotiation Process .....	123
7.5	Annotation Test and Discussion .....	127
7.5.1	Experimental Setting .....	127
7.5.2	Results and Discussion .....	129
7.6	Conclusion .....	131
	References .....	132
<b>8</b>	<b>Knowledge in Action: Integrating Cognitive Architectures and Ontologies</b> .....	135
	Alessandro Oltramari and Christian Lebiere	
8.1	Introduction .....	135
8.2	Knowledge Mechanisms Meet Contents in Visual Intelligence ....	137
8.2.1	Mechanisms: Cognitive Architectures as Modules of Knowledge Production .....	137

8.2.2	Contents: Ontologies as Declarative Knowledge Resources .....	138
8.2.3	Human Visual Intelligence .....	139
8.3	<i>Making Sense of Visual Data</i> .....	141
8.3.1	HOMinE: Model and Implementation .....	142
8.3.2	The Cognitive Engine .....	146
8.3.3	Recognition Task .....	147
8.3.4	Description Task .....	149
8.4	Evaluation .....	150
8.5	Conclusions and Future Work .....	152
	References .....	152

### Part III Enhancing NLP with Ontologies

<b>9</b>	<b>Use of Ontology, Lexicon and Fact Repository for Reference Resolution in Ontological Semantics</b> .....	157
	Marjorie McShane and Sergei Nirenburg	
9.1	Introduction .....	157
9.2	Our View of Reference Resolution Versus Others .....	159
9.3	The OntoAgent Environment and Its Resources .....	161
9.3.1	Comparing OntoAgent Static Knowledge Resources with Others .....	164
9.3.2	The OntoSem Text Analyzer .....	165
9.4	The Reference Resolution Algorithm .....	166
9.4.1	Stage 1: Proper Name Analysis During Preprocessing ...	166
9.4.2	Stage 2: Detection of Potentially Missing Elements in the Syntactic Parse .....	167
9.4.3	Stage 3: Reference Processing During Basic Semantic Analysis .....	168
9.4.4	Stage 4: Running Lexically Recorded Meaning Procedures .....	172
9.4.5	Stage 5: Dedicated Reference Resolution Module .....	172
9.5	Final Thoughts: Semantics in Reference Resolution .....	181
	References .....	183
<b>10</b>	<b>Ontology-Based Semantic Interpretation via Grammar Constraints</b> .....	187
	Smaranda Muresan	
10.1	Introduction .....	187
10.2	Lexicalized Well-Founded Grammar .....	188
10.2.1	Semantic Molecule: A Syntactic-Semantic Representation .....	189
10.2.2	Semantic Composition and Interpretation as Grammar Constraints .....	191
10.2.3	LWFG Learning Model .....	192

10.3	Ontology-Based Semantic Interpretation .....	194
10.3.1	Levels of Representation .....	194
10.3.2	The Local Ontology-Based Semantic Interpreter .....	196
10.3.3	Global Semantic Interpreter .....	198
10.4	Knowledge Acquisition and Querying Experiments .....	199
10.4.1	Acquisition of Terminological Knowledge from Consumer Health Definitions .....	200
10.4.2	Natural Language Querying .....	202
10.5	Ambiguity Handling .....	203
10.6	Conclusions .....	205
	References .....	205
<b>11</b>	<b>How Ontology Based Information Retrieval Systems May Benefit from Lexical Text Analysis .....</b>	<b>209</b>
	Sylvie Ranwez, Benjamin Duthil, Mohameth François Sy, Jacky Montmain, Patrick Augereau, and Vincent Ranwez	
11.1	Introduction .....	210
11.2	Related Work .....	211
11.2.1	Conceptual Versus Keyword-Based IRSs .....	212
11.2.2	Hybrid Ontology Based Information Retrieval System ...	213
11.2.3	Concept Identification Through Lexical Analysis .....	218
11.3	Concept Identification Through Lexical Analysis: The “Synopsis” Approach .....	219
11.3.1	Concept Characterization .....	220
11.3.2	Thematic Extraction .....	222
11.4	Human Accessibility Enhanced at the Crossroads of Ontology and Lexicology .....	223
11.4.1	An Example of Concept-Based IRS: OBIRS .....	223
11.4.2	Ontology and Lexical Resource Interfacing Within Hybrid IRSs .....	225
11.5	Evaluation: User Feedback on a Real Case Study .....	226
11.6	Conclusion and Perspectives .....	227
	References .....	228

