Lecture Notes in Computer Science **9114**

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

More information about this series at http://www.springer.com/series/7409

Philipp Cimiano · Flavius Frasincar
Geert-Jan Houben · Daniel Schwabe (Eds.)

Engineering the Web in the Big Data Era

15th International Conference, ICWE 2015 Rotterdam, The Netherlands, June 23–26, 2015 Proceedings

Springer

Editors
Philipp Cimiano
Bielefeld University
Bielefeld
Germany

Flavius Frasincar
Eramus University Rotterdam
Rotterdam
The Netherlands

Geert-Jan Houben
Delft University of Technology
Delft
The Netherlands

Daniel Schwabe
Pontifical Catholic University
of Rio de Janeiro
Rio de Janeiro
Brazil

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-19889-7 ISBN 978-3-319-19890-3 (eBook)
DOI 10.1007/978-3-319-19890-3

Library of Congress Control Number: 2015942219

LNCS Sublibrary: SL3 – Information Systems and Application, incl. Internet/Web, and HCI

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

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

Preface

The Web has become such an important tool in our daily lives that it is difficult to imagine living in a world without it. In addition to the original goal of satisfying an information need, the Web has grown to become a popular platform for doing business, education, entertainment, communication, etc. Developing software for the Web needs to meet an increasing number of challenges that have not been previously faced by any software platform, in good part spurred by its functional versatility, as well as its ubiquity fostered by the ever-growing presence of Web-enabled electronic devices. Web engineering aims to address these challenges by adapting software engineering to the Web context, as well as providing new methodologies, techniques, and tools to address Web application design, development, and maintenance.

This volume contains the full research papers, short papers, industry papers, the descriptions of posters, demonstrations, and tutorials, and extended abstracts for the keynotes of the 15th International Conference on Web Engineering (ICWE 2015), held during June 23–25, 2015, in Rotterdam, The Netherlands.

ICWE is the flagship conference for the Web engineering community. Previous editions of ICWE took place at Toulouse, France (2014), Aalborg, Denmark (2013), Berlin, Germany (2012), Paphos, Cyprus (2011), Vienna, Austria (2010), San Sebastian, Spain (2009), Yorktown Heights, NY, USA (2008), Como, Italy (2007), Palo Alto, CA, USA (2006), Sydney, Australia (2005), Munich, Germany (2004), Oviedo, Spain (2003), Santa Fe, Argentina (2002), and Cáceres, Spain (2001). The 2015 edition of ICWE was centered on the theme of "Engineering the Web in the Big Data Era," hereby highlighting the impact big data has on Web engineering research today. Big data promises new data usages contributing to a change in our daily practices. As the Web is a valuable producer and consumer of big data, it is imperative to analyze the consequences and impact of the big data paradigm in the Web engineering field.

ICWE 2015 presented eight research tracks, namely, Web Application Modeling and Engineering, Mobile Web Applications, Social Web Applications, Semantic Web Applications, Quality and Accessibility Aspects of Web Applications, Web Applications Composition and Mashups, Web User Interfaces, and Security and Privacy in Web Applications, and an industry track aimed at concentrating expertise and contributing to the identity of the Web engineering community.

The ICWE 2015 edition received 110 submissions distributed over nine tracks (eight research tracks plus one industry track), out of which the Program Committee selected 26 full research papers (24% acceptance rate), 11 short papers (34% acceptance rate), and seven industry papers. Additionally, the Program Committee accepted 11 demonstrations, six posters, and four contributions to the PhD symposium, where PhD students received personalized advice and guidance on their work from senior researchers in the Web engineering field. Also accepted were two tutorials lecturing on the advanced topics of (1) "Enabling the Web of (Linked Open) Data," and (2) "The Web of Data for E-Commerce: Schema.org and GoodRelations for Researchers and

Practitioners," three workshops entitled (1) First International Workshop on Natural Language Processing for Informal Text (NLPIT 2015), (2) First Workshop on Pervasive Web Technologies, Trends and Challenges (PEWET 2015), and (3) First International Workshop in Mining the Social Web (SoWeMine 2015), and one challenge on Rapid Mashup.

