

# Social Semantic Web Mining

# Synthesis Lectures on the Semantic Web: Theory and Technology

## Editors

**Ying Ding**, *Indiana University*

**Paul Groth**, *VU University Amsterdam*

## Founding Editor Emeritus

**James Hendler**, *Rensselaer Polytechnic Institute*

Synthesis Lectures on the Semantic Web: Theory and Application is edited by Ying Ding of Indiana University and Paul Groth of VU University Amsterdam. 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

### 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 2015

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.

Social Semantic Web Mining

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

ISBN: 978-3-031-79458-2      paperback

ISBN: 978-3-031-79459-9      ebook

DOI 10.1007/978-3-031-79459-9

A Publication in the Springer series

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

Lecture #10

Series Editors: Ying Ding, *Indiana University*

Paul Groth, *VU University Amsterdam*

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

Series ISSN

Print 2160-4711    Electronic 2160-472X

# Social Semantic Web Mining

Tope Omitola  
University of Southampton

Sebastián A. Ríos  
University of Chile

John G. Breslin  
National University of Ireland Galway

*SYNTHESIS LECTURES ON THE SEMANTIC WEB: THEORY AND  
TECHNOLOGY #10*

## ABSTRACT

The past ten years have seen a rapid growth in the numbers of people signing up to use Web-based social networks (hundreds of millions of new members are now joining the main services each year) with a large amount of content being shared on these networks (tens of billions of content items are shared each month). With this growth in usage and data being generated, there are many opportunities to discover the knowledge that is often inherent but somewhat hidden in these networks. Web mining techniques are being used to derive this hidden knowledge. In addition, the Semantic Web, including the Linked Data initiative to connect previously disconnected datasets, is making it possible to connect data from across various social spaces through common representations and agreed upon terms for people, content items, etc.

In this book, we detail some current research being carried out to semantically represent the implicit and explicit structures on the Social Web, along with the techniques being used to elicit relevant knowledge from these structures, and we present the mechanisms that can be used to intelligently mesh these semantic representations with intelligent knowledge discovery processes. We begin this book with an overview of the origins of the Web, and then show how web intelligence can be derived from a combination of web and Social Web mining. We give an overview of the Social and Semantic Webs, followed by a description of the combined Social Semantic Web (along with some of the possibilities it affords), and the various semantic representation formats for the data created in social networks and on social media sites.

Provenance and provenance mining is an important aspect here, especially when data is combined from multiple services. We will expand on the subject of provenance and especially its importance in relation to social data. We will describe extensions to social semantic vocabularies specifically designed for community mining purposes (SIOCM). In the last three chapters, we describe how the combination of web intelligence and social semantic data can be used to derive knowledge from the Social Web, starting at the community level (macro), and then moving through group mining (meso) to user profile mining (micro).

## KEYWORDS

World Wide Web, Web Mining, Web Intelligence, Semantic Web, Social Web, Social Semantic Web, Provenance, Knowledge Discovery, Knowledge Management

*We would like to dedicate this book to our families,  
who have supported us and been patient with us throughout  
the writing of this book.*

# Contents

<b>Acknowledgments</b> .....	<b>xiii</b>
<b>Grant Aid</b> .....	<b>xv</b>
<b>1 Introduction and the Web</b> .....	<b>1</b>
1.1 Introduction .....	1
1.2 The World Wide Web .....	2
1.2.1 History and Evolution of the Web .....	2
1.2.2 A Short Explanation of How the Web Works .....	3
1.3 Global Impact of the Internet and the Web .....	4
1.4 Web Mining, the Social Web, and the Semantic Web .....	5
1.5 Conclusion .....	6
<b>2 Web Mining</b> .....	<b>7</b>
2.1 Mining the World Wide Web .....	7
2.1.1 Different Types of Web Mining Processes .....	8
2.1.2 Directed vs. Undirected Web Mining Processes .....	10
2.1.3 Supervised vs. Unsupervised Algorithms .....	11
2.2 Traditional Framework for Offline Website Enhancements .....	11
2.2.1 Resource Discovery/Data Selection .....	14
2.2.2 Extraction/Preprocessing .....	15
2.2.3 Data Generalization .....	19
2.2.4 Analysis/Evaluation .....	19
2.3 Clustering Algorithms for Web Mining .....	19
2.3.1 Clustering Methods .....	20
2.3.2 Self-organizing Feature Maps .....	20
2.3.3 K-means Clustering Algorithm .....	22
2.3.4 Decision Trees .....	24
2.4 Dissimilarity and Similarity Measures .....	25
2.5 Latent Semantics Using LSA Techniques .....	27
2.6 Conclusion .....	30



<b>3</b>	<b>The Social Web</b>	<b>31</b>
3.1	What Is the Social Web?	31
3.1.1	A Brief History of the Social Web	32
3.1.2	Online Social Networks	33
3.1.3	Social Media Creation and Sharing	34
3.1.4	Tagging, Folksonomies, and Hashtags	34
3.1.5	Crowdsourcing and Citizen Sensors	35
3.1.6	Limitations with Social Spaces	35
3.2	Conclusion	36
<b>4</b>	<b>The Semantic Web</b>	<b>37</b>
4.1	Introduction	37
4.1.1	From Syntax to Semantics	37
4.1.2	A Great Big Graph of Metadata and Vocabularies	39
4.1.3	Issues with Vocabulary Creation	40
4.1.4	Representation Formats	41
4.1.5	Semantic Web and SEO	43
4.1.6	Comparisons with Microformats and Microdata	43
4.2	Conclusion	43
<b>5</b>	<b>The Social Semantic Web</b>	<b>45</b>
5.1	Introduction	45
5.2	The Social Semantic Web	45
5.3	Some Potential Uses of the Social Semantic Web	45
5.4	Integrating Existing Social Spaces on the Web	47
5.5	Extension to Further Social Spaces	48
5.6	Standard, Interoperable Descriptions of Social Data	49
5.7	The Long Tail of Information Domains	49
5.8	Social Semantic Web Vocabularies	50
5.8.1	FOAF—Friend-of-a-Friend	50
5.8.2	hCard and XFN	52
5.8.3	SIOC—Semantically Interlinked Online Communities	53
5.8.4	Other Ontologies	58
5.9	Conclusion	59

