

Health Information Science

Series Editor

Yanchun Zhang

More information about this series at <http://www.springer.com/series/11944>

Kerstin Denecke

Health Web Science

Social Media Data for Healthcare

Kerstin Denecke
Innovation Center for Computer Assisted Surgery
Universität Leipzig
Leipzig, Sachsen, Germany

Health Information Science
ISBN 978-3-319-20581-6 ISBN 978-3-319-20582-3 (eBook)
DOI 10.1007/978-3-319-20582-3

Library of Congress Control Number: 2015945592

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

“In the past, we went to the medical doctor (Dr. med.) when we became sick. Today, we are asking Dr. Google.”

Preface

The web became a rich source of personal information during the last years. People are twittering, blogging and chatting online. Feelings, experiences or latest news are posted everywhere in the web. Face-to-face communication seems to be replaced or at least extended by the new communication tools. Since this new communication style provides insights into thoughts of individuals and their behaviour, reflects trends and situations in the real world, it became focus of research, politics and economics. For instance, first hints to disease outbreaks or political changes can be identified from this data. Beyond, companies discover the benefits of analysing social web communication for developing and adapting marketing strategies.

The developments of the web and mobile communications led to the rise of a new research field summarized under the term *Web Science*. What is Web Science and more specifically, what characterizes Health Web Science—the topic of this book?

The American computer scientist Ben Shneiderman claims that “Web Science” is a “new way of thinking about computer science” [1]. Tim Berners-Lee becomes more specific: “Web Science must be inherently interdisciplinary; its goal is both, to understand the growth of the web and to create approaches that allow new powerful and more beneficial patterns to occur” [2]. For short, not only processing the information available on the web is part of the Web Science research agenda, but also social issues such as trust, reputation and privacy play an important role. Starting from this understanding, we define *Health Web Science* as follows:

Health Web Science is the medico-socio-technical science that investigates how the web evolves with respect to health issues, how health related data provided through the web can be processed and how tools that make use of web technology can be used in healthcare.

In other words, Health Web Science is about patients, doctors and health carers, the information they are providing in the web, the tools they can use to communicate on health issues and about methods to support in analysing the provided health information. Health Web Science targets at understanding how the web shapes and is shaped by medicine, healthcare and people, and how the web interacts and impacts on health [3].

With an increased interest in e-Health, Health 2.0, Medicine 2.0 and the recent birth of the discipline of Web Science, this book targets at introducing the field of Health Web Science with its facets in use cases and at presenting methods for gathering information from written medical social media data. How these methods can be exploited for social media analysis will be shown through example applications. The influence factors discussed at the end of the book serve as entry point for future developments in research and discussion.

This book is the result of my research activities at the L3S Research Center in Hannover and at the Innovation Center for Computer Assisted Surgery in Leipzig. Collaborations with colleagues in the EU-funded research projects LivingKnowledge and M-Eco motivated to dig into depth of this extremely interesting field of Health Web Science. Thanks to Yihan Deng who provided code to analyse linguistic characteristics in Chap. 6. The work presented in Chap. 11 is a result of the work of the M-Eco consortium whom I would like to thank for the interesting collaboration. Beyond, I would like to thank the members of the IMIA Social Media Working group who provide a very productive platform for discussion and joint research. Last but not least, I would like to thank my friends and family for support on multiple levels as well as for many valuable suggestions for improving the presentation in this book.

Braunschweig, Germany
April 2015

Kerstin Denecke

Contents

Part I Introduction

1	The Medical Web and Its Authors	3
1.1	What is Medical Social Media?	3
1.2	Personas.....	4
2	Applications of Social Media in Healthcare	7
2.1	Social Media Support for Patient Treatment	8
2.2	Social Media for Preventing Diseases and Promoting Health Activity	9
2.3	Social Media Tools for Networking, Information Gathering and Exchange	10
2.4	Social Media Tools for Managing Knowledge	12
2.5	Research and Health Monitoring Using Social Media Data.....	13
3	Objectives and Structure of the Book	17
3.1	Objectives.....	17
3.2	Research Questions and Structure of the Book	18
3.3	Material and Methods	19

Part II Medical Social Media: Content, Characteristics, Appearance

4	Types of Medical Social Media	25
5	Excursus: Content and Language in Clinical Narratives	29
5.1	Content of Clinical Documents	29
5.2	The Sublanguage in Clinical Narratives.....	31
6	Content and Language in Medical Social Media	33
6.1	Content of Medical Blogs.....	33
6.2	Word Usage in Medical Social Media: A Linguistic Analysis	36
6.2.1	Sublanguage Analysis of Medical Social Media.....	36
6.2.2	Statistics of Word Classes	37