The excellent program would not have been possible without the support of the many people who helped with the organization of this event. We would like to thank all the track, tutorial, demonstrations/poster, workshop, PhD symposium, sponsorship, publicity, metadata, and registration chairs for their hard work. Our thanks also goes to Enrique Alfonseca (Google Research), Peter Mika (Yahoo Research), and Mike Thelwall (University of Wolverhampton), who accepted to be our keynote speakers. Special thanks are extended to Oscar Diaz and Peter Dolog for their advice and encouragement in setting up ICWE 2015 in Rotterdam. We would like to also thank Michael Krug and Martin Gaedke for hosting the conference website, and Kim Schouten and Damir Vandic for building and maintaining the website. We are grateful to our local organizers Ursula David and Marianne Kroek-Buijs for their logistical support, and Springer for publishing this volume. In addition, we thank the reviewers for their meticulous work that allowed us to select the best papers to be presented at ICWE 2015. Last, but not least, we would like to thank the authors who submitted their work to this conference and all the participants who contributed to the success of this event.

June 2015

Philipp Cimiano
Flavius Frasincar
Geert-Jan Houben
Daniel Schwabe

Organization

Technical Committee

General Chair

Flavius Frasincar	Erasmus University Rotterdam, The Netherlands

Vice General Chair

Geert-Jan Houben	Delft University of Technology, The Netherlands

Program Committee Chairs

Philipp Cimiano	Bielefeld University, Germany
Daniel Schwabe	Pontifical Catholic University of Rio de Janiero, Brazil

Track Chairs

Web Application Modeling and Engineering

Oscar Pastor	Universitat Politècnica de València, Spain
Marco Winckler	University Paul Sabatier, France

Mobile Web Applications

In-Young Ko	Korea Advanced Institute of Science and Technology, South Korea
Volker Gruhn	University of Duisburg-Essen, Germany

Social Web Applications

Alessandro Bozzon	Delft University of Technology, The Netherlands
Sven Casteleyn	University Jaume I of Castellon, Spain

Semantic Web Applications

Sören Auer	University of Bonn, Germany
Heiko Paulheim	University of Mannheim, Germany

Quality and Accessibility Aspects of Web Applications

Luis Olsina	National University of La Pampa, Argentina
Markel Vigo	The University of Manchester, UK

Web Applications Composition and Mashups

Cinzia Cappiello	Politecnico di Milano, Italy
Cesare Pautasso	University of Lugano, Switzerland

Web User Interfaces

Ali Mesbah	University of British Columbia, Canada
Fabio Paterno	University of Pisa, Italy

Security and Privacy in Web Applications

Martin Gaedke	Chemnitz University of Technology, Germany
Christian Hammer	Saarland University, Germany

Industry

Marco Brambilla	Politecnico di Milano, Italy
Maja Vukovic	IBM T.J. Watson Research Center, USA

Tutorials Chairs

Franciska de Jong	Erasmus University Rotterdam, The Netherlands
Moira Norrie	ETH Zurich, Switzerland

Workshops Chairs

Florian Daniel	University of Trento, Italy
Oscar Diaz	University of the Basque Country, Spain

Demonstrations and Posters Chairs

Roberto De Virgilio	Roma Tre University, Italy
Gustavo Rossi	National University of La Plata, Argentina

PhD Symposium Chairs

Irene Garrigos	University of Alicante, Spain
Manuel Wimmer	Vienna University of Technology, Austria

Program Committee

Web Application Modeling and Engineering

Luciano Baresi	Politecnico di Milano, Italy
Hubert Baumeister	Technical University of Denmark, Denmark
Devis Bianchini	University of Brescia, Italy
Alessandro Bozzon	Delft University of Technology, The Netherlands
Marco Brambilla	Politecnico di Milano, Italy

