IJCNLP 2011

The Fifth International Joint Conference on Natural Language Processing

November 8-13, 2011
Shangri-La Hotel
Chiang Mai, Thailand

AFNLP

IJCNLP 2011

Proceedings of the Fifth International Joint Conference on Natural Language Processing

November 8 – 13, 2011
Chiang Mai, Thailand

We wish to thank our sponsors

Gold Sponsors

www.google.com

www.baidu.com

The Office of Naval Research (ONR)

The Asian Office of Aerospace Research and Development (AOARD)

Department of Systems Engineering and Engineering Managment, The Chinese University of Hong Kong

Silver Sponsors

Microsoft Corporation

Bronze Sponsors

Chinese and Oriental Languages Information Processing Society (COLIPS)

Supporter

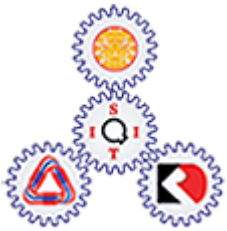
Thailand Convention and Exhibition Bureau (TCEB)

We wish to thank our sponsors

Organizers

Asian Federation of Natural Language Processing (AFNLP)

National Electronics and Computer Technology Center (NECTEC), Thailand



Sirindhorn International Institute of Technology (SIIT), Thailand

Rajamangala University of Technology Lanna (RMUTL), Thailand

Maejo University, Thailand

Chiang Mai University (CMU), Thailand

©2011 Asian Federation of Natural Language Proceesing

ISBN 978-974-466-564-5

FOREWORD

IJCNLP2011, where Anna(s) meet the king(s) for sharing knowledge in natural language

IJCNLP2011 is held in Chiang Mai. It is a historic city situated in the northern part of Thailand. Organizing the conference in this part of Asia made us think of the classic movie "The King and I" (1956), where King Mongkut of Siam invited Anna Leonowens an Anglo-Indian school teacher to Siam to teach his family English. Similar to the movie, IJCNLP2011 brings together scientists and practitioners from the East and West in pursuit of the knowledge of natural language processing (NLP).

Virach, Hitoshi and I compiled this passage collaboratively online using our own iPads. Despite us being physically apart, in Thailand, Japan and Hong Kong respectively, our collaborative editorial work went smoothly with virtually no distance. The increasing popularity of smart handheld devices, such as iPhones and iPads has practically made the world flat. The hurdles and boundaries between people have effectively been lifted enabling friends and relatives over the globe to keep in close contact with each other. We use email, blog, facebook and twitter regularly and ubiquitously for communications. Non-traditional they may be, the languages for communication over these channels are natural as they are used by the netizens (human) for information exchange. Processing of these natural languages is inevitably unconventional and the task is challenging, which requires much innovation. For this reason, NLP is a key research area both in the industry and in universities worldwide. Therefore, it is not surprising that we have received over 500 submissions from different countries around the world in this year's IJCNLP. This number is in fact the largest in the history of the conference series.

Organizing a conference of the scale of IJCNLP2011 (with over 300 participants) is never easy. We worked closely as a team in the past ten months. It is really not easy for us to express our gratitude to any one individual. The names of the hard working conference officers, the track chairs, the workshop chairs, the tutors as well as the reviewers are enlisted in the proceedings. We owe everyone a billion. Without their hard work IJCNLP2011 would never have reached this stage. So please help me praise and thank them when you meet them in the conference.

Chiang Mai is a cultural city full of history and traditions, with many famous attractions such as its melodious colloquial language, Lanna style of clothing, mellow taste of food, etc. During the conference period, we will experience the "Loi Krathong Festival" where people float krathong (floating basket) on a river to pay respect to the spirit of the waters. IJCNLP2011 in November Chiang Mai is unique. It coincides with the unforgettable Lanna Festival. Locally known as "Yi Peng", the festival will bring to you a memorable cultural experience. You will witness a multitude of Lanna-style sky lanterns (khom loi, literally "floating lanterns") gently rising in the air. These lanterns resemble large flocks of giant fluorescent jellyfish gracefully floating by through the sky. Honestly, these attractions are just too good to be missed.

Dear friends and colleagues of the world NLP communities, honorable guests of Chiang Mai, we are glad to see you in IJCNLP2011. We hope you find the technical program useful to your research and can discover something insightful at the end. And before closing, as one often said "seeing is believing", we urge you to spare some time after the conference to explore and to enjoy the city.

Ka Poon Kap (thank you)

Kam-Fai Wong, General Chair, The Chinese University of Hong Kong (CUHK), China

Virach Sornlertlamvanich, Organization Co-Chair, National Electronics and Computer Technology Center (NECTEC), Thailand

Hitoshi Isahara, Organization Co-Chair, Toyohashi University of Technology, Japan

November 7, 2011

PREFACE

As the flagship conference of the Asian Federation of Natural Language Processing (AFNLP), IJCNLP has now rapidly grown into a renowned international event. IJCNLP 2011 covers a broad spectrum of technical areas related to natural language processing. The conference includes full papers, short papers, demonstrations, a student research workshop, as well as pre- and post-conference tutorials and workshops.

This year, we received a record 478 valid paper submissions, which is well beyond our initial expectations. This represents an increasing interest of research on NLP and the growing reputation of IJCNLP as an international event. The 478 submissions include 385 full-paper submissions and 93 short-paper submissions from more than 40 countries. Specifically, approximately 61% of the papers are from 16 countries and areas in Asia Pacific, 22% from 16 countries in Europe, 14% from the United States and Canada; we also have 2% of the papers from the Middle East and Africa, and 1% from South America.

We would like to thank all the authors for submitting papers to IJCNLP 2011. The significant increase in the number of submissions and the wide range of demographic areas represent a rapid growth of our field. We would also like to thank the 22 area chairs and 474 program committee members for writing over 1400 reviews and meta-reviews and for paving the way for the final paper selection. Of all 478 submissions, a total of 176 papers were accepted, representing a healthy 36% acceptance rate. The accepted papers are comprised of 149 full papers (8+ pages), of which 107 are presented orally and 42 as posters, and 27 short papers (4+ pages) where 25 are presented orally and 2 as posters. We are extremely grateful to the area chairs and program committee members for all their hard work, without which the preparation of this program would not be possible.

We are delighted to have invited three strategic keynote speakers addressing different application aspects of NLP for the Web in IJCNLP2011. Mathew Lease will talk about "crowdsourcing", which is a trendy and effective means to perform a task that requires hundreds/thousands of people, such as corpus tagging. Wai Lam will present the latest techniques for information extraction, which is essential for today's Internet business. And last but not the least, Mengqiu Wang, Vice President of Baidu, the largest Internet search company in China, will share with us the recent trends in search and social network technologies and how NLP techniques can be applied to improve performance in the real world. These speeches will surely be informative and enlightening to the audience leading to many innovative research ideas. We are excited about it and are looking forward to them. Best paper awards will be announced in the last session of the conference as well.

We thank General Chair Kam-Fai Wong, the Local Arrangements Committee headed by Virach Sornlertlamvanich and Hitoshi Isahara, and the AFNLP Conference Coordination Committee chaired by Yuji Matsumoto, for their help and advice. Thanks to Min Zhang and Sudeshria Sarkar, the Publication Co-Chairs for putting the proceedings together, and all the other committee chairs for their work.

We hope that you enjoy the conference!

Haifeng Wang, Baidu

David Yarowsky, Johns Hopkins University

November 7, 2011

Honorary Conference Chair

Chaiyong Eurviriyanukul, Rajamangala University of Technology Lanna, Thailand
Chongrak Polprasert, Sirindhorn International Institute of Technology, Thailand
Thaweesak Koanantakool, NSTDA, Thailand

General Chair

Kam-Fai Wong, The Chinese University of Hong Kong, China

Program Co-Chairs:

Haifeng Wang, Baidu, China
David Yarowsky, John Hopkins University, USA

Organisation Co-Chairs:

Virach Sornlertlamvanich, NECTEC, Thailand
Hitoshi Isahara, Toyohashi University of Technology, Japan

Workshop Co-Chairs:

Sivaji Bandyopadhyay, Jadavpur University, India
Jong Park, KAIST, Korea
Noriko Kando, NII, Japan

Tutorial Co-Chairs:

Kentaro Inui, Tohoku University, Japan
Wei Gao, The Chinese University of Hong Kong, China
Dawei Song, Robert Gordon University, UK

Demonstration Co-Chairs:

Ken Church, Johns Hopkins University, USA
Yunqing Xia, Tsinghua University, China

Publication Co-Chairs:

Min Zhang, I2R, Singapore
Sudeshna Sarkar, IIT Kharagpur, India

Finance Co-Chairs:

Vilas Wuwongse, AIT, Thailand
Gary Lee, POSTECH, Korea

Sponsorship Co-Chairs:

Asanee Kawtrakul, Kasetsart University, Thailand
Methinee Sirikrai, NECTEC, Thailand
Hiromi Nakaiwa, NTT, Japan

Publicity Committee:

Steven Bird, University of Melbourne, Australia
Le Sun, CIPS, China
Kevin Knight, USC, USA
Nicoletta Calzolari, Istituto di Linguistica Computazionale del CNR, Italy
Thanaruk Theeramunkong, SIIT, Thailand

Webmasters:

Swit Phuvipadawat, Tokyo Institute of Technology, Japan
Wirat Chinnan, SIIT, Thailand

Area Chairs:

Discourse, Dialogue and Pragmatics
David Schlangen, The University of Potsdam, Germany

Generation /Summarization
Xiaojun Wan, Peking University, China

Information Extraction
Wenjie Li, The Hong Kong Polytechnic University, Hong Kong

Information Retrieval
Gareth Jones, Dublin City University, Ireland

Language Resource
Eneko Agirre, University of the Basque Country, Spain

Machine Translation
David Chiang, USC-ISI, USA
Min Zhang, Institute for Infocomm Research, Singapore
Hua Wu, Baidu, China

Phonology/morphology, POS tagging and chunking, Word Segmentation
Richard Sproat, Oregon Health & Science University, USA
Gary Lee, Pohang University of Science and Technology, Korea

Question Answering
Jun Zhao, Institute of Automation, Chinese Academy of Sciences, China

