

# Natural Language Processing for the Semantic Web

# Synthesis Lectures on the Semantic Web: Theory and Technology

## Editor

**Ying Ding, *Indiana University***

**Paul Groth, *Elsevier Labs***

*Synthesis Lectures on the Semantic Web: Theory and Application* is edited by Ying Ding of Indiana University and Paul Groth of Elsevier Labs. Whether you call it the Semantic Web, Linked Data, or Web 3.0, a new generation of Web technologies is offering major advances in the evolution of the World Wide Web. As the first generation of this technology transitions out of the laboratory, new research is exploring how the growing Web of Data will change our world. While topics such as ontology-building and logics remain vital, new areas such as the use of semantics in Web search, the linking and use of open data on the Web, and future applications that will be supported by these technologies are becoming important research areas in their own right. Whether they be scientists, engineers or practitioners, Web users increasingly need to understand not just the new technologies of the Semantic Web, but to understand the principles by which those technologies work, and the best practices for assembling systems that integrate the different languages, resources, and functionalities that will be important in keeping the Web the rapidly expanding, and constantly changing, information space that has changed our lives.

Topics to be included:

- Semantic Web Principles from linked-data to ontology design
- Key Semantic Web technologies and algorithms
- Semantic Search and language technologies
- The Emerging "Web of Data" and its use in industry, government and university applications
- Trust, Social networking and collaboration technologies for the Semantic Web
- The economics of Semantic Web application adoption and use
- Publishing and Science on the Semantic Web
- Semantic Web in health care and life sciences

### Natural Language Processing for the Semantic Web

Diana Maynard, Kalina Bontcheva, and Isabelle Augenstein

2016

### The Epistemology of Intelligent Semantic Web Systems

Mathieu d'Aquin and Enrico Motta

2016

### Entity Resolution in the Web of Data

Vassilis Christophides, Vasilis Efthymiou, and Kostas Stefanidis

2015

### Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description

Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixer

2015

### Semantic Mining of Social Networks

Jie Tang and Juanzi Li

2015

### Social Semantic Web Mining

Tope Omitola, Sebastián A. Ríos, and John G. Breslin

2015

### Semantic Breakthrough in Drug Discovery

Bin Chen, Huijun Wang, Ying Ding, and David Wild

2014

### Semantics in Mobile Sensing

Zhixian Yan and Dipanjan Chakraborty

2014

### Provenance: An Introduction to PROV

Luc Moreau and Paul Groth

2013

### Resource-Oriented Architecture Patterns for Webs of Data

Brian Sletten

2013

### Aaron Swartz's A Programmable Web: An Unfinished Work

Aaron Swartz

2013

### Incentive-Centric Semantic Web Application Engineering

Elena Simperl, Roberta Cuel, and Martin Stein

2013

### Publishing and Using Cultural Heritage Linked Data on the Semantic Web

Eero Hyvönen

2012

### VIVO: A Semantic Approach to Scholarly Networking and Discovery

Katy Börner, Michael Conlon, Jon Corson-Rikert, and Ying Ding

2012

### Linked Data: Evolving the Web into a Global Data Space

Tom Heath and Christian Bizer

2011

© Springer Nature Switzerland AG 2022  
Reprint of original edition © Morgan & Claypool 2017

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.

Natural Language Processing for the Semantic Web

Diana Maynard, Kalina Bontcheva, and Isabelle Augenstein

ISBN: 978-3-031-79473-5      paperback  
ISBN: 978-3-031-79474-2      ebook

DOI 10.1007/978-3-031-79474-2

A Publication in the Springer series

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND TECHNOLOGY*

Lecture #15

Series Editors: Ying Ding, *Indiana University*

Paul Groth, *Elsevier Labs*

Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*

Series ISSN

Print 2160-4711    Electronic 2160-472X

# Natural Language Processing for the Semantic Web

Diana Maynard  
University of Sheffield

Kalina Bontcheva  
University of Sheffield

Isabelle Augenstein  
University College London

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND  
TECHNOLOGY #15*

## ABSTRACT

This book introduces core natural language processing (NLP) technologies to non-experts in an easily accessible way, as a series of building blocks that lead the user to understand key technologies, why they are required, and how to integrate them into Semantic Web applications. Natural language processing and Semantic Web technologies have different, but complementary roles in data management. Combining these two technologies enables structured and unstructured data to merge seamlessly. Semantic Web technologies aim to convert unstructured data to meaningful representations, which benefit enormously from the use of NLP technologies, thereby enabling applications such as connecting text to Linked Open Data, connecting texts to each other, semantic searching, information visualization, and modeling of user behavior in online networks.