Jordi Cabot	Inria-École des Mines de Nantes, France
Richard Chbeir	LIUPPA Laboratory, France
Florian Daniel	University of Trento, Italy
Oscar Diaz	University of the Basque Country, Spain
Schahram Dustdar	Vienna University of Technology, Austria
Marina Egea	Indra, Spain
Sergio Firmenich	UNLP and CONICET, Argentina
Flavius Frasincar	Erasmus University Rotterdam, The Netherlands
Piero Fraternali	Politecnico di Milano, Italy
Jose Luis Garrido	University of Granada, Spain
Irene Garrigos	University of Alicante, Spain
Sujoy Gupta	AppDirect, USA
Guy-Vincent Jourdan	University of Ottawa, Canada
Gerti Kappel	Vienna University of Technology, Austria
Alexander Knapp	Universität Augsburg, Germany
Maristella Matera	Politecnico di Milano, Italy
Santiago Melia	Universidad de Alicante, Spain
Jose Ignacio Panach Navarrete	Universitat de València, Spain
Vicente Pelechano	Universitat Politècnica de València, Spain
Alfonso Pierantonio	University of L'Aquila, Italy
Werner Retschitzegger	Johannes Kepler University Linz, Austria
Gustavo Rossi	UNLP, Argentina
Fernando Sánchez	Universidad de Extremadura, Spain
Daniel Schwabe	Pontifical Catholic University of Rio de Janiero, Brazil
Francisco Valverde	Universitat Politècnica de València, Spain
Manuel Wimmer	Vienna University of Technology, Austria
Gefei Zhang	Celonis GmbH, Germany
Jürgen Ziegler	University of Duisburg-Essen, Germany

Mobile Web Applications

Marcos Baez	University of Trento, Italy
Matthias Book	University of Iceland, Iceland
Dickson K.W. Chiu	The University of Hong Kong, SAR China
Olga De Troyer	Vrije Universiteit Brussel, Belgium
Federico Michele Facca	CREATE-NET, Italy
Angel Jimenez-Molina	University of Chile, Chile
Philipp Leitner	University of Zurich, Switzerland
Vicente Pelechano	Universitat Politècnica de València, Spain
Thomas Richter	Rhein-Waal University of Applied Sciences, Germany
Gustavo Rossi	UNLP, Argentina
Quan Z. Sheng	The University of Adelaide, Australia
Beat Signer	Vrije Universiteit Brussel, Belgium
Giovanni Toffetti Carughi	IBM Research Haifa, Israel

Social Web Applications

Ioannis Anagnostopoulos	University of Central Greece, Greece
Maria Bielikova	Slovak University of Technology in Bratislava, Slovakia
Marco Brambilla	Politecnico di Milano, Italy
Richard Chbeir	LIUPPA Laboratory, France
Alexandra Cristea	University of Warwick, UK
Gianluca Demartini	University of Sheffield, UK
Tommaso di Noia	Polytechnic University of Bari, Italy
Vania Dimitrova	University of Leeds, UK
Peter Dolog	Aalborg University, Denmark
Jutta Eckstein	IT communication, Germany
Filomena Ferruci	Università di Salerno, Italy
Geert-Jan Houben	Delft University of Technology, The Netherlands
Roman Klinger	Universität Stuttgart, Germany
Zakaria Maamar	Zayed University, UAE
Lourdes Moreno Lopez	Universidad Carlos III de Madrid, Spain
George Pallis	University of Cyprus, Cyprus
Jan Paralic	Technical University Kosice, Slovakia
Iv Ramakrishnan	Stony Brook, USA
Filippo Rica	University of Genoa, Italy
Thomas Richter	Rhein-Waal University of Applied Sciences, Germany
Elena Simperl	University of Southampton, UK
Philipp Singer	Knowledge Management Institute, Austria
William Van Woensel	University of Dalhousie, Canada
Vincent Wade	Dublin University, Ireland
Erik Wilde	EMC Corporation, USA
Guandong Xu	University of Technology Sydney, Australia
Gefei Zhang	Ludwig Maximilians University of Munich, Germany