Semantics
Pushpak Bhattacharyya, Indian Institute of Technology, India
Hinrich Schuetze, University of Stuttgart, Germany

Sentiment Analysis, Opinion Mining and Text Classification
Rafael Banchs, Institute for Infocomm Research, Singapore
Theresa Wilson, Johns Hopkins University, USA

Spoken Language Processing
Chung-Hsien Wu, National Cheng Kung University, Taiwan

Statistical and ML Methods
Miles Osborne, The University of Edinburgh, UK
David Smith, University of Massachusetts Amherst, USA

Syntax and Parsing
Stephen Clark, University of Cambridge, UK
Yusuke Miyao, National Institute of Informatics, Japan

Text Mining and NLP Applications
Juanzi Li, Tsinghua University, China
Patrick Pantel, Microsoft Research, USA

Reviewers

Ahmed Abbasi, Omri Abend, Akiko Aizawa, Ahmet Aker, Enrique Alfonseca, DAUD ALI, Ben Allison, Robin Aly, Alina Andreevskaia, Masayuki Asahara, Ai Azuma

Jing Bai, Alexandra Balahur, Timothy Baldwin, Kalika Bali, Carmen Banea, Srinivas Bangalore, Mohit Bansal, Marco Bbaroni, Roberto Basili, Timo Baumann, Emily Bender, Shane Bergsma, Pushpak Bhattacharyya, Dan Bikel, Wang Bin, Lexi Birch, Michael Bloodgood, Phil Blunsom, Nate Bodenstab, Ester Boldrini, Gemma Boleda, Danushka Bollegala, Luc Boruta, Stefan Bott, Chris Brew, Sam Brody, Julian Brooke, Paul Buitelaar, Miriam Butt

Aoife Cahill, Li Cai, Yi Cai, Nicoletta Calzolari, Jaime Carbonell, Marine Carpuat, John Carroll, Paula Carvalho, Suleyman Cetintas, Debasri Chakrabarti, Nate Chambers, Niladri Chatterjee, Wanxiang Che, Berlin Chen, Boxing Chen, Chia-Ping Chen, Hsin-Hsi Chen, Wenliang Chen, Ying Chen, Yufeng Chen, Pu-Jen Cheng, Colin Cherry, Jackie Chi KiCheung, Key-Sun Choi, Monojit Choudhury, Christos Christodoulopoulos, Kenneth Church, Alex Clark, Shay Cohen, Trevor Cohn, Gao Cong, Marta R. Costa-jussa, Paul Crook, Montse Cuadros, Ronan Cummins

Robert Damper, Kareem Darwish, Dipanjan Das, Niladri Dash, Adrià de Gispert, Daniel de Kok, Eric De La Clergerie, Stijn De Saeger, Steve DeNeefe, Pascal Denis, Ann Devitt, Arantza Diaz de Ilarraza, Anne Diekema, Markus Dreyer, Rebecca Dridan, Jinhua Du, Xiangyu Duan, Amit Dubey, Kevin Duh, Chris Dyer, Michal Dziemianko

Jacob Eisenstein, Michael Elhadad, Micha Elsner, Martin Emms

Angela Fahrni, Hui Fang, Yi Fang, Li Fangtao, Christiane Fellbaum, Raquel Fernandez, Colum Foley, Jennifer Foster, Timothy Fowler, Stella Frank, Guohong Fu, Atsushi Fujii, Kotaro Funakoshi, Hagen Fürstenau

Matthias Galle, Michael Gamon, Michaela Geierhos, Eugenie Giesbrecht, Alastair Gill, Roxana Girju, Bruno Golenia, Carlos Gomez-Rodriguez, Zhengxian Gong, Matt Gormley, Amit Goyal, João Graça, Jens Grivolla, Iryna Gurevych

Stephanie Haas, Barry Haddow, Eva Hajicova, David Hall, Keith Hall, Xianpei Han, Kazuo Hara, Donna Harman, Kazi Hasan, Chikara Hashimoto, Koiti Hasida, Eva Hasler, Samer Hassan, Claudia Hauff, Xiaodong He, Yulan He, Zhongjun He, Carlos Henriquez, Tsutomu Hirao, Hieu Hoang, Tracy Holloway King, Matthew Honnibal, Mark Hopkins, Meishan Hu, Chien-Lin Huang, Fei Huang, Minlie Huang, Ruizhang Huang, Xiaojiang Huang, Xuanjing Huang, Yun Huang, Zhongqiang Huang

Francisco Iacobelli, Diana Inkpen, Aminul Islam, Ruben Izquierdo

Heng Ji, Sittichai Jiampojamarn, Hongfei Jiang, Wenbin Jiang, Xing Jiang, Cai Jie, Rong Jin, Richard Johansson, Hanmin Jung

Sarvnaz Karimi, Daisuke Kawahara, Jun'ichi Kazama, Liadh Kelly, Maxim Khalilov, Mitesh Khapra, Adam Kilgarriff, Byeongchang Kim, Irwin King, Alistair Knott, Philipp Koehn, Rob Koeling, Oskar Kohonen, Mamoru Komachi, Grzegorz Kondrak, Fang Kong, Valia Kordoni, Lili Kotlerman, Zornitsa Kozareva, Wessel Kraaij, Parton Kristen, Lun-Wei Ku, Sudip Kumar Naskar, June-Jei Kuo, Kow Kuroda, Sadao KUROHASH, Kui-Lam Kwok, Han Kyoung-Soo

Sobha Lalitha Devi, Wai Lam, Joel Lang, Jun Lang, Matt Lease, Cheongjae Lee, Jung-Tae Lee, Sungjin Lee, Tan Lee, Russell Lee-goldman, Alessandro Lenci, Johannes Leveling, Abby Levenberg, Gina-Anne Levow, Baoli Li, Daifeng Li, Haizhou Li, linlin li, Mu Li, Qing Li, Shoushan Li, Sujian Li, Yunyao Li, Shasha Liao, Yuan-Fu Liao, Chin-Yew Lin, Pierre Lison, Ken Litkowski, Marina Litvak, Bing Liu, Fei Liu, Feifan Liu, Kang Liu, Pengyuan Liu, Qun Liu, Shui Liu, Xiaohua Liu, Yang Liu (UT Dallas), Yang Liu (ICT CAS), Yi Liu, Ying Liu, Yiqun Liu, Zhanyi Liu, Hector Llorens, Elena Lloret, Wai-Kit Lo, QIU Long, Adam Lopez, Yajuan Lu

Bin Ma, Yanjun Ma, Walid Magdy, OKUMURA Manabu, Suresh Manandhar, Maria Antonia Marti, David Martinez, Andre Martins, Yuji Matsumoto, Yutaka Matsuo, Takuya Matsuzaki, Mike Maxwell, Jonathan May, Diana McCarthy, David McClosky, Ryan McDonald, Paul McNamee, Beata Megyesi, Donald Metzler, Haitao Mi, Lukas Michelbacher, Dipti Mishra Sharma, Mandar Mitra, Daichi Mochihashi, Saif Mohammed, Behrang Mohit, Karo Moilanen, Christian Monson, Paul Morarescu, Jin'ichi Murakami, Sung Hyon Myaeng

Seung-Hoon Na, Masaaki Nagata, Mikio Nakano, Preslav Nakov, Jason Naradowsky, Vivi Nastase, Roberto Navigli, Mark-Jan Nederhof, Ani Nenkova, Vincent Ng, Truc-Vien T. Nguyen, Eric Nichols, Tadashi Nomoto, Scott Nowson, Andreas Nuernberger, Pierre Nugues

Diarmuid O Seaghdha, Brendan O'Connor, Neil O'Hare, Stephan Oepen, Kemal Oflazer, Kemal Oflazer, Alice Oh, Naoaki Okazaki, Constantin Orasan, Arantxa Otegi, Myle Ott, Jahna Otterbacher, You Ouyang

Alexandre Patry, Soma Paul, Adam Pease, Ted Peders, Wei Peng, Gerald Penn, Sasa Petrovic, Christian Pietsch, Juan Pino, Matt Post, John Prager, Daniel Preotiuc, Matthew Purver

Vahed Qazvinian, Guang Qiu, Chris Quirk

Altaf Rahman, Ganesh Ramakrishnan, Karthik Raman, AnanthakrishnRamanathan, Sujith Ravi, Bunescu Razvan, Jonathon Read, Marta Recasens, Jeremy Reffin, Roi Reichart, Jason Riesa, Verena Rieser, Arndt Riester, Stefan Riezler, German Rigau, Laura Rimell, Carlos Rodriguez, Kepa Rodriguez, Robert Ross, Michael Roth, Sasha Rush

Kenji Sagae, Benoît Sagot, Agnes Sandor, Anoop Sarkar, Sudeshna Sarkar, Ryohei Sasano, Roser Sauri, Helmut Schmid, Satoshi Sekine, Arulmozi Selvaraj, Pavel Serdyukov, Gao Sheng, Masashi Shimbo, Darla Shockley, Luo Si, Khalil Sima'an, Ben Snyder, Ruihua Song, Young-In Song, Sebastian Spiegler, Valentin Spitkovsky, Caroline Sporleder, Manfred Stede, Mark Steedman, Mark Stevenson, Nicola Stokes, Veselin Stoyanov, Michael Strube, Jian Su, Keh-Yih Su, Zhifang Sui, Aixin Sun, Jun Sun, Weiwei Sun, Mihai Surdeanu

Oscar Tackstrom, Hiroya Takamura, Jianhua Tao, Joel Tetreault, Stefan Thater, Jörg Tiedemann, Ivan Titov, Takenobu Tokunaga, Kentaro Torisawa, Lamia Tounsi, Kristina Toutanova, Roy Tromble, Reut Tsarfaty, Yuen-Hsien Tseng, Hajime Tsukada

Christina Unger, Takehito Utsuro

Antal van den Bosch, Gertjan van Noord, Vasudeva Varma, Silvia Vazquez, Tony Veale, Olga Vechtomova, Sriram Venkatapathy, Yannick Versley, Jesus Vilares, Sami Virpioja, Andreas Vlachos, Piek Vossen