The first half of this book describes the basic NLP processing tools: tokenization, part-of-speech tagging, and morphological analysis, in addition to the main tools required for an information extraction system (named entity recognition and relation extraction) which build on these components. The second half of the book explains how Semantic Web and NLP technologies can enhance each other, for example via semantic annotation, ontology linking, and population. These chapters also discuss sentiment analysis, a key component in making sense of textual data, and the difficulties of performing NLP on social media, as well as some proposed solutions. The book finishes by investigating some applications of these tools, focusing on semantic search and visualization, modeling user behavior, and an outlook on the future.

## KEYWORDS

natural language processing, semantic web, semantic search, social media analysis, text mining, linked data, entity linking, information extraction, sentiment analysis

*This book is a timely exposition of natural language processing and its role and importance for those seeking to apply semantic technologies. Clearly written with good coverage of the key topics and a comprehensive bibliography, the text will be invaluable for semantic web practitioners and more widely.*

*Prof John Davies  
BT Research & Technology  
Adastral Park UK  
November 2016*

# Contents

	<b>Acknowledgments</b> . . . . .	<b>xiii</b>
<b>1</b>	<b>Introduction</b> . . . . .	<b>1</b>
1.1	Information Extraction . . . . .	2
1.2	Ambiguity . . . . .	4
1.3	Performance . . . . .	5
1.4	Structure of the Book . . . . .	7
<b>2</b>	<b>Linguistic Processing</b> . . . . .	<b>9</b>
2.1	Introduction . . . . .	9
2.2	Approaches to Linguistic Processing . . . . .	9
2.3	NLP Pipelines . . . . .	10
2.4	Tokenization . . . . .	12
2.5	Sentence Splitting . . . . .	14
2.6	POS Tagging . . . . .	15
2.7	Morphological Analysis and Stemming . . . . .	16
2.7.1	Stemming . . . . .	17
2.8	Syntactic Parsing . . . . .	19
2.9	Chunking . . . . .	21
2.10	Summary . . . . .	23
<b>3</b>	<b>Named Entity Recognition and Classification</b> . . . . .	<b>25</b>
3.1	Introduction . . . . .	25
3.2	Types of Named Entities . . . . .	26
3.3	Named Entity Evaluations and Corpora . . . . .	27
3.4	Challenges in NERC . . . . .	27
3.5	Related Tasks . . . . .	29
3.6	Approaches to NERC . . . . .	30
3.6.1	Rule-based Approaches to NERC . . . . .	30
3.6.2	Supervised Learning Methods for NERC . . . . .	31
3.7	Tools for NERC . . . . .	33



3.8	NERC on Social Media .....	34
3.9	Performance .....	34
3.10	Summary .....	35
<b>4</b>	<b>Relation Extraction .....</b>	<b>37</b>
4.1	Introduction .....	37
4.2	Relation Extraction Pipeline .....	37
4.3	Relationship between Relation Extraction and other IE Tasks .....	39
4.4	The Role of Knowledge Bases in Relation Extraction .....	40
4.5	Relation Schemas .....	41
4.6	Relation Extraction Methods .....	42
4.6.1	Bootstrapping Approaches .....	42
4.7	Rule-based Approaches .....	44
4.8	Supervised Approaches .....	45
4.9	Unsupervised Approaches .....	46
4.10	Distant Supervision Approaches .....	47
4.10.1	Universal Schemas .....	48
4.10.2	Hybrid Approaches .....	49
4.11	Performance .....	49
4.12	Summary .....	50
<b>5</b>	<b>Entity Linking .....</b>	<b>53</b>
5.1	Named Entity Linking and Semantic Linking .....	54
5.2	NEL Datasets .....	54
5.3	LOD-based Approaches .....	55
5.3.1	DBpedia Spotlight .....	55
5.3.2	YODIE: A LOD-based Entity Disambiguation Framework .....	56
5.3.3	Other Key LOD-based Approaches .....	57
5.4	Commercial Entity Linking Services .....	58
5.5	NEL for Social Media Content .....	59
5.6	Discussion .....	60
<b>6</b>	<b>Automated Ontology Development .....</b>	<b>61</b>
6.1	Introduction .....	61
6.2	Basic Principles .....	61
6.3	Term Extraction .....	63