## **Part IV Sentiment Analysis Thorough Lexicon and Ontologies**

<b>12</b>	<b>Detecting Implicit Emotion Expressions from Text Using Ontological Resources and Lexical Learning .....</b>	<b>235</b>
	Alexandra Balahur, Jesús M. Hermida, and Hristo Tanev	
12.1	Introduction .....	235
12.2	Related Work .....	237
12.2.1	Appraisal Theories .....	237
12.2.2	Affect Detection and Classification in Natural Language Processing .....	237
12.2.3	Knowledge Bases for NLP Applications .....	238

12.2.4	Lexical Learning.....	238
12.2.5	Linking Ontologies with Lexical Resources .....	239
12.3	The EmotiNet Knowledge Base .....	239
12.3.1	Self-Reported Affect and the ISEAR Data Set .....	240
12.3.2	Building the EmotiNet Knowledge Base .....	240
12.3.3	Preliminary Extensions of EmotiNet .....	242
12.4	Further Extensions of EmotiNet with Lexical and Ontological Resources.....	244
12.4.1	Extending EmotiNet with Additional Emotion-Triggering Situations.....	244
12.4.2	Extending EmotiNet Using Ontopopulis .....	245
12.5	Evaluation .....	248
12.6	Discussion, Conclusions and Future Work.....	251
	References.....	253
<b>13</b>	<b>The Agile Cliché: Using Flexible Stereotypes as Building Blocks in the Construction of an Affective Lexicon.....</b>	<b>257</b>
	Tony Veale	
13.1	Introduction.....	257
13.2	Related Work and Ideas .....	259
13.3	Finding Stereotypes on the Web .....	261
13.3.1	Web-derived Models of Typical Behavior .....	263
13.3.2	Mutual Reinforcement Among Properties.....	265
13.4	Estimating Lexical Affect.....	266
13.5	In the Mood for Affective Search.....	269
13.6	Empirical Evaluation .....	270
13.6.1	Bottom Level: Properties and Behaviors of Stereotypes .....	270
13.6.2	Top Level: Stereotypical Concepts .....	271
13.6.3	Separating Words by Affect: Two Views .....	272
13.7	Conclusions .....	273
	References.....	274
	<b>Index.....</b>	<b>277</b>

# List of Figures

Fig. 2.1	The core of the <i>lemon</i> model .....	14
Fig. 2.2	Representing and integrating annotations for syntax and frame-semantics in a directed graph .....	15
Fig. 3.1	The LMM semiotic triangle .....	29
Fig. 3.2	The LingNet metamodel .....	35
Fig. 3.3	The LingNet model extension .....	36
Fig. 3.4	The LingNet architecture .....	38
Fig. 3.5	Refinement module for OverlapRelation .....	39
Fig. 3.6	LingNet’s linguistic and terminological domain coverage .....	40
Fig. 4.1	Example of mapping of natural language into a task vocabulary ..	44
Fig. 4.2	Concept of ‘school’ in the ontology .....	54
Fig. 4.3	The modelling of senses within the <i>lemon</i> ontology-lexicon model .....	56
Fig. 5.1	Overview of the KYOTO architecture .....	70
Fig. 5.2	Terms in KAF (in <i>blue</i> ) expanded with ontological tags. Ontological classes from direct mappings are marked with ‘*’ and implied ontological classes are marked with ‘**’ and in <i>red</i> .....	71
Fig. 5.3	Example of a Kybot profile .....	78
Fig. 5.4	Output structure resulting from a Kybot profile .....	78
Fig. 5.5	Events and participants extracted from the terms <i>migratory fish</i> and <i>spawn</i> .....	79
Fig. 5.6	Search results in table form for the query <i>infection of frogs</i> .....	85
Fig. 5.7	Search results on Google map for filtered results of the query <i>infection of frogs</i> .....	86
Fig. 6.1	Relating resources, mentions, entities and contexts in the KNOWLEDGESTORE .....	96
Fig. 6.2	The KNOWLEDGESTORE representation layers (only the most relevant attributes are shown).....	97