Stephen Wan, Bin Wang, Bo Wang, Dingding Wang, Hsin-Min Wang, Ting Wang, Wei Wang, Zhichun Wang, Taro Watanabe, Yotaro Watanabe, Bonnie Webber, Furu Wei, Richard Wicentowski, Shuly Wintner, Kristian Woodsend, Gang Wu, Zhiyong Wu

Yunqing Xia, Tong Xiao, Xin Xin, Deyi Xiong, Qiu Xipeng, Jun Xu, Ruifeng Xu

Christopher Yang, Grace Yang, Muyun Yang, Yuhang Yang, Zi Yang, Benajiba Yassine, Mark Yatskar, Patrick Ye, Jui-Feng Yeh, Ainur Yessenalina, Scott Wen-tauYih, Bei Yu, Hong Yu

Taras Zagibalov, Benat Zapirain, Alessandra Zarcone, Duo Zhang, Hao Zhang, Jiajun Zhang, Jing Zhang, Lanbo Zhang, Lei Zhang, Min Zhang, Qi Zhang, Yi Zhang (UCSC), Yi Zhang (DFKI), Yue

Zhang, Shiqi Zhao, Tiejun Zhao, Haitao Zheng, Zhi Zhong, Bowen Zhou, Dong Zhou, GuoDong Zhou, Qiang Zhou, Yu Zhou, Muhua Zhu, Xiaodan Zhu, Chengqing Zong

Table of Contents

Part A: Full Papers

Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers
Sonal Gupta and Christopher Manning 1

Dependency-directed Tree Kernel-based Protein-Protein Interaction Extraction from Biomedical Literature
Longhua Qian and Guodong Zhou 10

Learning Logical Structures of Paragraphs in Legal Articles
Ngo Xuan Bach, Nguyen Le Minh, Tran Thi Oanh and Akira Shimazu 20

Extracting Pre-ordering Rules from Predicate-Argument Structures
Xianchao Wu, Katsuhito Sudoh, Kevin Duh, Hajime Tsukada and Masaaki Nagata 29

Context-Sensitive Syntactic Source-Reordering by Statistical Transduction
Maxim Khalilov and Khalil Sima'an 38

Discriminative Phrase-based Lexicalized Reordering Models using Weighted Reordering Graphs
Wang Ling, João Graça, David Martins de Matos, Isabel Trancoso and Alan W Black 47

Active Learning Strategies for Support Vector Machines, Application to Temporal Relation Classification
Seyed Abolghasem Mirroshandel, Gholamreza Ghassem-Sani and Alexis Nasr 56

A Fast Accurate Two-stage Training Algorithm for L1-regularized CRFs with Heuristic Line Search Strategy
Jinlong Zhou, Xipeng Qiu and Xuanjing Huang 65

Automatic Topic Model Adaptation for Sentiment Analysis in Structured Domains
Geoffrey Levine and Gerald DeJong 75

Multi-modal Reference Resolution in Situated Dialogue by Integrating Linguistic and Extra-Linguistic Clues
Ryu Iida, Masaaki Yasuhara and Takenobu Tokunaga 84

Single and multi-objective optimization for feature selection in anaphora resolution
Sriparna Saha, Asif Ekbal, Olga Uryupina and Massimo Poesio 93

A Unified Event Coreference Resolution by Integrating Multiple Resolvers
Bin Chen, Jian Su, Sinno Jialin Pan and Chew Lim Tan 102

Handling verb phrase morphology in highly inflected Indian languages for Machine Translation
Ankur Gandhe, Rashmi Gangadharaiah, Karthik Visweswariah and Ananthakrishnan Ramanathan 111

Japanese Pronunciation Prediction as Phrasal Statistical Machine Translation
Jun Hatori and Hisami Suzuki 120

Comparing Two Techniques for Learning Transliteration Models Using a Parallel Corpus
Hassan Sajjad, Nadir Durrani, Helmut Schmid and Alexander Fraser 129

A Semantic-Specific Model for Chinese Named Entity Translation
Yufeng Chen and Chengqing Zong 138

Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners
Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto 147

Modality Specific Meta Features for Authorship Attribution in Web Forum Posts
Thamar Solorio, Sangita Pillay, Sindhu Raghavan and Manuel Montes-Gomez 156

Keyphrase Extraction from Online News Using Binary Integer Programming
Zhuoye Ding, Qi Zhang and Xuanjing Huang 165

Improving Related Entity Finding via Incorporating Homepages and Recognizing Fine-grained Entities
Youzheng Wu, Chiori Hori, Hisashi Kawai and Hideki Kashioka 174

Enhancing Active Learning for Semantic Role Labeling via Compressed Dependency Trees
Chenhua Chen, Alexis Palmer and Caroline Sporleder 183

Semantic Role Labeling Without Treebanks?
Stephen Boxwell, Chris Brew, Jason Baldridge, Dennis Mehay and Sujith Ravi 192

Japanese Predicate Argument Structure Analysis Exploiting Argument Position and Type
Yuta Hayashibe, Mamoru Komachi and Yuji Matsumoto 201

An Empirical Study on Compositionality in Compound Nouns
Siva Reddy, Diana McCarthy and Suresh Manandhar 210

Feature-Rich Log-Linear Lexical Model for Latent Variable PCFG Grammars
Zhongqiang Huang and Mary Harper 219

Improving Dependency Parsing with Fined-Grained Features
Guangyou Zhou, Li Cai, Kang Liu and Jun Zhao 228

Natural Language Programming Using Class Sequential Rules
Cohan Sujay Carlos 237

Treeblazing: Using External Treebanks to Filter Parse Forests for Parse Selection and Treebanking
Andrew MacKinlay, Rebecca Dridan, Dan Flickinger, Stephan Oepen and Timothy Baldwin . . 246

Cross-Language Entity Linking
Paul McNamee, James Mayfield, Dawn Lawrie, Douglas Oard and David Doermann 255

Generating Chinese Named Entity Data from a Parallel Corpus
Ruiji Fu, Bing Qin and Ting Liu 264

Learning the Latent Topics for Question Retrieval in Community QA
Li Cai, Guangyou Zhou, Kang Liu and Jun Zhao 273

Identifying Event Descriptions using Co-training with Online News Summaries
William Yang Wang, Kapil Thadani and Kathleen McKeown 282

Automatic Labeling of Voiced Consonants for Morphological Analysis of Modern Japanese Literature
Teruaki Oka, Mamoru Komachi, Toshinobu Ogiso and Yuji Matsumoto 292

S^3 - Statistical Sandhi Splitting
Abhiram Natarajan and Eugene Charniak 301

Improving Chinese Word Segmentation and POS Tagging with Semi-supervised Methods Using Large Auto-Analyzed Data
Yiou Wang, Jun'ichi Kazama, Yoshimasa Tsuruoka, Wenliang Chen, Yujie Zhang and Kentaro Torisawa 309

CODACT: Towards Identifying Orthographic Variants in Dialectal Arabic
Pradeep Dasigi and Mona Diab 318

Enhancing the HL-SOT Approach to Sentiment Analysis via a Localized Feature Selection Framework
Wei and Jon Atle Gulla 327

Fine-Grained Sentiment Analysis with Structural Features
Cäcilia Zirn, Mathias Niepert, Heiner Stuckenschmidt and Michael Strube 336

Predicting Opinion Dependency Relations for Opinion Analysis
Lun-Wei Ku, Ting-Hao Huang and Hsin-Hsi Chen 345

Detecting and Blocking False Sentiment Propagation
Hye-Jin Min and Jong C. Park 354

Efficient induction of probabilistic word classes with LDA
Grzegorz Chrupala 363

Quality-biased Ranking of Short Texts in Microblogging Services
Minlie Huang, Yi Yang and Xiaoyan Zhu 373

Labeling Unlabeled Data using Cross-Language Guided Clustering
Sachindra Joshi, Danish Contractor and Sumit Negi 383

Extracting Relation Descriptors with Conditional Random Fields
Yaliang Li, Jing Jiang, Hai Leong Chieu and Kian Ming A. Chai 392

Attribute Extraction from Synthetic Web Search Queries
Marius Pasca 401

Japanese Abbreviation Expansion with Query and Clickthrough Logs
Kei Uchiumi, Mamoru Komachi, Keigo Machinaga, Toshiyuki Maezawa, Toshinori Satou and Yoshinori Kobayashi 410

Mining Parallel Documents Using Low Bandwidth and High Precision CLIR from the Heterogeneous Web
Simon Shi, Pascale Fung, Emmanuel Prochasson, Chi-kiu Lo and Dekai Wu 420

Crawling Back and Forth: Using Back and Out Links to Locate Bilingual Sites
Luciano Barbosa, Srinivas Bangalore and Vivek Kumar Rangarajan Sridhar 429

Grammar Induction from Text Using Small Syntactic Prototypes
Prachya Boonkwan and Mark Steedman 438

Transferring Syntactic Relations from English to Hindi Using Alignments on Local Word Groups
Aswarth Dara, Prashanth Mannem, Hemanth Sagar Bayyarapu and Avinesh PVS 447

Generative Modeling of Coordination by Factoring Parallelism and Selectional Preferences
Daisuke Kawahara and Sadao Kurohashi 456

Syntactic Parsing for Ranking-Based Coreference Resolution
Altaf Rahman and Vincent Ng 465

TriS: A Statistical Sentence Simplifier with Log-linear Models and Margin-based Discriminative Training
Nguyen Bach, Qin Gao, Stephan Vogel and Alex Waibel 474

Social Summarization via Automatically Discovered Social Context
Po Hu, Cheng Sun, Longfei Wu, Donghong Ji and Chong Teng 483

Simultaneous Clustering and Noise Detection for Theme-based Summarization
Xiaoyan Cai, Renxian Zhang, Dehong Gao and Wenjie Li 491

Extractive Summarization Method for Contact Center Dialogues based on Call Logs
Akihiro Tamura, Kai Ishikawa, Masahiro Saikou and Masaaki Tsuchida 500

Indexing Spoken Documents with Hierarchical Semantic Structures: Semantic Tree-to-string Alignment Models
Xiaodan Zhu, Colin Cherry and Gerald Penn 509