Semantic Web Applications

Christian Bizer	University of Mannheim, Germany
Sarven Capadisli	University of Bonn, Germany
Pieter Colpaert	Ghent University, Belgium
Oscar Corcho	Universidad Politècnica de Madrid, Spain
Jose Manuel Gomez-Perez	Intelligent Software Components S.A., Spain
Thomas Gottron	University of Koblenz-Landau, Germany
Paul Groth	VU University, Amsterdam
Peter Haase	Fluid Operations, Germany
Eero Hyvönen	Aalto University and University of Helsinki, Finland
Tomi Kauppinen	Aalto University School of Science, Finland
Christoph Lange	University of Bonn, Germany
Robert Meusel	University of Mannheim, Germany
Daniel Oberle	SAP, Germany

Harald Sack	University of Potsdam, Germany
Axel Schulz	TU Darmstadt, Germany
Raphaël Troncy	EURECOM, France
Ruben Verborgh	Ghent University, Belgium
Maria Esther Vidal	Universidad Simon Bolivar, Venezuela
Evelyne Viegas	Microsoft Research, USA

Quality and Accessibility Aspects of Web Applications

Silvia Abrahao	Universitat Politècnica de València, Spain
Amaia Aizpurua	University of the Basque Country, Spain
Myriam Arrue	University of the Basque Country, Spain
Domenico Bianculli	University of Luxembourg, Luxembourg
Giorgio Brajnik	University of Udine, Italy
Andrew Brown	BBC, UK
Cinzia Cappiello	Politecnico di Milano, Italy
Luis Carrio	University of Lisbon, Portugal
Alex Chen	Nanyang Technological University, China
Nádia Fernandes	University of Lisbon, Portugal
Simon Harper	The University of Manchester, UK
Caroline Jay	The University of Manchester, UK
Maristella Matera	Politecnico di Milano, Italy
Emilia Mendes	The University of Auckland, New Zealand
Cu Duy Nguyen	University of Luxembourg, Luxembourg
Hironobu Takagi	IBM Research, Japan
Arie van Deursen	Delft University of Technology, The Netherlands
Marco Winckler	Université Paul Sabatier, France
Yeliz Yesilada	The University of Manchester, UK

Web Applications Composition and Mashups

Saeed Aghaee	University of Lugano, Switzerland
Boualem Benatallah	University of New South Wales, Australia
Christoph Bussler	Oracle Corporation, USA
Sven Casteleyn	University Jaume I of Castellon, Spain
Florian Daniel	University of Trento, Italy
Oscar Diaz	University of the Basque Country, Spain
Hao Han	Kanagawa University, Japan
Geert-Jan Houben	Delft University of Technology, The Netherlands
Agnes Koschmider	Karlsruhe Institute of Technology, Germany
Peep Küngas	University of Tartu, Estonia
Maristella Matera	Politecnico di Milano, Italy
Nikolay Mehandjiev	The University of Manchester, UK
Michael Weiss	Carleton University, Canada
Eric Wohlstadter	University of British Columbia, Canada

Web User Interfaces

Carlos Duarte	University of Lisbon, Portugal
Cristina Gena	University of Turin, Italy
William G.J. Halfond	University of Southern California, USA
Guy-Vincent Jourdan	University of Ottawa, Canada
Josip Maras	University of Split, Croatia
Maristella Matera	Politecnico di Milano, Italy
Carmen Santoro	ISTI-CNR, Italy
Nikolaos Tsantalis	Concordia University, Canada
Marco Winckler	Université Paul Sabatier, France

