Computational Analysis of Storylines

Event structures are central in linguistics and artificial intelligence research: People can easily refer to changes in the world, identify their participants, distinguish relevant information, and have expectations of what can happen next. Part of this process is based on mechanisms similar to narratives, which are at the heart of information sharing. But it remains difficult to automatically detect events or automatically construct stories from such event representations. This book explores how to handle today’s massive news streams and provides multidimensional, multimodal, and distributed approaches, like automated deep learning, to capture events and narrative structures involved in a “story.” This overview of the current state-of-the-art on event extraction, temporal and casual relations, and storyline extraction aims to establish a new multidisciplinary research community with a common terminology and research agenda. Graduate students and researchers in natural language processing, computational linguistics, and media studies will benefit from this book.

Tommaso Caselli is an Assistant Professor in Computational Semantics at the University of Groningen. He received his PhD in computational linguistics on temporal processing of texts from the University of Pisa. His main research areas are in discourse processing, event extraction, and (event) sentiment analysis. He is one of the founders of the “Event and Stories in the News” workshop series and is currently working on developing computational models and natural language processing tools to extract plot structures from news. He took part in organizing semantic evaluation campaigns in natural language processing for English and Italian.

Eduard Hovy is a Research Professor at the Language Technology Institute at Carnegie Mellon University. He was awarded honorary doctorates from the National Distance Education University (UNED) in Madrid in 2013 and the University of Antwerp in 2015. He is one of the initial 17 Fellows of the Association for Computational Linguistics (ACL). His research contributions include the co-development of the ROUGE text summarization evaluation method, the BLANC coreference evaluation method, the Omega ontology, the Webclopedia QA Typology, the FEMTI machine translation evaluation classification, the DAP text harvesting method, the OntoNotes corpus, and a model of structured distributional semantics.

Martha Palmer is a Professor at the University of Colorado in Linguistics, Computer Science, and Cognitive Science. She is a AAAI Fellow and an ACL Fellow. She works on trying to capture elements of the meanings of words that can comprise automatic representations of complex sentences and documents. She is a co-editor of Linguistic Issues in Language Technology and has been on the CLJ Editorial Board and a co-editor of JNLE. She is a past President of the Association for Computational Linguistics, past Chair of SIGLEX and SIGHAN, and was the Director of the 2011 Linguistics Institute held in Boulder, Colorado.
PIEK VOSSEN is Professor at Vrije Universiteit Amsterdam. He is the co-founder and co-president of the Global Wordnet Association, organizing the international Wordnet conferences since 2002. In 2013, he received the Dutch Spinoza prize for his research. He used this prize to launch a series of projects that included the structuring of news streams using storylines and reader/writer perspectives. Vossen’s current main research focuses on cross-document event co-reference and perspective modeling of multiple sources with respect to event data and modeling event implications, as well as event timelines and storylines.
Studies in Natural Language Processing

Series Editor:
Chu-Ren Huang, The Hong Kong Polytechnic University

Associate Series Editor:
Qi Su, Peking University

Editorial Board Members:
Nianwen Xue, Brandeis University
Maarten de Rijke, University of Amsterdam
Lori Levin, Carnegie Mellon University
Alessandro Lenci, Universita degli Studi, Pisa
Francis Bond, Nanyang Technological University

This series offers widely accessible accounts of the state of the art in natural language processing. Established on the foundations of formal language theory and statistical learning, natural language processing is burgeoning with the widespread use of large annotated corpora, rich models of linguistic structure, and rigorous evaluation methods. New multilingual and multimodal language technologies have been stimulated by the growth of the web and pervasive computing devices. The series strikes a balance between statistical versus symbolic methods; deep versus shallow processing; rationalism versus empiricism; and fundamental science versus engineering. Each volume sheds light on these pervasive themes, delving into theoretical foundations and current applications. The series is aimed at a broad audience who are directly or indirectly involved in natural language processing, from fields including corpus linguistics, psycholinguistics, information retrieval, machine learning, spoken language, human-computer interaction, robotics, language learning, ontologies, and databases.