Structured and Extended Named Entity Evaluation in Automatic Speech Transcriptions
Olivier Galibert, Sophie Rosset, Cyril Grouin, Pierre Zweigenbaum and Ludovic Quintard 518

Normalising Audio Transcriptions for Unwritten Languages
Adel Foda and Steven Bird 527

Similarity Based Language Model Construction for Voice Activated Open-Domain Question Answering
Istvan Varga, Kiyonori Ohtake, Kentaro Torisawa, Stijn De Saeger, Teruhisa Misu, Shigeki Matsuda and Jun'ichi Kazama 536

The application of chordal graphs to inferring phylogenetic trees of languages
Jessica Enright and Grzegorz Kondrak 545

Cross-domain Feature Selection for Language Identification
Marco Lui and Timothy Baldwin 553

A Wikipedia-LDA Model for Entity Linking with Batch Size Changing Instance Selection
Wei Zhang, Jian Su and Chew-Lim Tan 562

Discovering Latent Concepts and Exploiting Ontological Features for Semantic Text Search
Vuong M. Ngo and Tru H. Cao 571

CLGVSM: Adapting Generalized Vector Space Model to Cross-lingual Document Clustering
Guoyu Tang, Yunqing Xia, Min Zhang, Haizhou Li and Fang Zheng 580

Thread Cleaning and Merging for Microblog Topic Detection
Jianfeng Zhang, Yunqing Xia, Bin Ma, Jianmin Yao and Yu Hong 589

Training a BN-based user model for dialogue simulation with missing data
Stéphane Rossignol, Olivier Pietquin and Michel Ianotto 598

Automatic identification of general and specific sentences by leveraging discourse annotations
Annie Louis and Ani Nenkova 605

A POS-based Ensemble Model for Cross-domain Sentiment Classification
Rui Xia and Chengqing Zong 614

Ensemble-style Self-training on Citation Classification
Cailing Dong and Ulrich Schäfer 623

Back to the Roots of Genres: Text Classification by Language Function
Henning Wachsmuth and Kathrin Bujna 632

Transductive Minimum Error Rate Training for Statistical Machine Translation
Yinggong Zhao, Shujie Liu, Yangsheng Ji, Jiajun Chen and Guodong Zhou 641

Distributed Minimum Error Rate Training of SMT using Particle Swarm Optimization
Jun Suzuki, Kevin Duh and Masaaki Nagata 649

Going Beyond Word Cooccurrences in Global Lexical Selection for Statistical Machine Translation using a Multilayer Perceptron
Alexandre Patry and Philippe Langlais 658

System Combination Using Discriminative Cross-Adaptation
Jacob Devlin, Antti-Veikko Rosti, Sankaranarayanan Ananthakrishnan and Spyros Matsoukas . 667

Word Sense Disambiguation by Combining Labeled Data Expansion and Semi-Supervised Learning Method
Sanae Fujita and Akinori Fujino 676

Combining ConceptNet and WordNet for Word Sense Disambiguation
Junpeng Chen and Juan Liu 686

It Takes Two to Tango: A Bilingual Unsupervised Approach for Estimating Sense Distributions using Expectation Maximization
Mitesh M Khapra, Salil Joshi and Pushpak Bhattacharyya 695

Dynamic and Static Prototype Vectors for Semantic Composition
Siva Reddy, Ioannis Klapaftis, Diana McCarthy and Suresh Manandhar 705

Using Prediction from Sentential Scope to Build a Pseudo Co-Testing Learner for Event Extraction
Shasha Liao and Ralph Grishman 714

Text Segmentation and Graph-based Method for Template Filling in Information Extraction
Ludovic Jean-Louis, Romaric Besançon and Olivier Ferret 723

Joint Distant and Direct Supervision for Relation Extraction
Truc-Vien T. Nguyen and Alessandro Moschitti 732

A Cross-lingual Annotation Projection-based Self-supervision Approach for Open Information Extraction
Seokhwan Kim, Minwoo Jeong, Jonghoon Lee and Gary Geunbae Lee 741

Exploring Difficulties in Parsing Imperatives and Questions
Tadayoshi Hara, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii 749

A Discriminative Approach to Japanese Zero Anaphora Resolution with Large-scale Lexicalized Case Frames
Ryohei Sasano and Sadao Kurohashi 758

An Empirical Comparison of Unknown Word Prediction Methods
Kostadin Cholakov, Gertjan van Noord, Valia Kordoni and Yi Zhang 767

Training Dependency Parsers from Partially Annotated Corpora
Daniel Flannery, Yusuke Miayo, Graham Neubig and Shinsuke Mori 776

A Breadth-First Representation for Tree Matching in Large Scale Forest-Based Translation
Sumukh Ghodke, Steven Bird and Rui Zhang 785

Bayesian Subtree Alignment Model based on Dependency Trees
Toshiaki Nakazawa and Sadao Kurohashi 794

Enriching SMT Training Data via Paraphrasing
Wei He, Shiqi Zhao, Haifeng Wang and Ting Liu 803

Translation Quality Indicators for Pivot-based Statistical MT
Michael Paul and Eiichiro Sumita 811

Source Error-Projection for Sample Selection in Phrase-Based SMT for Resource-Poor Languages
Sankaranarayanan Ananthakrishnan, Shiv Vitaladevuni, Rohit Prasad and Prem Natarajan 819

A Named Entity Recognition Method based on Decomposition and Concatenation of Word Chunks
Tomoya Iwakura, Hiroya Takamura and Manabu Okumura 828

Extract Chinese Unknown Words from a Large-scale Corpus Using Morphological and Distributional Evidences
Kaixu Zhang, Ruining Wang, Ping Xue and Maosong Sun 837

Entity Disambiguation Using a Markov-Logic Network
Hong-Jie Dai, Richard Tzong-Han Tsai and Wen-Lian Hsu 846

Named Entity Recognition in Chinese News Comments on the Web
Xiaojun Wan, Liang Zong, Xiaojiang Huang, Tengfei Ma, Houping Jia, Yuqian Wu and Jianguo Xiao 856

Clustering Semantically Equivalent Words into Cognate Sets in Multilingual Lists
Bradley Hauer and Grzegorz Kondrak 865

Extending WordNet with Hypernyms and Siblings Acquired from Wikipedia
Ichiro Yamada, Jong-Hoon Oh, Chikara Hashimoto, Kentaro Torisawa, Jun'ichi Kazama, Stijn De Saeger and Takuya Kawada 874

What Psycholinguists Know About Chemistry: Aligning Wiktionary and WordNet for Increased Domain Coverage
Christian M. Meyer and Iryna Gurevych 883

From News to Comment: Resources and Benchmarks for Parsing the Language of Web 2.0
Jennifer Foster, Ozlem Cetinoglu, Joachim Wagner, Joseph Le Roux, Joakim Nivre, Deirdre Hogan and Josef van Genabith 893

Toward Finding Semantic Relations not Written in a Single Sentence: An Inference Method using Auto-Discovered Rules
Masaaki Tsuchida, Kentaro Torisawa, Stijn De Saeger, Jong Hoon Oh, Jun'ichi Kazama, Chikara Hashimoto and Hayato Ohwada 902

Fleshing it out: A Supervised Approach to MWE-token and MWE-type Classification
Richard Fothergill and Timothy Baldwin 911

Identification of relations between answers with global constraints for Community-based Question Answering services
Hikaru Yokono, Takaaki Hasegawa, Genichiro Kikui and Manabu Okumura 920

Automatically Generating Questions from Queries for Community-based Question Answering
Shiqi Zhao, Haifeng Wang, Chao Li, Ting Liu and Yi Guan 929

Question classification based on an extended class sequential rule model
Zijing Hui, Juan Liu and Lumei Ouyang 938

K2Q: Generating Natural Language Questions from Keywords with User Refinements
Zhicheng Zheng, Xiance Si, Edward Chang and Xiaoyan Zhu 947

Answering Complex Questions via Exploiting Social Q&A Collection
Youzheng Wu, Chiori Hori, Hisashi Kawai and Hideki Kashioka 956

Safety Information Mining — What can NLP do in a disaster—
Graham Neubig, Yuichiroh Matsubayashi, Masato Hagiwara and Koji Murakami 965

A Character-Level Machine Translation Approach for Normalization of SMS Abbreviations
Deana Pennell and Yang Liu 974

Using Text Reviews for Product Entity Completion
Mrinmaya Sachan, Tanveer Faruquie, L. V. Subramaniam and Mukesh Mohania 983

Mining bilingual topic hierarchies from unaligned text
Sumit Negi 992

Efficient Near-Duplicate Detection for Q&A Forum
Yan Wu, Qi Zhang and Xuanjing Huang 1001

A Graph-based Method for Entity Linking
Yuhang Guo, Wanxiang Che, Ting Liu and Sheng Li 1010

Harvesting Related Entities with a Search Engine
Shuqi Sun, Shiqi Zhao, Muyun Yang, Haifeng Wang and Sheng Li 1019

Acquiring Strongly-related Events using Predicate-argument Co-occurring Statistics and Case Frames
Tomohide Shibata and Sadao Kurohashi 1028

Relevance Feedback using Latent Information
Jun Harashima and Sadao Kurohashi 1037

Passage Retrieval for Information Extraction using Distant Supervision
Wei Xu, Ralph Grishman and Le Zhao 1046

Using Context Inference to Improve Sentence Ordering for Multi-document Summarization
Peifeng Li, Guangxi Deng and Qiaoming Zhu 1055

Enhancing extraction based summarization with outside word space
Christian Smith and Arne Jönsson 1062

Shallow Discourse Parsing with Conditional Random Fields
Sucheta Ghosh, Richard Johansson, Giuseppe Riccardi and Sara Tonelli 1071

Relational Lasso —An Improved Method Using the Relations Among Features—
Kotaro Kitagawa and Kumiko Tanaka-Ishii 1080

Enhance Top-down method with Meta-Classification for Very Large-scale Hierarchical Classification
Xiao-lin Wang, Hai Zhao and Bao-Liang Lu 1089

Using Syntactic and Shallow Semantic Kernels to Improve Multi-Modality Manifold-Ranking for Topic-Focused Multi-Document Summarization
Yllias Chali, Sadid A. Hasan and Kaisar Imam 1098