Security and Privacy in Web Applications

Adam Doupe	Arizona State University, USA
Schahram Dustdar	Vienna University of Technology, Austria
Martin Johns	SAP Research, Germany
Aniket Kate	Saarland University, Germany
Sergio Maffeis	Imperial College London, UK
Maristella Matera	Politecnico di Milano, Italy
Mohammadreza Mousavi	Halmstad University, Sweden
Phu H. Phung	University of Illinois at Chicago, USA
Raphael M. Reischuk	Saarland University, Germany
Nicola Zannone	Eindhoven University of Technology, The Netherlands

Industry

Marco Brambilla	Politecnico di Milano, Italy
Marina Egea	Indra, Spain
Maja Vukovic	IBM T.J. Watson Research Center, USA

Additional Reviewers

Panos Alexopoulos
Ritwik Banerjee
Seyed-Mehdi-Reza Beheshti
Juergen Cito
Ronald Denaux
Hariton Efstathiades
Birhanu Eshete
Bin Fu
Magnus Knuth
Christoph Pinkel
Dinesh Reddy
Petar Ristoski
Dominique Ritze
Roberto Rodriguez-Echeverria
Jessica Rosati
Paolo Tomeo
Joerg Waitelonis
Youquan Wang
Jie Yang

Local Organizing Committee

Local Arrangements Chairs

Ursula David	Erasmus University Rotterdam, The Netherlands
Marianne Kroek-Buijs	Erasmus University Rotterdam, The Netherlands

Registrations and Finance Chair

Susan Dijkstra	Congress Bureau, The Netherlands
Irene van der Wel	Congress Bureau, The Netherlands

Sponshorship Chairs

Viorel Milea	Erasmus University Rotterdam, The Netherlands
Damir Vandic	Erasmus University Rotterdam, The Netherlands

Publicity Chair

Guandong Xu	University of Technology Sydney, Australia

Proceedings Chair

Katja Temnow	Bielefeld University, Germany

Metadata Chair

Cord Wiljes	Bielefeld University, Germany

Website Chairs

Kim Schouten	Erasmus University Rotterdam, The Netherlands
Damir Vandic	Erasmus University Rotterdam, The Netherlands

ICWE Steering Committee Liaisons

Oscar Diaz	University of the Basque Country, Spain
Peter Dolog	Aalborg University, Denmark

Sponsoring Institutions

The conference chairs and conference organizers would like to thank our sponsors:

Keynotes

News Understanding for Knowledge Graph Freshness

Enrique Alfonseca

Google Inc.
ealfonseca@google.com

Abstract. This talk will describe ongoing work at Google Zurich towards event understanding and headline generation from news collections. From a web-scale corpus of English news, we mine and cluster syntactic patterns using sentence compression techniques, each cluster constituting an event description. At inference time, we can query the model with the patterns observed in an unseen news collection, identify the event that better captures the gist of the collection and automatically produce updates for the knowledge graph, and retrieve the most appropriate pattern to generate a headline.

Keywords: Information extraction · Event understanding · Sentence compression

Most Open Information Extraction (Open-IE) systems [2] extract textual relational patterns between entities automatically [3, 8] and optionally organize them into paraphrase clusters. These pattern clusters have been found to be useful for different applications, including Question Answering [4, 7] and relation extraction [5, 9].

A related Open-IE problem is that of automatically extracting and paraphrasing **event patterns:** those that describe changes in the state or attribute values of one or several entities. Applications include news understanding for e.g. generating news alerts, keeping structured knowledge bases up-to-date, or automatically generating headlines for news or news collections.

An existing approach lo learn paraphrases of event patterns is to build on the following weak supervision signal: news articles that were published on the same day and mention the same entities should contain good paraphrase candidates. A core component of these models is the automatic extraction of extraction patterns from the original sentences. Most previous works generate these patterns using heuristics, i.e., hand-written rules or lexico-syntactic regular expressions [1, 3, 8]. In our work we have explored the use of *Sentence compression* and *memory-based* extractors as viable alternatives, allowing us to trade-off coverage and accuracy for the extracted patterns.