Also in the series

Douglas E. Appelt, Planning English Sentences
Madeleine Bates and Ralph M. Weischedel (eds.), Challenges in Natural Language Processing
Steven Bird, Computational Phonology
Peter Bosch and Rob van der Sandt, Focus
The Language of Word Meaning: (editors) Federica Busa and Pierrette Bouillon
Inheritance, Defaults and the Lexicon: (editors) Ted Briscoe, Valeria de Paiva, and Ann Copestake
Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Varile,
   Annie Zaenen, Antonio Zampolli, and Victor Zue (eds.), Survey of the State
   of the Art in Human Language Technology
David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky (eds.),
   Natural Language Parsing
Ralph Grishman, Computational Linguistics
Graeme Hirst, Semantic Interpretation and the Resolution of Ambiguity
András Kornai, Extended Finite State Models of Language
Kathleen R. McKeown, Text Generation
Martha Stone Palmer, Semantic Processing for Finite Domains
Terry Patten, Systemic Text Generation as Problem Solving
Ehud Reiter and Robert Dale, Building Natural Language Generation Systems
Manny Rayner, David Carter, Pierette Bouillon, Vassilis Digalakis, and
   Matis Wiren (eds.), The Spoken Language Translator
Michael Rosner and Roderick Johnson (eds.), Computational
   Lexical Semantics
Richard Sproat, A Computational Theory of Writing Systems
George Anton Kiraz, Computational Nonlinear Morphology
Nicholas Asher and Alex Lascarides, Logics of Conversation
Margaret Masterman (edited by Yorick Wilks), Language, Cohesion and Form
Walter Daelemans and Antal van den Bosch, Memory-Based
   Language Processing
Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci,
   Alessandro Oltramari, and Laurent Prévot (eds.), Ontology and the Lexicon:
   A Natural Language Processing Perspective
Thierry Poibeau and Aline Villavicencio (eds.), Language, Cognition, and
   Computational Models
Bind Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions,
   Second Edition
Marcos Zampieri and Preslav Nakov, Similar Languages, Varieties, and
   Dialects: A Computational Perspective
Computational Analysis of Storylines
Making Sense of Events

Edited by

TOMMASO CASELLI
University of Groningen

EDUARD HOVY
Carnegie Mellon University, Pennsylvania

MARTHA PALMER
University of Colorado at Boulder

PIEK VOSSEN
Vrije Universiteit, Amsterdam
## Contents

<table>
<thead>
<tr>
<th>List of Contributors</th>
<th>page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction and Overview</td>
<td>1</td>
</tr>
</tbody>
</table>

### PART ONE FOUNDATIONAL COMPONENTS OF STORYLINES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The Role of Event-Based Representations and Reasoning in Language</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>James Pustejovsky</td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Introduction</td>
<td>23</td>
</tr>
<tr>
<td>1.2</td>
<td>Introducing Situations and Events</td>
<td>25</td>
</tr>
<tr>
<td>1.3</td>
<td>Modeling the Substructure of Events</td>
<td>33</td>
</tr>
<tr>
<td>1.4</td>
<td>Enriching VerbNet with Event Dynamics</td>
<td>38</td>
</tr>
<tr>
<td>1.5</td>
<td>Conclusions</td>
<td>42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>The Rich Event Ontology: Ontological Hub for Event Representations</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Claire Bonial, Susan W. Brown, Martha Palmer, and Ghazaleh Kazeminejad</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Introduction</td>
<td>47</td>
</tr>
<tr>
<td>2.2</td>
<td>Ontologies</td>
<td>48</td>
</tr>
<tr>
<td>2.3</td>
<td>Semantic Role Labeling</td>
<td>56</td>
</tr>
<tr>
<td>2.4</td>
<td>Conclusions, Gaps, and Future Work</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Decomposing Events and Storylines</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>William Croft, Pavlína Kalin, and Michael Regan</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction: Events within Stories and Events within Events</td>
<td>67</td>
</tr>
<tr>
<td>3.2</td>
<td>Constructions as Well as Verbs Determine the Internal Structure of Events</td>
<td>69</td>
</tr>
</tbody>
</table>

© in this web service Cambridge University Press
3.3 Time and Qualitative State (Change) 71
3.4 Causation 74
3.5 Annotation Scheme 78
3.6 Visualization 80
3.7 Conclusion 82

4 Extracting and Aligning Timelines 87
  Mark A. Finlayson, Andres Cremisini, and Mustafa Ocal
  4.1 Introduction 87
  4.2 Extracting Timelines 89
  4.3 Aligning Timelines 95
  4.4 Bringing It All Together 100
  4.5 Conclusion 101

5 Event Causality 106
  Paramita Mirza
  5.1 Introduction 106
  5.2 Modelling Causal Relations 108
  5.3 Causal Annotation in Natural Language Text 110
  5.4 Extracting Event Causality 114
  5.5 Causal Commonsense Discovery 117
  5.6 Conclusions 120

6 A Narratology-Based Framework for Storyline Extraction 125
  Piek Vossen, Tommaso Caselli, and Roxane Segers
  6.1 Introduction 125
  6.2 A Narratology-Grounded Framework for Storylines Identification 126
  6.3 From Theory to Data: Annotating Causelines and Storylines 130
  6.4 Validating Causelines and Extracting Storylines 134
  6.5 Conclusion 137

PART TWO CONNECTING THE DOTS: RESOURCES, TOOLS, AND REPRESENTATIONS 143

7 The Richer Event Description Corpus for Event–Event Relations 145
  Tim O’Gorman, Kristin Wright-Bettner, and Martha Palmer
  7.1 Introduction 145
  7.2 A Comparison of Event Annotation Choices 146
  7.3 Long-Distance Relations in RED: Contains, Causality, and Coreference 155
Contents