Automatic Determination of a Domain Adaptation Method for Word Sense Disambiguation Using Decision Tree Learning
Kanako Komiya and Manabu Okumura 1107

Learning from Chinese-English Parallel Data for Chinese Tense Prediction
Feifan Liu, Fei Liu and Yang Liu 1116

Jointly Extracting Japanese Predicate-Argument Relation with Markov Logic
Katsumasa Yoshikawa, Masayuki Asahara and Yuji Matsumoto 1125

Word Meaning in Context: A Simple and Effective Vector Model
Stefan Thater, Hagen Fürstenau and Manfred Pinkal 1134

Automatic Analysis of Semantic Coherence in Academic Abstracts Written in Portuguese
Vinícius Mourão Alves de Souza and Valéria Delisandra Feltrim 1144

Sentence Subjectivity Detection with Weakly-Supervised Learning
Chenghua Lin, Yulan He and Richard Everson 1153

Opinion Expression Mining by Exploiting Keyphrase Extraction
Gábor Berend 1162

Extracting Resource Terms for Sentiment Analysis
Lei Zhang and Bing Liu 1171

Towards Context-Based Subjectivity Analysis
Farah Benamara, Baptiste Chardon, Yannick Mathieu and Vladimir Popescu 1180

Compression Methods by Code Mapping and Code Dividing for Chinese Dictionary Stored in a Double-Array Trie
Huidan Liu, Minghua Nuo, Longlong Ma, Jian Wu and Yeping He 1189

Functional Elements and POS Categories
Qiuye Zhao and Mitch Marcus 1198

Joint Alignment and Artificial Data Generation: An Empirical Study of Pivot-based Machine Transliteration
Min Zhang, Xiangyu Duan, Ming Liu, Yunqing Xia and Haizhou Li 1207

Incremental Joint POS Tagging and Dependency Parsing in Chinese
Jun Hatori, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii 1216

Extending the adverbial coverage of a NLP oriented resource for French
Elsa Tolone and Stavroula Voyatzi 1225

Linguistic Phenomena, Analyses, and Representations: Understanding Conversion between Treebanks
Rajesh Bhatt, Owen Rambow and Fei Xia 1234

Automatic Transformation of the Thai Categorial Grammar Treebank to Dependency Trees
Christian Rishøj, Taneth Ruangrajitpakorn, Prachya Boonkwan and Thepchai Supnithi 1243

Parse Reranking Based on Higher-Order Lexical Dependencies
Zhiguo Wang and Chengqing Zong 1251

Improving Part-of-speech Tagging for Context-free Parsing
Xiao Chen and Chunyu Kit 1260

Models Cascade for Tree-Structured Named Entity Detection
Marco Dinarelli and Sophie Rosset 1269

Clausal parsing helps data-driven dependency parsing: Experiments with Hindi
Samar Husain, Phani Gadde, Joakim Nivre and Rajeev Sangal 1279

Word-reordering for Statistical Machine Translation Using Trigram Language Model
Jing He and Hongyu Liang 1288

Extracting Hierarchical Rules from a Weighted Alignment Matrix
Zhaopeng Tu, Yang Liu, Qun Liu and Shouxun Lin 1294

Integration of Reduplicated Multiword Expressions and Named Entities in a Phrase Based Statistical Machine Translation System
Thoudam Doren Singh and Sivaji Bandyopadhyay 1304

Regularizing Mono- and Bi-Word Models for Word Alignment
Thomas Schoenemann 1313

Parametric Weighting of Parallel Data for Statistical Machine Translation
Kashif Shah, Loïc Barrault and Holger Schwenk 1323

An Effective and Robust Framework for Transliteration Exploration
EA-EE JAN, Niyu Ge, Shih-Hsiang Lin and Berlin Chen 1332

Part B: Short Papers

An Evaluation of Alternative Strategies for Implementing Dialogue Policies Using Statistical Classification and Hand-Authored Rules
David DeVault, Anton Leuski and Kenji Sagae 1341

Reducing Asymmetry between language-pairs to Improve Alignment and Translation Quality
Rashmi Gangadharaiah 1346

Clause-Based Reordering Constraints to Improve Statistical Machine Translation
Ananthakrishnan Ramanathan, Pushpak Bhattacharyya, Karthik Visweswariah, Kushal Ladha and Ankur Gandhe 1351

Generalized Minimum Bayes Risk System Combination
Kevin Duh, Katsuhito Sudoh, Xianchao Wu, Hajime Tsukada and Masaaki Nagata 1356

Enhancing scarce-resource language translation through pivot combinations
Marta R. Costa-jussà, Carlos Henríquez and Rafael E. Banchs 1361

A Baseline System for Chinese Near-Synonym Choice
Liang-Chih Yu, Wei-Nan Chien and Shih-Ting Chen 1366

Cluster Labelling based on Concepts in a Machine-Readable Dictionary
Fumiyo Fukumoto and Yoshimi Suzuki 1371

Text Patterns and Compression Models for Semantic Class Learning
Chung-Yao Chuang, Yi-Hsun Lee and Wen-Lian Hsu 1376

Potts Model on the Case Fillers for Word Sense Disambiguation
Hiroya Takamura and Manabu Okumura 1382

Improving Word Sense Induction by Exploiting Semantic Relevance
Zhenzhong Zhang and Le Sun 1387

Predicting Word Clipping with Latent Semantic Analysis
Julian Brooke, Tong Wang and Graeme Hirst 1392

A Semantic Relatedness Measure Based on Combined Encyclopedic, Ontological and Collocational Knowledge
Yannis Haralambous and Vitaly Klyuev 1397

Going Beyond Text: A Hybrid Image-Text Approach for Measuring Word Relatedness
Chee Wee Leong and Rada Mihalcea 1403

Domain Independent Model for Product Attribute Extraction from User Reviews using Wikipedia
Sudheer Kovelamudi, Sethu Ramalingam, Arpit Sood and Vasudeva Varma 1408

Finding Problem Solving Threads in Online Forum
Zhonghua Qu and Yang Liu 1413

Compiling Learner Corpus Data of Linguistic Output and Language Processing in Speaking, Listening, Writing, and Reading
Katsunori Kotani, Takehiko Yoshimi, Hiroaki Nanjo and Hitoshi Isahara 1418

Mining the Sentiment Expectation of Nouns Using Bootstrapping Method
Miaomiao Wen and Yunfang Wu 1423

An Analysis of Questions in a Q&A Site Resubmitted Based on Indications of Unclear Points of Original Questions
Masahiro Kojima, Yasuhiko Watanabe and Yoshihiro Okada 1428

Diversifying Information Needs in Results of Question Retrieval
Yaoyun Zhang, Xiaolong Wang, Xuan Wang, Ruifeng Xu, Jun Xu and ShiXi Fan 1432

Beyond Normalization: Pragmatics of Word Form in Text Messages
Tyler Baldwin and Joyce Chai 1437

Chinese Discourse Relation Recognition
Hen-Hsen Huang and Hsin-Hsi Chen 1442

Improving Chinese POS Tagging with Dependency Parsing
Zhenghua Li, Wanxiang Che and Ting Liu 1447

Exploring self training for Hindi dependency parsing
Rahul Goutam and Bharat Ram Ambati 1452

Reduction of Search Space to Annotate Monolingual Corpora
Prajol Shrestha, Christine Jacquin and Beatrice Daille 1457

Toward a Parallel Corpus of Spoken Cantonese and Written Chinese
John Lee 1462

Query Expansion for IR using Knowledge-Based Relatedness
Arantxa Otegi, Xabier Arregi and Eneko Agirre 1467

Word Sense Disambiguation Corpora Acquisition via Confirmation Code
Wanxiang Che and Ting Liu 1472

Opinion Expression Mining by Exploiting Keyphrase Extraction

Gábor Berend
Department of Informatics,
University of Szeged
2. Árpád tér, H-6720, Szeged, Hungary
berendg@inf.u-szeged.hu

Abstract

In this paper, we shall introduce a system for extracting the keyphrases for the reason of authors' opinion from product reviews. The datasets for two fairly different product review domains related to movies and mobile phones were constructed semi-automatically based on the pros and cons entered by the authors. The system illustrates that the classic supervised keyphrase extraction approach – mostly used for scientific genre previously – could be adapted for opinion-related keyphrases. Besides adapting the original framework to this special task through defining novel, task-specific features, an efficient way of representing keyphrase candidates will be demonstrated as well. The paper also provides a comparison of the effectiveness of the standard keyphrase extraction features and that of the system designed for the special task of opinion expression mining.

1 Introduction

The amount of community-generated contents on the Web has been steadily growing and most of the end-user contents (e.g. blogs and customer reviews) are likely to deal with the author's emotions and opinions towards some subject. The automatic analysis of such material is useful for both companies and consumers. Companies can easily get an overview of what people think of their products and services and what their most important strengths and weaknesses are while users can have access to information from the Web before purchasing some product.

In this paper we will introduce a system which assigns pro and con keyphrases (free-text annotation) to product reviews. When dealing with product reviews, our definition of keyphrases is the set of phrases that make the opinion-holder feel negative or positive towards a given product, i.e. they should be the reason why the author likes or dislikes the product in question (e.g. *cheap price*, *convenient user interface*). Here, we adapted the general keyphrase extraction procedure from the scientific publications domain (Witten et al., 1999; Turney, 2003) to the extraction of opinion-reasoning features. However, our task is rather different since we aim at identifying the reasons for opinions, instead of keyphrases that represent the content of the whole document.

The supervised keyphrase extractor to be introduced here was trained on the pros and cons assigned to the reviews by their authors on the `epinions.com` site. These pros and cons are ill-structured free-text annotations and their length, depth and style are extremely heterogeneous. In order to have clean gold-standard corpora, we manually revised the segmentation and the contents of the pros and cons, and obtained sets of tag-like keyphrases.

2 Related work

There have been many studies on opinion mining (Turney, 2002; Pang et al., 2002; Titov and McDonald, 2008; Liu and Seneff, 2009). Our approach relates to previous work on the extraction of reasons for opinions. Most of these papers treat the task of mining reasons from product reviews as one of identifying sentences that express the author's negative or positive feelings (Hu and Liu, 2004a; Popescu and Etzioni, 2005). This paper is clearly distinguishable from them as our goal is to find the reasons for opinions expressed by phrases and we aim the task of phrase extraction instead of sentence recognition.