Fig. 6.3	The KNOWLEDGESTORE approach for processing information content .....	99
Fig. 6.4	Context-driven mention–entity linking algorithm .....	104
Fig. 6.5	Example of entity-based search for query “Schumacher” in the <i>TrentinoMedia</i> application .....	108
Fig. 6.6	Example of enrichments of mentions “Michael Schumacher” and “Ungheria” (Hungary) in a news article with context <i>{World, 02 Aug 2010, Formula 1}</i> , from the <i>TrentinoMedia</i> application .....	109
Fig. 7.1	The annotation process framework. The annotation system incorporates the semantic model (Application Profile, below) and the external vocabularies, thus enforcing the correctness of the metadata encoded by hand by the human annotators and their translation into a formal language .....	118
Fig. 7.2	The template for annotating story incidents within the Cadmos system; the incident is described by an ontological concept, a semantic frame and its participants.....	119
Fig. 7.3	The architecture of the Cadmos system .....	122
Fig. 7.4	The disambiguation model proposed in this chapter includes several knowledge bases: MultiWordNet, WordNet, FrameNet and YAGOSUMO .....	124
Fig. 7.5	The annotation of an example incident from North By Northwest (a policeman questions Eve about Roger) .....	126
Fig. 8.1	Information from the environment is processed through the different modules of ACT-R.....	138
Fig. 8.2	An excerpt of DOLCE top level .....	143
Fig. 8.3	HOMinE backbone taxonomy .....	145
Fig. 8.4	CMU Mind’s eye architecture .....	146
Fig. 8.5	A Diagram of the Cognitive Engine (colored boxes in the <i>bottom</i> represents the preprocessing algorithms and IAR system) .....	147
Fig. 8.6	EAR processing schema .....	148
Fig. 8.7	Equation for Bayesian activation pattern matching.....	149
Fig. 8.8	Description performance.....	151
Fig. 9.1	The OntoSem text analyzer .....	158
Fig. 10.1	(1) Elementary semantic molecules for the adjective <i>formal</i> ( <b>a</b> ) and the noun <i>proposal</i> ( <b>b</b> ); (2) A derived semantic molecule for the noun phrase <i>formal proposal</i> .....	189
Fig. 10.2	Levels of representation for the utterance <i>a virus that does not persist in the blood serum</i> .....	195
Fig. 10.3	Terminological knowledge acquired from consumer health definitions .....	201

Fig. 10.4	Examples of precise and vague questions, their OKR representations and the concept-level answers .....	202
Fig. 10.5	Two OKRs for <i>I saw the man with the telescope</i> .....	204
Fig. 11.1	Overview of our CoLexIR approach .....	211
Fig. 11.2	Semantic continuum: the classification of hybrid IRSs as proposed by Giunchiglia et al. [21] .....	215
Fig. 11.3	Example of sliding window $\mathcal{F}'$ (with size = 1). The dots between two nouns symbolize the possible presence of any word that is not a common noun .....	222
Fig. 11.4	Example of sensitivity analysis. Five discontinuities are observed, they correspond to rough changes of semantic level in the description of a document .....	223
Fig. 11.5	<i>CoLexIR</i> interface displays selected document histograms in a semantic map according to their relevance scores w.r.t the query (symbolized by the <i>question mark</i> ) (B). The query concepts and their weights are provided (C) as well as query parameters and color code legend (A). Match explanation of a document is proposed as well as a link towards the whole document (E). Document passages related to the query concepts are available in a pop-up (D) .....	226
Fig. 12.1	EmotiNet ontology cores .....	241
Fig. 12.2	Main concepts and examples of instances in the EmotiNet KB .....	241
Fig. 12.3	Example of action chain extracted from the ISEAR corpus and added to the EmotiNet KB .....	242
Fig. 12.4	Precision action similarity .....	248
Fig. 12.5	Recall action similarity .....	249
Fig. 12.6	Precision emotion similarity .....	250
Fig. 12.7	Recall emotion similarity .....	251