Two state-of-the-art event paraphrasing systems that are based on this assumption are NewsSpike [11] and Heady/Idest [1, 6, 10]. This talk will focus on the latter, which has evolved from a Bayesian model (a Noisy-OR neural network) into a deep neural network model, where patterns are represented with a distributed representation in an embedding space, with the assumption that closeness in that space correlates with semantic closeness, with significant improvements in model accuracy.

References

1. Alfonseca, E., Pighin, D., Garrido, G.: HEADY: News headline abstraction through event pattern clustering. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, pp. 1243–1253, 4–9 August 2013
2. Banko, M., Cafarella, M.J., Soderland, S., Broadhead, M., Etzioni, O.: Open information extraction from the Web. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, pp. 2670–2676, 6–12 January 2007
3. Fader, A., Soderland, S., Etzioni, O.: Identifying relations for open information extraction. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, UK, pp. 1535–1545, 27–29 July 2011 (2011)
4. Fader, A., Zettlemoyer, L.S., Etzioni, O.: Paraphrase-driven learning for open question answering. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, pp. 1608–1618, 4–9 August 2013 (2013)
5. Grycner, A., Weikum, G.: Harpy: hypernyms and alignment of relational paraphrases. In: Proceedings of the 25rd International Conference on Computational Linguistics, Dublin, Ireland, pp. 2195–2204, 23–29 August 2014 (2014)
6. Krause, S., Alfonseca, E., Filippova, K., Pighin, D.: Idest: learning a distributed representation for event patterns. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics Human Language Technologies (2015)
7. Lin, D., Pantel, P.: Discovery of inference rules for question-answering. Nat. Lang. Eng. **7**(4), 343–360 (2001)
8. Mausam, Schmitz, M., Bart, R., Soderland, S., Etzioni, O.: Open language learning for information extraction. In: Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing, Jeju Island, Korea, pp. 523–534, 12–14 July 2012 (2012)
9. Moro, A., Navigli, R.: Wisenet: building a wikipedia-based semantic network with ontologized relations. In: Proceedings of the 21st ACM International Conference on Information and knowledge management (CIKM 2012), Maui, USA, pp. 1672–1676. ACM, 29 October–02 November 2012
10. Pighin, D., Colnolti, M., Alfonseca, E., Filippova, K.: Modelling events through memory-based, Open-IE patterns for abstractive summarization. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, USA, pp. 892–901, 22–27 June 2014
11. Zhang, C., Weld, D.S.: Harvesting parallel news streams to generate paraphrases of event relations. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, pp. 1776–1786, 18–21 October 2013

Making the Web Searchable

Peter Mika

Yahoo Labs
125 Shaftesbury Avenue
London, UK
pmika@yahoo-inc.com

Abstract. The key idea of the Semantic Web is to make information on the Web easily consumable by machines. As machines start to understand web pages as sources of data, search on the Web will move well beyond the current paradigm of retrieving pages by keywords. Instead, search engines will start to answer complex queries based on the cumulative knowledge of the Web. In this presentation, we will review the brief history of Semantic Search in academic research and in developments across the search industry. We also look ahead to highlight the research challenges that have surfaced and remain unsolved.

Improving access to the vast amount of human knowledge on the Web can be achieved in two principally different ways. Information Retrieval is traditionally concerned with improving the process of retrieval, given the content of the Web, as well as information on the structure of the Web and usage patterns of Web users [2]. Largely complementary to these efforts, the Semantic Web [1] has emerged as an effort to re-engineer the Web in a way that its contents become easier for machines to process. In this presentation, we will discuss the convergence of these approaches in what is known as Semantic Search, i.e. improving retrieval by exploiting the explicit semantics of information on the Web.

We begin by summarizing the relatively brief history of the scientific and engineering efforts towards realizing the Semantic Web, focusing in particular on the standards and tools for describing the content of web pages using additional metadata embedded in HTML. We will highlight the significant extent to which embedded metadata is already adopted on the Web, driven by an increasing number of applications but also facilitated by increasing collaboration around initiatives such as `schema.org`.