6.2.3	Sentiment Terms	40
6.2.4	Semantic Categories	41
6.2.5	Semantic Patterns	42
6.2.6	Additional Characteristics of Blogs.....	43
6.3	Implications for Automatic Processing.....	44

Part III Information Discovery from Medical Social Media

7	Textanalytic-Enabled Healthcare Applications: Requirements and Architecture	51
7.1	Textanalysis Tasks and Requirements	51
7.1.1	End-User Requirements	52
7.1.2	System Developer Requirements	53
7.1.3	Architecture Requirements	53
7.2	Existing NLP Tools and Frameworks	54
7.3	Architecture for Textanalytic-Enabled Applications	55
7.3.1	Client	56
7.3.2	Server	56
7.3.3	Processing Services.....	57
7.3.4	Resources	58
7.4	Benefits and Shortcomings	59
8	Information Extraction from Medical Social Media	61
8.1	Problem Description: Extraction of Medical Concepts.....	61
8.2	Tools and Resources for Information Extraction	63
8.2.1	Information Extraction Process	63
8.2.2	Medical Ontologies.....	65
8.2.3	Information Extraction Tools for Medical and Biomedical Texts.....	66
8.3	Extracting Medical Concepts from Medical Social Media: A Qualitative Study	68
8.3.1	Study Design.....	68
8.3.2	Mapping Observations	69
8.4	Error Analysis and Open Issues	70
8.4.1	Mapping Errors	70
8.4.2	Discussion of Applicability of MetaMap and cTakes on Social Media	71
8.4.3	Improvements of NER Tools for Processing Medical Social Media	72
8.4.4	Discussion of the Study Design.....	73
9	Relation Extraction	75
9.1	Extracting Relationships: Definitions and Approaches.....	75
9.2	ReVerb for Medical Relation Extraction	77
9.3	Quality of Relation Extraction.....	78
9.3.1	Evaluation Setup and Objectives.....	78

9.3.2	Relation Types in the Data Set	78
9.3.3	Evaluation Results for Relation Extraction.....	80
10	Sentiment Analysis from Medical Texts	83
10.1	The Notions of Sentiment in Medicine.....	83
10.2	Approaches to Sentiment Analysis	86
10.2.1	General Overview	86
10.2.2	Sentiment Lexicons.....	87
10.2.3	Sentiment Analysis in Medical Context	88
10.2.4	Summary of Medical Opinion Mining Approaches.....	91
10.3	Application Areas	92
10.4	Future Research Aspects of Sentiment Analysis in Medicine.....	96
10.4.1	Domain-Specific Sentiment Source.....	96
10.4.2	Level of Analysis	97
10.4.3	Explicit and Implicit Sentiments	98
10.4.4	Medical Sentiment Analysis: The Future.....	98
 Part IV Applications		
11	Social Media for Health Monitoring.....	101
11.1	M-Eco System	102
11.1.1	Content Collection and Document Analysis Component.....	103
11.1.2	Signal Generation Component	104
11.1.3	Recommendation Component.....	105
11.1.4	User Interface and Visualization	106
11.2	Experiences and Lessons Learnt.....	107
12	Diversity-Aware Search	109
12.1	Diversity in Web Search and Medical Social Media.....	110
12.2	Diversity Dimensions: Information Type and Aspects Considered	111
12.3	Architecture for Diversity-Aware Search in Medical Social Media.....	111
12.3.1	Content Collector and Indexer	112
12.3.2	Domain Filter	113
12.3.3	Concept Annotator.....	113
12.3.4	Diversity Assessor	113
12.3.5	Ranking	114
12.3.6	Visualization	115
12.4	User Satisfaction with Diverse Retrieval Engine: Experiments	115
12.4.1	Data Set	115
12.4.2	Evaluation Objectives and Setup.....	115
12.4.3	Results	117
12.5	Diversity in Medical Social Media: Lessons Learnt.....	118

Part V Influence Factors

13 Implementation Obstacles	123
13.1 A Broad Range of Challenges	123
13.2 Legal Issues.....	124
14 Ethics in Health Web Science	127
14.1 Ethics in the Big Data Era	128
14.2 Ethical Issues of Social Media in Healthcare	129
14.2.1 Ethical Issues in Medical Social Media Research.....	130
14.2.2 Social Media and the Patient-Doctor Relationship.....	131
14.3 Implications for Implementing Social Media Applications	133
15 Integrating Clinical and Social Media Data	137
15.1 Digital Patient Modelling and Model-Based Decision Support	138
15.2 Requirements for Data Integration	140
15.3 Data Integration Using Digital Patient Modelling	141
15.3.1 Generating Probabilistic Graphical Models	142
15.3.2 From a General Model to Patient-Specific Models	142
15.3.3 Vision of Digital Patient Models in Practice	144
16 Concluding Remarks	147
Glossary	149
References.....	155
Index.....	167