This work differs in important aspects even from the frequent pattern mining-based approach of (Hu and Liu, 2004b) since they regarded the main task of mining opinion features with respect

Proceedings of the 5th International Joint Conference on Natural Language Processing, pages 1162–1170,
Chiang Mai, Thailand, November 8 – 13, 2011. ©2011 AFNLP

to a group of products, not individually at review-level as we did. Even if an opinion feature phrase is feasible for a given product-type, it is not necessary that all of its occurrence are accompanied with sentiments expressed towards it (e.g. *The phone comes in red and black colors*, where *color* could be an appropriate product feature, but not an opinion-forming phrase).

A similar task to pro and con extraction gathers the key aspects from document sets, which has also gained interest recently (Sullivan, 2008; Branavan et al., 2008; Liu and Seneff, 2009). Existing aspect extraction systems first identify a number of aspects throughout the whole review set, then they automatically assign items from this pre-recognized set of aspects to each unseen review. Hence, they work at the corpus level and restrict themselves to using only a pre-defined number of aspects.

The approach presented here differs from these studies in the sense that it looks for the reason phrases themselves review by review, instead of multi-labeling some aspects. These approaches are intended for applications used by companies who would like to obtain a general overview about a product or would like to monitor the polarity relating to their products in a particular community. In contrast, we introduce here a keyphrase extraction-based approach which works at the document level as it extracts keyphrases from reviews which are handled independently of each other. This approach is more appropriate for the consumers, who would like to be informed before purchasing some product.

The work of Kim and Hovy (2006) lies probably the closest to our one. They addressed the task of extracting con and pro sentences, i.e. the sentences on why the reviewers liked or disliked the product. They also note that such pro and con expressions can differ from positive and negative opinion expressions as factual sentences can also be reason sentences (e.g. *Video drains battery.*). Here the difference is that they extracted sentences, but we targeted phrase extraction.

Most of the keyphrase extraction approaches (Witten et al., 1999; Turney, 2003; Medelyan et al., 2009; Kim et al., 2010) work on the scientific domain and extract phrases from one document that are the most characteristic of its content. In these supervised approaches keyphrase extraction is regarded as a classification task, in which certain n-grams of a specific document function as keyphrase candidates, and the task is to classify them as proper or improper keyphrases. Here, our task formalization of keyphrase extraction is adapted from this line of research for opinion mining and we focus on the extraction of phrases from product reviews that also bear subjectivity and induce sentiments in its author. As community generated pros and cons can provide abundant training samples and our goal is to extract the users' own words, here we also follow this supervised keyphrase extraction procedure.

3 Opinion Phrase Extraction Framework

Here, we employed a supervised machine learning approach for the extraction of reason keyphrases from a given review. Candidate terms were extracted from the text of the review and those present in the extracted set of pros and cons were regarded as positive examples during training and evaluation. Maximum Entropy classifiers were trained and the keyphrase candidates with the highest posteriori probabilities were selected to be keyphrases for a review of a test document in question. In the following subsections we will describe how keyphrase candidates and the feature space representing them were constructed.

3.1 Candidate term generation

One key aspect in keyphrase extraction is the way keyphrase candidates are selected and represented. As usually the number of potentially extracted n-grams and that of genuine keyphrases among them show high imbalancedness, keyphrase candidates are worth to be filtered, instead of using any successive n-grams. For this reason we limited the maximal length of the extracted phrases to at most 4 tokens and also required that the phrases should begin with either a non-stopword adjective, verb or noun and should end to either a non-stopword noun or adjective.

As for the filtration of the candidate set, a new step is introduced here, which omits normalized phrases that had only such occurrences which contained stopwords. This simple step proved effective in excluding many non-proper opinion phrases (i.e. increasing the maximal precision achievable) at the cost of discarding only a small proportion of proper phrases (i.e. slightly decreasing the best recall achievable).

Once we had the keyphrase candidates, they had

to be brought to a normalized form. The normalization of an n-gram consisted of lowercasing and Porter-stemming each of the lemmatized forms of its tokens, then putting these stems into alphabetical order (while omitting the stems of stopword tokens). With this kind of representation it was then possible to handle two orthographically different, but semantically equivalent phrases, such as '*the screen is tiny*' and '*TINY screen*' in the same way.

Previous works on keyphrase extraction also usually carry out this step of normalization, however, here we did it in such a manner that a mapping to each of the original orthographic forms of a normalized form and its corresponding context (i.e. the sentences containing it) was preserved at the same time and that could be successfully utilized at later processing steps.

To provide an alternative way of normalizing phrases, experiments relying on the usage of WordNet (Fellbaum, 1998) were also conducted. In these settings the normalized form of a single token was determined by first searching for all its synsets (in the case of verbs, these were such noun synsets that were in derivative relation with the synsets of the verb word form). Then instead of Porter-stemming the original token, its most frequent word form was stemmed, based on the estimated frequencies of WordNet for all the word forms of the synsets of the original token. In this way two – originally differently stemmed – word forms, such as *decide* and *decision* could be stemmed to the same root forms. Another advantage of this procedure is that it is able to handle semantic similarity to some extent.

The remaining parts of the normalization procedure were left unchanged (i.e. lowercasing and alphabetical ordering of the normalized forms of the individual tokens). Later, in the Results section, the effect of this kind of normalization will be shown.

Candidate terms were handled at the review level instead of occurrence level. This means that each normalized occurrence of a keyphrase candidate was gathered from the document and the feature values for the candidate term aggregate over its occurrences.

3.2 Feature representation

We constructed a rich feature set to represent the review-level keyphrase candidates. The feature space incorporates features calculated on the basis of the normalized phrases themselves, but more importantly, thanks to the mapping between the normalized phrase forms and their original occurrences, new contextual and orthographic features were possible to incorporate.

Features that could be generally used for any kind of keyphrase extracting task (e.g. that makes use of multiword expressions or character suffixes in a special way) and ones designed especially for the novel task of opinion phrase extraction (e.g. that uses SentiWordNet to determine polarity) as well as the standard features of keyphrase extraction are both introduced in the following.

Standard Features Since we assumed that the underlying principles of extracting opinionated phrases are quite similar to that of extracting standard (most of the time scientific) keyphrases, features of the standard setting were applied in this task as well. The most common ones, introduced by KEA (Witten et al., 1999) are the **Tf-idf** value and the **relative position** of the first occurrence of a candidate phrase within a document. We should note that KEA is primarily designed for keyphrase extraction from scientific publications and whereas the position of the first occurrence might be indicative in research papers, product reviews usually do not contain a summarizing "abstract" at the beginning. For these reasons we chose these features as the ones which form our baseline system. **Phrase length** is also a common feature, which was defined here as the number of the non-stopword tokens of an opinion phrase candidate.

Linguistic and orthographic features Since certain POS-codes are more frequent than others among genuine keyphrases, features generated by POS-codes belonging to an occurrence of a normalized phrase were applied. As **POS-code** sequences seem to be more informative, instead of simply indicating which POS-codes were assigned to any orthographic alternation of a normalized keyphrase candidate, it would be desirable to store the POS-code sequences in their full length as well. However, doing so might affect dimensionality in a negative way (especially when having few training data), i.e. the number of all the possible POS-code sequences ranging from lengths of 1 to 4 is too much. To overcome this issue, positional information was added to the POS-code features derived from the tokens of an n-gram. Features

of POS-codes that were assigned to a token being itself a 1-token long keyphrase candidate, at the beginning, at the end, in between an n-gram, got a prefix `S-`, `B-`, `E-` and `I-`, respectively. For instance, the phrase *cheap/JJ phone/NN* induces the features {*B-JJ*, *E-NN*}, whereas the 1-token-long phrase *cheap/JJ* induces the feature {*S-JJ*}. Finally, numeric values for a normalized candidate phrase were assigned based on the distribution of the different POS-related features of all the running-text forms of a normalized phrase.

We introduced features exploiting the syntactic context of a candidate with parse trees. For an n-gram with respect to all the sentences it was contained in a given document, this feature stored the average and the minimal depths of those **NP-rooted trees** that contained the whole n-gram in its yield. These features are intended to express the "noun phraseness" of the phrase.

Features generated from the **character suffixes** of the individual tokens of the occurrences of a normalized keyphrase candidate were also employed. Character suffix features also incorporated positional information, similarly as it was done in the case of POS features. The suffixes themselves came from the last 2 and 3 characters of the tokens constructing an n-gram. For instance, the features induced by (and thus assigned with true value) for the phrase *ch**eap** ph**one*** are {*B-eap*, *B-ap*, *E-one*, *E-ne*}.

Opinionated phrases often bear special orthographic characteristics, e.g. in the case of *so sl**ooo**w* or *CHEAP*. Due to the fact that the original forms of the phrases are stored in our representation, it was possible to construct two features for this phenomenon: the first feature is responsible for **character run**s (i.e. more than 2 of the same consecutive characters), and an other is responsible for **strange capitalization** (i.e. the presence of uppercase characters besides the initial one). The S-,B-,E-,I- prefixes were applied here as well, just like in the case of the **Named Entity** feature, which represented if a token was part of NE (with its type as well).

World knowledge-based features Features relying on the outer resources of Wikipedia and SentiWordNet were also exploited during our experiments. They were useful as world knowledge could be incorporated by their means.

Multiword expressions are lexical items that can be decomposed into single words and display idiosyncratic features (Sag et al., 2002), in other words, they are lexical items that contain space.

To measure the added value of MWEs in the task of opinion phrase extraction, a set of features was designed that indicated whether a certain phrase candidate (1) is an MWE on its own (e.g. *ease of use*), (2) can be composed from more MWEs on the list (e.g. *mobile internet access*), or is just the (3) superstring of at least one MWE from the list (e.g. *send text messages*). In order to be able to make such decisions, a wide list of MWEs was constructed from Wikipedia (dump 2011-01-07): all the links and formatted (i.e. bold or italic) text were gathered that were at least two tokens in length, started with lowercase letters and contained only English characters or some punctuation. Finally, an alignment of the elements of the list and the contexts of the reviews of the dataset was carried out (taking care of linguistic alternations and POS-tag matchings).