6.3.1	Approaches Using Distributional Knowledge	64
6.3.2	Approaches Using Contextual Knowledge	65
6.4	Relation Extraction	66
6.4.1	Clustering Methods	66
6.4.2	Semantic Relations	66
6.4.3	Lexico-syntactic Patterns	68
6.4.4	Statistical Techniques	68
6.5	Enriching Ontologies	69
6.6	Ontology Development Tools	70
6.6.1	Text2Onto	70
6.6.2	SPRAT	70
6.6.3	FRED	70
6.6.4	Semi-automatic Ontology Creation	71
6.7	Summary	71
<b>7</b>	<b>Sentiment Analysis</b>	<b>73</b>
7.1	Introduction	73
7.2	Issues in Opinion Mining	75
7.3	Opinion-Mining Subtasks	76
7.3.1	Polarity Recognition	76
7.3.2	Opinion Target Detection	76
7.3.3	Opinion Holder Detection	77
7.3.4	Sentiment Aggregation	77
7.3.5	Further Linguistic Subcomponents	78
7.4	Emotion Detection	79
7.5	Methods for Opinion Mining	81
7.6	Opinion Mining and Ontologies	83
7.7	Opinion-Mining Tools	85
7.8	Summary	86
<b>8</b>	<b>NLP for Social Media</b>	<b>87</b>
8.1	Social Media Streams: Characteristics, Challenges, and Opportunities	88
8.2	Ontologies for Representing Social Media Semantics	90
8.3	Semantic Annotation of Social Media	92
8.3.1	Keyphrase Extraction	92
8.3.2	Ontology-based Entity Recognition in Social Media	93

8.3.3	Event Detection	99
8.3.4	Sentiment Detection and Opinion Mining	100
8.3.5	Cross-media Linking	101
8.3.6	Rumor Analysis	102
8.3.7	Discussion	103
<b>9</b>	<b>Applications</b>	<b>107</b>
9.1	Semantic Search	107
9.1.1	What is Semantic Search?	108
9.1.2	Why Semantic Full-text Search?	109
9.1.3	Semantic Search Queries	110
9.1.4	Relevance Scoring and Retrieval	111
9.1.5	Semantic Search Full-text Platforms	111
9.1.6	Ontology-based Faceted Search	114
9.1.7	Form-based Semantic Search Interfaces	116
9.1.8	Semantic Search over Social Media Streams	118
9.2	Semantic-Based User Modeling	122
9.2.1	Constructing Social Semantic User Models from Semantic Annotations	122
9.2.2	Discussion	125
9.3	Filtering and Recommendations for Social Media Streams	125
9.4	Browsing and Visualization of Social Media Streams	126
9.5	Discussion and Future Work	132
<b>10</b>	<b>Conclusions</b>	<b>135</b>
10.1	Summary	135
10.2	Future Directions	135
10.2.1	Cross-media Aggregation and Multilinguality	136
10.2.2	Integration and Background Knowledge	137
10.2.3	Scalability and Robustness	137
10.2.4	Evaluation, Shared Datasets, and Crowdsourcing	138
	<b>Bibliography</b>	<b>141</b>
	<b>Authors' Biographies</b>	<b>179</b>

# Acknowledgments

This work was supported by funding from PHEME, DecarboNet, COMRADES, uComp, and the Engineering and Physical Sciences Research Council (grant EP/I004327/1). The authors also wish to thank colleagues from the GATE team, listed here in alphabetical order: Hamish Cunningham, Leon Derczynski, Genevieve Gorrell, Mark Greenwood, Johann Petrak, Angus Roberts, Ian Roberts, Dominic Rout, Wim Peters; and the many other colleagues who have contributed fruitful discussions.

Diana Maynard, Kalina Bontcheva, and Isabelle Augenstein  
November 2016