Publications of the JRC's *Europe Media Monitor* team

Status March 2017

You may find more recent **EMM-Publications in the JRC's PUBSY system**

Forthcoming

- Steinberger Ralf, Martin Atkinson, Teófilo Garcia, Erik van der Goot, Jens Linge, Charles Macmillan, Hristo Tanev, Marco Verile & Gerhard Wagner (forthcoming). **EMM: Supporting the analyst by turning multilingual text into structured data**. Chapter of the book ‚Transparenz aus Verantwortung: Neue Herausforderungen für die digitale Datenanalyse'. Berlin, Erich Schmidt Verlag.
- Jacquet Guillaume, Maud Ehrmann, Jakub Piskorski, Hristo Tanev & Ralf Steinberger (forthcoming). **Cross-lingual linking of multi-word entities and language-dependent learning of multi-word entity patterns**. In Yannick Parmentier & Jakub Waszczuk (eds.), Representation and Parsing of Multiword Expressions, pp. 229-255. Berlin, Language Sciences Press.
- Erjavec T., J. Piskorski, L. Pivovarova, J. Šnajder, J. Steinberger & R. Yangarber (forthcoming). **Proceedings of the 6th Biennial Workshop on Balto-Slavic Natural Language Processing** (BSNLP 2017). Held at EACL 2017, Valencia, Spain, 4 April 2017.
- Piskorski Jakub, Lidia Pivovarova, Jan Šnajder, Josef Steinberger & Roman Yangarber (forthcoming). **The first cross-lingual challenge on recognition, normalization and matching of Named Entities in Slavic languages**. Proceedings of the 6th biannual workshop on Balto-Slavonic Natural Language Processing (BSNLP'2017).

2016

- ➔ Ehrmann Maud, Guillaume Jacquet & Ralf Steinberger (2017). JRC-Names: Multilingual Entity Name Variants and Titles as Linked Data. Semantic Web 8 (2), pp. 283-295.
- Jacquet Guillaume, Maud Ehrmann, Ralf Steinberger & Jaakko Väyrynen (2016). **Cross-lingual linking of multi-word entities and their corresponding acronyms**, LREC'2016, Portorož, Slovenia, 23-28 May 2016. (Read online)
- Balahur Alexandra (2016). **OPAL at SemEval Task 4: the Challenge of Porting a Sentiment Analysis System to the "Real" World**. Proceedings of SemEval-2016, pp. 262-265 (Read online)
- Balahur Alexandra & Hristo Tanev (2016). **Detecting Implicit Expressions of Affect from Text using Semantic Knowledge on Common Concept Properties**. Proceedings of LREC'2016, pp 1165-1170. (Read online)
- Riccardo Flavia, M. del Manso, M. Caporali, C. Napoli, J. Linge, E. Mantica, M. Verile, A. Piatta, M. Pompa, L. Vellucci, V. Costanzo, A. Bastiampillai, E. Gabrielli, M. Gramegna & S. Declich (2016). **Event-based surveillance during EXPO Milan 2015: rationale, tools, procedures and initial results**. Health Security, Vol. 14 No. 3 p. 161-172.

2015

- ➔ Steinberger Ralf, Aldo Podavini, Alexandra Balahur, Guillaume Jacquet, Hristo Tanev, Jens Linge, Martin Atkinson, Michele Chinosi, Vanni Zavarella, Yaniv Steiner & Erik van der Goot (2015). Observing Trends in Automated Multilingual Media Analysis. Proceedings of the Symposium on New Frontiers of Automated Content Analysis in the Social Sciences (ACA'2015), Zürich, Switzerland, 1-3 July 2015.
- Zavarella Vanni, Hristo Tanev, Ralf Steinberger & Erik van der Goot (2015). **An Ontology-Based Approach to Social Media Mining for Crisis Management**. CEUR Workshop Proceedings (CEUR-WS.org), pp. 55-66. (Read online).
- Nardo Michela & Erik van der Goot (2015). **Does web anticipate stocks? Analysis for a subset of systemically important banks**. EUR - Scientific and Technical Research Report EUR 27023. (Read online)
- Balahur Alexandra & Perea Ortega José Manuel (2015). **Sentiment Analysis System Adaptation for Multilingual Processing: the Case of Tweets**. Information Processing & Management 51:4 (2015), pp 547-556. (Read online)
- Alomar Oscar; A Battle, JM Brunetti, Roberto Garcia, R Gil; A Granollers, S Jimenez, A Lavina, Jens Linge, Marco Pautasso, C Reverte, J Riuvadets, Agnès Rortais, Giuseppe Stancanelli, S Volani & Sybren Vos (2015).