We then turn to discussing what these developments mean for search engine providers and other consumers of web content, starting from the very first application of embedded metadata, i.e. displaying more visual and interactive summaries of web pages [6]. We will explain in more detail the workings of entity retrieval and recommendation systems that power the “Knowledge Graph” experiences in large scale search engines such as Yahoo Search, and some of the research challenges associated, including the focused crawling of structured data [8], information integration [5, 7], ranking [3, 9] and query interpretation [4].

We will close by discussing some of the future directions in Semantic Search, in particular the challenge of going beyond entity-based, informational experiences to solving more complex tasks that require interaction with the user. We will also point

out the increasing shift in web search from desktop to mobile, and some of the opportunities and challenges of the mobile context.

References

1. Antoniou, G., Groth, P., van Harmelen, F., Hoekstra, R.: A Semantic Web Primer, 3rd edn. MIT Press (2012)
2. Baeza-Yates, R.A., Ribeiro-Neto, B.A.: Modern Information Retrieval, 2nd edn. ACM Press / Addison-Wesley (2011)
3. Blanco, R., Mika, P., Vigna, S.: Effective and efficient entity search in RDF data. In: Proceedings of 10th International Semantic Web Conference, Bonn, Germany, pp. 83–97 (2011)
4. Blanco, R., Ottaviano, G., Meij, E.: Fast and space-efficient entity linking for queries. In: Proceedings of the Eighth ACM WSDM Conference, Shanghai, China, pp. 179–188 (2015)
5. Dalton, J., Blanco, R., Mika, P.: Coreference aware web object retrieval. In: Proceedings of the 20th ACM CIKM Conference, Glasgow, United Kingdom, pp. 211–220 (2011)
6. Haas, K., Mika, P., Tarjan, P., Blanco, R.: Enhanced results for web search. In: Proceeding of the 34th International ACM SIGIR Conference, Beijing, China, pp. 725–734 (2011)
7. Herzig, D.M., Mika, P., Blanco, R., Tran, T.: Federated entity search using on-the-fly consolidation. In: Proceedings of the 12th International Semantic Web Conference, Sydney, NSW, Australia, pp. 167–183 (2013)
8. Meusel, R., Mika, P., Blanco, R.: Focused crawling for structured data. In: Proceedings of the 23rd ACM CIKM Conference, Shanghai, China, pp. 1039–1048 (2014)
9. Pound, J., Mika, P., Zaragoza, H.: Ad-hoc object retrieval in the web of data. In: Proceedings of the 19th WWW Conference, Raleigh, North Carolina, USA, pp. 771–780 (2010)

Sentiment Strength Detection for Social Media Text: Artificial Agents, Answer Ranking and Art Installations

Mike Thelwall

Statistical Cybermetrics Research Group,
University of Wolverhampton,
Wolverhampton, UK
M.Thelwall@wlv.ac.uk

Abstract. This talk will describe a simple, fast, intuitive and flexible lexical method to detect sentiment strength in short informal text and will illustrate it with a range of research and commercial applications. Implemented in the software SentiStrength and optimised for social web text, the method can process 14,000 tweets per second with human level accuracy in many cases. The talk will demonstrate the program and its evaluations and show how it has been translated from the original English version to many other languages. It is free for researchers to use and its current applications include art installations, answer ranking, opinion mining, stock market prediction, customer feedback analysis, and sentiment-aware artificial agents.

Keywords: Sentiment analysis · Opinion mining · Social web · Twitter

The sentiment strength detection program SentiStrength [4] was developed in response to a need for the sentiment analysis of the huge number of informal texts written every day in social network sites, Twitter and elsewhere on the social web. Emotion psychology suggests that the most fundamental affective distinction is between positive and negative sentiment, with finer grained expressions of emotion being more culturally dependant. Moreover, humans process positive and negative sentiment in parallel to some extent. SentiStrength was therefore designed to detect the strength of positive and negative sentiment separately in social web texts.