A more sophisticated surface-based feature used external information as well on the individual tokens of a phrase. It relied on the **sentiment scores** of SentiWordNet (Esuli et al., 2010), a publicly available database that contains a subset of the synsets of the Princeton Wordnet with positivity, negativity and neutrality scores assigned to each one, depending on the use of its sentiment orientation (which can be regarded as the probability of a phrase belonging to a synset being mentioned in a positive, negative or neutral context). These scores were utilized for the calculation of the sentiment orientations of each token of a keyphrase candidate. Surface-based SentiWordnet-calculated feature values for a keyphrase candidate included the *maximal positivity and negativity and subjectivity* scores of the individual tokens and the *total sum* over all the tokens of one phrase.

Sentence-based features were also defined based on SentiWordNet as it was also used to check for the presence of **indicator terms** within the sentences containing a candidate phrase. Those word forms were gathered from SentiWordNet, for which the sum of the average positivity and negativity sentiments scores among all its synsets were above 0.5 (i.e. the ones that are more likely to have some kind of polarity). Then for a given keyphrase candidate of a given document, a true value was assigned to the SentiWordNet-derived indicator features that had at least one

co-occurrence within the same sentence with the keyphrase candidate in the same document.

SentiWordnet was also used to investigate the entire sentences that contained a phrase candidate. This kind of feature calculated the sum of every sentiment score in each sentence where a given phrase candidate was present. Then the mean and the deviation of the sum of the sentiment scores were calculated for each token of the phrase-containing sentences and assigned to the phrase candidate. The mean of the sentiment scores of the individual sentences yielded a general score on the **sentiment orientation** of the sentences containing a candidate phrase, while higher values for the **deviation** was intended to capture cases when a reviewer writes both factual (i.e. uses few opinionated words) and non-factual (i.e. uses more emotional phrases and opinions) sentences about a product.

Finally, Wikipedia was also used to incorporate semantic features from its category hierarchy. (Wikipedia categories form a taxonomy, indicating which article belongs to which (sub)category). In the case of a candidate phrase all the nominal parts of the normalized titles of **Wikipedia categories** for its related Wikipedia articles were added as separate binary features to the feature space. The normalization of the Wikipedia category names was similar to that of keyphrase candidates. For instance, given the candidate phrase '*service quality*' the feature *wiki_control_qual* is set to true since the Wikipedia article named *Service quality* is in the category *Quality control*.

Document and corpus-level features Among document-level features, the **standard deviation of the relative positions** compared to the document length was a measure to be computed. Higher values of the deviation in the position means that the reviewer keeps repeating some phrase from the beginning to the end of the review, which might indicate that this phrase is of higher importance for them.

As verbs often contribute to the sentiment polarity of the noun phrases they accompany (e.g. '*I adore its fancy screen.*' versus '*I bought this phone one year ago.*'), a set of features was introduced to deal with the **indicative verbs** in the context of candidate phrase occurrences within their document. For this feature to be calculated we took those verbs as indicators that occurred at least 100 times in the whole training dataset. When calculating a feature value for an opinionated-phrase candidate, the algorithm matched all of its occurrences in a document against every indicator verb. For the calculation of the feature value for a given phrase candidate – indicator verb pair, a syntactic distance value was first defined. This syntactic distance was equal to the minimal height of the subtree which contained both the keyphrase candidate and the indicator verb itself to the left among all the sentences associated with a document that contained the keyphrase candidate. The feature value was then determined by simply taking the reciprocal of this semantic distance. This way, the feature value was scaled between 0 and 1. (Note that for indicator verbs that were not present in any of the sentences containing a phrase candidate associated with a document, the semantic distance value was defined to be infinity, the limit value of the reciprocal of which is 0.)

Quite general characteristics of reason-expressing phrases can also be captured at the corpus level. Simply using the number of times an argument phrase aspirant was assigned to a review as a proper phrase on the training dataset was also taken into account as a **corpus-level** feature since the same proper opinion phrases can easily reoccur regarding products of the same type.

4 Experiments

Experiments were carried out on two fairly different types of product reviews, namely mobile phones and movies. We use standard keyphrase extraction evaluation metrics and baselines for evaluating our pros and cons extractor system.

4.1 Datasets

In our experiments, we crawled two quite different domains of product reviews, i.e. mobile phone and movie reviews from the review portal `epinions.com`. For both domains, 2000 reviews were crawled from `epinions.com` and an additional of 50 and 75 reviews for measuring inter-annotator agreement, respectively. This corpus is quite noisy (similarly to other user-generated contents); run-on sentences and improper punctuation were common, as well as grammatically incorrect sentences since reviews were often written by non-native English speakers. [1]

[1] All the data used in our experiments are available at `http://rgai.inf.u-szeged.hu/proCon`

	Mobiles	Movies
Number of reviews	2009	1962
Avg. sentence/review	31.9	29.8
Avg. tokens/sentence	16.1	17.0
Avg. keyphrases/review	4.7	3.2
Avg. keyphrase candidates/review	130.38	135.89

Table 1: Size-related statistics of the corpora

The list of pros and cons was inconsistent too in the sense that some reviewers used full sentences to express their opinions, while usually a few token-long phrases were given by others. The segmentation of their elements was marked in various ways among reviews (e.g. comma, semicolon, ampersand or the *and* token) and even differed sometimes within the very same review. There were many general or uninformative pros and cons (like *none* or *everything* as a pro phrase) as well.

In order to have a consistent gold-standard annotation for training and evaluation, we manually refined the pros and cons of the reviews in the corpora. In the first step, the automatic preprocessing of the segmentation of pros and cons was checked by human annotators. Our automatic segmentation method split the lines containing pros and cons along the most frequent separators. This segmentation was corrected by the annotators in 7.5% of the reviews. Then the human annotators also marked the general pros and cons (11.1% of the pro and con phrases) and the reviews without any identified keyphrases were discarded.

4.2 Evaluation issues

Keyphrase extraction systems are traditionally evaluated on the top-n ranked keyphrase candidates for each document by F-score (Kim et al., 2010), which combines the precision and recall of the correct keyphrases' class. Evaluation is carried out in a strict manner as a top-ranked keyphrase candidate is accepted if it has exactly the same standardized form as one of the keyphrases assigned to the review. The ranking of the phrase candidates was based on a probability estimation of a candidate belonging to the positive keyphrase class. Results reported here were obtained using 5-fold cross validation using Maximum Entropy classifier.

As we treated the mining of pros and cons as a supervised keyphrase extraction task, we conducted measurements with KEA (Witten et al., 1999), which is one of the most cited publicly available automatic keyphrase extraction system. However, we should note that due to the fact that our phrase extraction and representation strategy (and even the determination of true positive instances to some extent) slightly differs from that of KEA, the added values of our features should rather be compared to our second Baseline System (BL_{WN}) which uses WordNet for candidate phrase normalization. The baseline systems use our framework, with the feature set of KEA, which consists of tf-idf feature and the relative first occurrence of a keyphrase candidate. The only difference among the two baseline systems is that BL does not apply the WordNet-based normalization of phrase candidates introduced in Section 3.1.

Since we had the same findings as Branavan et al. (2008) that authors often omit several opinion forming aspects from their pros and cons listings that they later include in their review, we decided to determine the complete lists of pros and cons manually, that is, to compose pro and con phrases on the basis of the reviews. Due to the highly subjective nature of sentiments, the determination of sentiment-affecting pro and con phrases was carried out by three linguists, who were asked to annotate a 25-document subset of the mobile phone dataset. Their averaged agreements for the determination of pro phrases are 0.701 and 0.533 for Dice's coefficient and Jaccard index, and 0.69 and 0.526 for cons, respectively.

4.3 Results

In our experiments all the linguistic processing of the product reviews were carried out using Stanford CoreNLP. It uses the Maximum Entropy POS-tagger of Toutanova and Manning (2000) and syntactic parsing works on the basis of Klein and Manning (2003). The ranking of the candidate keyphrases was based on the posteriori probabilities of the MALLET implementation (McCallum, 2002) of Maximum Entropy classifier (le Cessie and van Houwelingen, 1992).

During the fully automatic evaluation, we followed strict evaluation (see 4.2) that is commonly utilized in scientific keyphrase extraction tasks. Table 2 contains the results of the strict evaluation for both domains. However, since strict evaluation is more likely to suit the evaluation of scientific keyphrase extraction better, i.e. semantically equivalent but different word forms are less common at that domain, we conducted human evaluation on the 25-document subset of the mobile