<b>6</b>	<b>Social Semantic Web Mining</b>	<b>61</b>
6.1	Introduction	61
6.2	Provenance	61
6.2.1	Rumors and Dissimulation in Online Social Networks	61
6.2.2	What is Provenance?	62
6.2.3	Modeling Provenance	64
6.2.4	Provenance of Social Data	64
6.2.5	Provenance on the Web	65
6.3	SIOCM	70
6.3.1	Ontological Representation of Online Social Communities	70
6.4	Conclusion	72
<b>7</b>	<b>Social Semantic Web Mining of Communities</b>	<b>73</b>
7.1	Introduction	73
7.2	Purpose Evolution	73
7.3	Goals as a Measure of Purpose Accomplishment	74
7.4	Mining Goals from Texts: Concept-based Text Mining	75
7.4.1	Fuzzy Logic for Goals Classification	75
7.4.2	Identification and Definition of Goals	76
7.5	Real Application of Community Purpose Monitoring	77
7.5.1	The Plexilandia Community	77
7.5.2	Concept-based Text Mining Application	78
7.5.3	Analysis of Results	78
7.5.4	Results Evaluation	79
7.5.5	Applying SIOCM to Store Goal Definitions	80
7.5.6	Extracting Topic-filtered Networks using SIOCM	81
7.6	Conclusion	83
<b>8</b>	<b>Social Semantic Web Mining of Groups</b>	<b>85</b>
8.1	Introduction	85
8.2	Previous Work	86
8.2.1	Topic-based Social Network Analysis	86
8.2.2	Social Network Analysis on the Dark Web	87
8.3	Methodology for Group Key Member Discovery	88
8.3.1	Basic Notation	88
8.3.2	Topic Modeling	89

8.3.3	Network Configuration . . . . .	90
8.3.4	Topic-based Network Filtering . . . . .	90
8.3.5	Network Construction . . . . .	91
8.4	Experimental Setup and Results . . . . .	92
8.4.1	Results and Discussion . . . . .	93
8.5	Conclusion . . . . .	99
<b>9</b>	<b>Social Semantic Web Mining of Users . . . . .</b>	<b>101</b>
9.1	Introduction . . . . .	101
9.2	Modeling a Distributed User Profile with Interests . . . . .	102
9.3	Related Work Mining User Profiles on the Social Semantic Web . . . . .	103
9.4	Representing User Interest Profiles . . . . .	104
9.5	Leveraging the Provenance of User Data . . . . .	107
9.6	Interests on the Web of Data . . . . .	108
9.7	Interest Mining on the Social Web . . . . .	108
9.7.1	Bag-of-words vs. Disambiguated Entities . . . . .	108
9.7.2	Time Decay of Interests . . . . .	109
9.7.3	Categories vs. Resources . . . . .	110
9.7.4	Provenance-based Features . . . . .	111
9.8	An Architecture for Aggregating User Profiles of Interests on the Social Web . . . . .	113
9.8.1	Evaluation of Aggregated User Profiles . . . . .	117
9.9	Conclusion . . . . .	118
<b>10</b>	<b>Conclusions . . . . .</b>	<b>119</b>
10.1	Summary . . . . .	119
10.2	Future Work . . . . .	119
	<b>Bibliography . . . . .</b>	<b>121</b>
	<b>Authors' Biographies . . . . .</b>	<b>137</b>

# Acknowledgments

Tope Omitola would like to express gratitude for the managerial support of Professor Sir Nigel Shadbolt of the Web and Internet Science Group, University of Southampton and of Dr. John Davies of British Telecommunications during the writing of this book. He would also like to acknowledge the support of EPSRC (Grant Number EP/K503770/1) during this time.

Sebastián Ríos would like to thank the continuous support from the Complex Engineering Systems Institute (ICM: P-05-004-F, CONICYT: FBO16) and the research grant for Initiation into Research (FONDECYT), project code 11090188, entitled “Semantic Web Mining Techniques to Study Enhancements of Virtual Communities.”

John Breslin would like to acknowledge input from his collaborators Dr. Fabrizio Orlandi and Dr. Alexandre Passant. He would also like to acknowledge that his work on this publication has emanated from research supported in part by research grants from Science Foundation Ireland (SFI) under Grant Number SFI/08/CE/I1380 (DERI Lón 2) and also under Grant Number SFI/12/RC/2289 (Insight).

Tope Omitola, Sebastián A. Ríos, and John G. Breslin  
January 2015

# Grant Aid

This publication was grant-aided by the Publications Fund of  
**National University of Ireland Galway**

Rinneadh maoiniú ar an bhfoilseacháin seo trí Chiste Foilseacháin  
**Ollscoil na hÉireann Gaillimh**