7.4 Studying RED Impact on Event Ordering 156
7.5 Conclusions 158

8 Low-Resource Event Extraction via Share-and-Transfer and Remaining Challenges 163
Heng Ji and Clare Voss
8.1 Introduction 163
8.2 Approach Overview 166
8.3 Share: Construction of Common Semantic Space 167
8.4 Transfer: From High- to Low-Resource Setting 175
8.5 Transfer Learning Performance 177
8.6 Remaining Challenges 177
8.7 Conclusions and Future Research Directions 182

9 Reading Certainty across Sources 187
Benjamin Miller
9.1 Introduction 187
9.2 Background 192
9.3 Methods 194
9.4 Results 195
9.5 Discussion 198
9.6 Conclusion 200

10 Narrative Homogeneity and Heterogeneity in Document Categories 203
Dan Simonson and Anthony R. Davis
10.1 Introduction: Narrative Schemas and Their Evaluations 203
10.2 Background 205
10.3 Data and Schema Generation 206
10.4 Evidence through NASTEA Task 208
10.5 Evidence through Schema Stability 213
10.6 Discussion 216
10.7 Conclusions 217

11 Exploring Machine Learning Techniques for Linking Event Templates 221
Jakub Piskorski, Fredi Šarić, Vanni Zavarella, and Martin Atkinson
11.1 Introduction 221
11.2 Task Description 224
11.3 Event Similarity Metrics 225
11.4 Experiments 230
11.5 Conclusions 236
12 Semantic Storytelling: From Experiments and Prototypes to a Technical Solution

Georg Rehm, Karolina Zaczynska, Peter Bourgonje, Malte Ostendorff, Julián Moreno-Schneider, Maria Berger, Jens Rauenbusch, André Schmidt, Mikka Wild, Joachim Böttger, Joachim Quantz, Jan Thomsen, and Rolf Fricke

12.1 Introduction: Technologies for Content Curation 240
12.2 Semantic Storytelling: Selected Components 242
12.3 Semantic Storytelling in Industry Use Cases 246
12.4 Towards a Flexible and Robust Technology Solution for Semantic Storytelling 250
12.5 Related Work 252
12.6 Conclusions and Future Work 254

Author Index 260
Contributors

André Schmidt
3pc Neue Kommunikation GmbH, Germany

Andres Cremisini
Florida International University, USA

Anthony R. Davis
Southern Oregon University, USA

Benjamin Miller
Emory University, USA

Claire Bonial
Army Research Lab, USA

Clare Voss
Army Research Lab, USA

Dan Simonson
BlackBoiler LLC, USA

Fredi Šarić
University of Zagreb, Croatia

Georg Rehm
DFKI GmbH, Germany
xii List of Contributors

**Ghazaleh Kazeminejad**
University of Colorado Boulder, USA

**Heng Ji**
University of Illinois at Urbana–Champaign, USA

**Jakub Piskorski**
Joint Research Centre of the European Commission, Italy

**James Pustejovsky**
Brandeis University, USA

**Jan Thomsen**
Condat AG, Germany

**Jens Rauenbusch**
3pc Neue Kommunikation GmbH, Germany

**Joachim Böttger**
ART+COM AG, Germany

**Joachim Quantz**
ART+COM AG, Germany

**Julián Moreno-Schneider**
DFKI GmbH, Germany

**Karolina Zaczyńska**
DFKI GmbH, Germany

**Kristin Wright-Bettner**
University of Colorado Boulder, USA

**Malte Ostendorff**
DFKI GmbH, Germany

**Maria Berger**
DFKI GmbH, Germany
List of Contributors

Mark A. Finlayson
Florida International University, USA

Martha Palmer
University of Colorado Boulder, USA

Martin Atkinson
Joint Research Centre of the European Commission, Italy

Michael Regan
University of New Mexico, USA

Mikka Wild
3pc Neue Kommunikation GmbH, Germany

Mustafa Ocal
Florida International University, USA

Paramita Mirza
Max Planck Institute for Informatics, Germany

Pavlína Kalm
University of New Mexico, USA

Peter Bourgonje
DFKI GmbH, Germany

Piek Vossen
Vrije Universiteit Amsterdam, the Netherlands

Rolf Fricke
Condat AG, Germany

Roxane Segers
Vrije Universiteit Amsterdam, the Netherlands

Susan W. Brown
University of Colorado Boulder, USA
List of Contributors

Tim O'Gorman
University of Massachusetts Amherst, USA

Tommaso Caselli
Rijksuniversiteit Groningen, the Netherlands

Vanni Zavarella
Joint Research Centre of the European Commission, Italy

William Croft
University of New Mexico, USA