Feature	Mobiles Top-5	Mobiles Top-10	Mobiles Top-15	Movies Top-5	Movies Top-10	Movies Top-15
KEA	1.72/1.84/1.77	1.42/3.04/1.94	1.39/4.48/2.12	1.21/1.93/1.49	0.98/3.13/1.5	0.89/4.26/1.48
BL	2.6/2.8/2.73	2.6/5.5/3.54	2.6/8.2/3.93	1.6/2.5/1.95	1.5/4.9/2.34	1.6/7.4/2.58
BL_{WN}	2.7/2.9/2.8	2.7/5.8/3.68	2.7/8.7/4.12	1.7/2.8/2.14	1.7/5.4/2.61	1.7/8.2/2.88
IV	3.1/3.4/3.25§	2.9/6.2/3.92	2.8/9.1/4.31	2.4/3.7/2.9†	2.0/6.3/3.04§	1.9/8.8/3.09
KF	2.6/2.8/2.71	2.7/5.9/3.73	2.7/8.7/4.11	1.7/2.7/2.09	1.7/5.4/2.59	1.7/8.2/2.87
$Length$	3.2/3.4/3.26 §	3.1/6.6/4.18†	2.9/9.3/4.4	2.1/3.3/2.6	2.0/6.4/3.08§	2.0/9.1/3.22§
MWE	4.7/5.0/4.88‡	3.8/8.0/5.11‡	3.4/10.8/5.12‡	2.3/3.6/2.81†	2.0/6.3/3.06†	1.9/9.1/3.18§
POS	4.6/4.9/4.71‡	4.2/9.0/5.77‡	3.9/12.6/5.98‡	2.9/4.6/3.57‡	2.8/8.7/4.18‡	2.5/11.7/4.1‡
SWN	6.0/6.4/6.2‡	4.9/10.4/6.65‡	4.3/13.6/6.49‡	3.7/6.0/4.6‡	3.1/9.8/4.73‡	2.8/13.1/4.59‡
$StDev$	3.9/4.2/4.06‡	3.8/8.1/5.15‡	3.5/11.2/5.33‡	2.9/4.6/3.59‡	2.6/8.1/3.9‡	2.5/11.6/4.07‡
$Orth.$	3.2/3.4/3.28§	3.1/6.7/4.27†	2.9/9.5/4.49	3.0/4.7/3.65‡	2.5/7.8/3.76‡	2.3/10.9/3.82‡
$Suffix$	11.5/12.2/11.83‡	8.6/18.2/11.66‡	6.9/22.0/10.54‡	6.8/10.7/8.34‡	5.2/16.4/7.91‡	4.3/20.1/7.08‡
$Syntax$	3.5/3.7/3.61‡	3.0/6.4/4.06	2.8/9.1/4.33	2.3/3.6/2.78†	2.0/6.1/2.97§	1.9/9.1/3.2§
$Wiki$	11.9/12.7/12.25‡	8.1/17.4/11.09‡	6.3/20.1/9.63‡	8.8/13.9/10.78‡	6.3/19.8/9.59‡	4.8/22.5/7.9‡
$COMB$	**14.8/15.7/15.27**‡	10.4/22.0/14.11‡	8.0/25.4/12.17‡	**10.0/15.8/12.22**‡	7.0/21.9/10.63‡	5.3/24.6/8.67‡

Table 2: Performance using different features in the form of Precision/Recall/F-score obtained. IV, KF, SWN and Orth. stands for indicator verbs, corpus-level keyphrase frequency, SentiWordNet and orthography-driven features, respectively. Symbols §, † and ‡ in the upper index of a result indicates that it is significantly better compared to the baseline system which uses the WordNet based candidate phrase normalization (BL_{WN}) at confidence levels of 0.1, 0.05 and 0.01, based on Student's t-test, respectively. As it was only the KF feature which did not yield any significant improvement at all, the combined system (COMB) incorporated all the features but KF.

phone domain. The results of the manual evaluation is shown in 3.

4.4 Discussion

The fact that the highest F-scores for keyphrases are achieved when the number of extracted phrases is around the average number of pro and con phrases per reviews (i.e. between 3 and 5) suggests that our ordering of keyphrase candidates is quite effective (since once we find the number of keyphrases a document has, performance cannot really grow anymore).

Comparing the nature of the task of extracting keyphrases from scientific publications and that of product reviews, we shall take two observations: firstly, keyphrases of scientific documents are more universal, i.e. once we have the knowledge that the expression *distributed computing* was a good keyphrase for one scientific document, we can be more confident about it being a proper keyphrase for other documents within the same domain as well, whereas in the case of opinion phrases such as *pink color* can easily be mentioned in either opinionated and non-opinionated contexts. Secondly, besides scientific keyphrases being more ***universal***, they are more ***deterministic*** in the sense that there are fewer ways to express good keyphrases, e.g. suppose *simulated annealing* is a proper keyphrase for a scientific document, it is unlikely that an automatic system would extract *imitated annealing*, whereas in the case of product review the gold standard keyphrases often differ from their mention in the text (e.g. *tiny keys* and *small keys*).

The above mentioned examples suggest that opinion phrase extraction is more difficult to be performed and evaluated compared to scientific keyphrase extraction. We should note that the best performing system at SemEval-2010 (Kim et al., 2010) that dealt with the much simpler task of scientific keyphrase extraction achieved an F-score of 19.3 when evaluated against author keywords at the top-15 level.

It should be also added here, that among the keyphrases regarded as false positives in our evaluations, there were many near misses due to synonymy, e.g. *tiny keys* and *small keys* or *slow Web* and *slow WAP*. To overcome the synonymy issue to some extent the WordNet-based rewriting of tokens was introduced, which brought improvement in the case of the baseline systems for both domains (so it was employed in the later experiments as well). Another source of false positive classifications was due to the incompleteness of the opinion aspect entered by the user, i.e. not all the important aspects are necessarily listed among the

	Top-5			Top-10			Top-15		
	Prec.	Recall	F-score	Prec.	Recall	F-score	Prec.	Recall	F-score
∪	72.8	20.63	32.14	66.8	33.54	44.66	63.47	46.88	53.92
∩	46.4	27.81	34.77	41.6	44.92	43.2	37.07	56.68	44.82
Author	34.4	22.29	27.05	31.6	35.43	33.4	28.8	45.14	35.17

Table 3: Results of the human evaluation. ∪, ∩ and Author means when the automatic keyphrases were matched against the union, intersection of the keyphrases of three independent annotators and the keyphrases of the original author, respectively.

pros and cons section, as described earlier. On the other side, many of the author-entered keyphrases were absent in the contents of a review in their same form: only 34,8% and 23,9% of gold standard keyphrases could be found in the texts having the same normalized form for the mobile phone and the movie domains, respectively, setting an upper bound for the recall values when evaluating based on strict matching.

To overcome all the previously mentioned shortcoming during automatic evaluation, human evaluation was performed and it showed that real life application of opinion phrase extraction could be of much higher utility than strict evaluation would suggest. This is due to the fact that human annotators had access to common sense knowledge and during the inspection of keyphrases they could resolve such cases that were impossible during automatic evaluation.

All the features were effective in the sense that expanding the baseline feature set by them separately resulted in better results. Moreover, in the majority of the cases improvements were of high significance (see Table 2). The added value of Wikipedia features (that are likely to work well in other domains as well) should be highlighted as well as the relatively poor effect of keyphrase frequency feature which normally works better in the case of standard scientific keyphrase extraction tasks. A possible reason for keyphrase frequency feature not being that effective in the opinion domain is that in the case of opinionated keyphrases, the presence of such a phrase that was marked as positive in one document is not necessarily marked the same way in other documents, e.g. because one author may write about the feature objectively while the other may write his strong opinions about the very same feature, using similar wording.

5 Conclusions

In this paper, we presented a pros and cons extraction system by pointing out the parallelism between the keyphrases of scientific papers – given by their author – and the pros and cons phrases – given by product reviewers. The WordNet-based phrase normalization and an extended stopword-based filtration of keyphrase candidates introduced here could be of possible use for any kind of phrase extraction tasks. Besides demonstrating their similarity, the main differences of the two tasks were also highlighted, and several ways to adopt to the specialties of opinion phrase extraction have been suggested by introducing a rich feature set, some of which could also be widely used (e.g. Wikipedia-based ones), and others are specifically designed to the special task of opinion phrase extraction (e.g. SentiWordNet-related ones).

Among the most important differences of opinion phrase extraction from scientific keyphrase extraction we should note that for product reviews the pure occurrence of a single phrase is less deterministic to be a keyphrase, i.e. some emotional context is necessary to treat them as genuine ones. Also, the language of reviews is more special since it tends to contain elements that are not present in other genres of documents, such as irony and sarcasm and offers more possibility to express identical things in different ways. In total, our results are competitive with those of other standard keyphrase extraction tasks even when applying strict normalized form matching evaluation. Moreover, human evaluation showed that when semantics are involved into the evaluation, results are significantly better than it is suggested by automatic evaluations.

Acknowledgments

This work was supported by the Project "TÁMOP-4.2.1/B-09/1/KONV-2010-0005 – Creating the Center of Excellence at the University of Szeged", supported by the European Union and co-financed by the European Regional Development Fund and by the project BELAMI financed by the National Innovation Office of the Hungarian government.

References

S.R.K. Branavan, Harr Chen, Jacob Eisenstein, and Regina Barzilay. 2008. Learning document-level semantic properties from free-text annotations. In *Proceedings of ACL-08: HLT*, pages 263–271, Columbus, Ohio. ACL.

Andrea Esuli, Stefano Baccianella, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. ELRA.

Christiane Fellbaum, editor. 1998. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London.

Minqing Hu and Bing Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Minqing Hu and Bing Liu. 2004b. Mining opinion features in customer reviews. In *Proceedings of the 19th national conference on Artifical intelligence*, AAAI'04, pages 755–760. AAAI Press.

Soo-Min Kim and Eduard Hovy. 2006. Automatic identification of pro and con reasons in online reviews. In *Proceedings of the COLING/ACL 2006 Main Conference Poster Sessions*, pages 483–490, Sydney, Australia. ACL.

Su Nam Kim, Olena Medelyan, Min-Yen Kan, and Timothy Baldwin. 2010. Semeval-2010 task 5: Automatic keyphrase extraction from scientific articles. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, SemEval '10, pages 21–26, Morristown, NJ, USA. ACL.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st ACL*, pages 423–430.

S. le Cessie and J.C. van Houwelingen. 1992. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191–201.

Jingjing Liu and Stephanie Seneff. 2009. Review sentiment scoring via a parse-and-paraphrase paradigm. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 161–169, Singapore. ACL.

Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu.

Olena Medelyan, Eibe Frank, and Ian H. Witten. 2009. Human-competitive tagging using automatic keyphrase extraction. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1318–1327, Singapore. ACL.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86, Morristown, NJ, USA. ACL.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 339–346, Vancouver, British Columbia, Canada. ACL.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of CICLing-2002*, pages 1–15, Mexico City, Mexico.

Todd Sullivan. 2008. Pro, con, and affinity tagging of product reviews. Technical Report 224n, Stanford CS.

Ivan Titov and Ryan McDonald. 2008. A joint model of text and aspect ratings for sentiment summarization. In *Proceedings of ACL-08: HLT*, pages 308–316, Columbus, Ohio. ACL.

Kristina Toutanova and Christopher D. Manning. 2000. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora*, EMNLP '00, pages 63–70, Stroudsburg, PA, USA. ACL.

Peter Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424.

Peter Turney. 2003. Coherent keyphrase extraction via web mining. In *Proceedings of IJCAI*, pages 434–439.

Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig. 1999. Kea: Practical automatic keyphrase extraction. In *ACM DL*, pages 254–255.