SentiStrength uses a primarily lexical approach to detect sentiment. It has a list of about 2700 word stems, each with a pre-defined positive or negative sentiment strength score. For example, if a text contains the word *excellent* then the list suggests the presence of strong (4 out of 5) positive sentiment. There are 18 additional rules for special cases, such as negation, questions, booster words (e.g., *very*), emoticons and sentiment spelling (e.g., *niiiiice*). This approach is fast (14,000 tweets/sec on a basic PC) and gives human-level accuracy on most of the social web texts that it has been tested on [6]. Its scores also tend to reflect the stated judgments and bodily responses (e.g., sweating, heartrate) of the people reading or writing it [1]. Because all sentiment resources, such as the sentiment term list, are stored in plain text files separately from the code, SentiStrength can be customised for specialised versions and other languages without the need for new code, except for agglutinative languages [8] and languages without word boundaries (e.g., Chinese).

In addition to being used in social science research to detect patterns of sentiment in the social web, SentiStrength has been used in a wide variety of other applications. The most prominent use is for detecting sentiment in the feedback given to question answerers in Yahoo Answers. The results boost the ranking of the answers given by people who receive more positive feedback [2]. SentiStrength also supports the logic of autonomous agents responding to human text. For example, users in one virtual world reported deeper engagement with a virtual bartender when his facial expression and body posture reacted to the sentiment in their conversation [3]. SentiStrength's ease of customisation has also allowed it to be used in applications when sentiment is commonly expressed in unusual ways. For example, it was used to drive lightshows on the London Eye during the London 2012 Olympics and on the Empire State Building during the Super Bowl 2014 by detecting sentiment in relevant tweets and translating the results into patterns of light. In both cases, simple changes were needed to detect context-specific expressions, such as *Go Hawks!*. SentiStrength has also been used to analyse tweets about prominent political and news events. It incorporates automatic methods to optimise its term list and term weightings for such topic specific applications [5] but struggles with sarcasm in online political arguments [6], and with using context from ongoing discussions to improve sentiment scores [7].

References

1. Kappas, A. Tsankova, E. Theunis, M., Kuester, D.: Cyberemotions: subjective and physiological responses elicited by contributing to online discussion forums. In: 51th Annual Meeting of the Society for Psychophysiological Research (2010)
2. Kucuktunc, O., Cambazoglu, B.B., Weber, I., Ferhatosmanoglu, H.: A large-scale sentiment analysis for Yahoo! Answers. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining (2012)
3. Skowron, M., Pirker, H., Rank, S., Paltoglou, G., Ahn, J., Gobron S.: No peanuts! Affective cues for the virtual bartender. In: Proceedings of the Florida Artificial Intelligence Research Society Conference. AAAI Press (2011)
4. Thelwall, M., Buckley, K., Paltoglou, G. Cai, D., Kappas, A.: Sentiment strength detection in short informal text. J. Am. Soc. Inf. Sci. Technol. **61**(12), 2544–2558 (2010)
5. Thelwall, M., et al.: Damping sentiment analysis in online communication: Discussions, monologs and dialogs. In: Gelbukh, A. (ed.) CICLing 2013, Part II. LNCS, vol. 7817, pp. 1–12. Springer, Heidelberg (2013)
6. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social web. J. Am. Soci. Inf. Sci. Technol. **63**(1), 163–173 (2012)
7. Thelwall, M., Buckley, K.: Topic-based sentiment analysis for the Social Web: the role of mood and issue-related words. J. Am. Soc. Inf. Sci. Technol. **64**(8), 1608–1617 (2013)
8. Vural, G., Cambazoglu, B.B., Senkul, P., Tokgoz O.: A framework for sentiment analysis in Turkish: application to polarity detection of movie reviews in Turkish. Comput. Inf. Sci. III, 437–445 (2013)

Contents

Social Web Applications

Semantic Web Applications

Web User Interfaces

Industry

PhD Symposium

Demonstrations