Development and testing of the media monitoring tool MedISys for early identification and reporting of existing and emerging plant health threats. EPPO Bulletin 45:2, pp. 288-293. ([Read online](#))

- Caselli Tommaso, Marieke Van Erp, Anne-Lyse Minard, Mark Finlayson, Ben Miller, Jordi Atserias; Alexandra Balahur & Piek Vossen (2015). **Proceedings of the First Workshop on Computing News Storylines** (CNewsStory 2015). ([Read online](#))
- Balahur Alexandra, Erik van der Goot, Piek Vossen & Andrés Montoyo (2015). **EMNLP2015 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis WASSA 2015 – Proceedings**. ([Read online](#))

2014

- Steinberger Ralf, Guillaume Jacquet & Leonida Della Rocca (2014). **Creation and Use of Multilingual Named Entity Variant Dictionaries**. In: Marc Van Campenhoudt, Nathalie Lemaire & Rita Temmerman (eds):Traduire aux confins du lexique : les nouveaux terrains de la terminologie (Translation at the Frontiers of the Lexicon: The New Fields of Terminology). Cahiers de Linguistique 2014-40/2, pp. 113-132. Éditions Modulaires Européennes, Fernelmont, Belgium. ([Read online](#))
➔ Steinberger Ralf, Mohamed Ebrahim, Alexandros Poulis, Manuel Carrasco-Benitez, Patrick Schlüter, Marek Przybyszewski & Signe Gilbro (2014). [An overview of the European Union's highly multilingual parallel corpora](#). Language Resources and Evaluation Journal (LRE). DOI: 10.1007/s10579-014-9277-0.
- Steinberger Josef, Ralf Steinberger, Hristo Tanev, Vanni Zavarella & Marco Turchi (2014). **Aspects of Multilingual News Summarisation**. In: Alessandro Fiori (ed.): Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding, pp. 277-294. IGI Global Information Science Reference Series, Hershey, PA, USA. ([Purchase online](#))
- Balahur Alexandra, Rada Mihalcea & Andrés Montoyo (2014). **Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications - Preface**. Computer Speech & Language 28(1), pp. 1-6. Elsevier.
- Balahur Alexandra & Marco Turchi (2013). **Comparative Experiments using supervised learning and machine translation for multilingual sentiment analysis**. Computer Speech & Language 28(1), pp. 56-75. Elsevier.
- Balahur Alexandra, Erik van der Goot, Ralf Steinberger & Andrés Montoyo (eds.) (2014). Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA'2014), held at ACL'2014. Baltimore (MD), USA, 27 June. ISBN: 978-1-941643-11-2. ([Read online](#))
- Pajzs Júlia, Ralf Steinberger, Maud Ehrmann, Mohamed Ebrahim, Leonida Della Rocca, Stefano Bucci, Eszter Simon, Tamás Váradi (2014). **Media Monitoring and Information Extraction for the Highly Inflected Agglutinative Language Hungarian**. Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014, pp. 2049-2056. ([Read online](#))
➔ Hajlaoui Najeh, David Kolovratnik, Jaakko Väyrynen, Daniel Várga, Ralf Steinberger (2014). [DCEP - Digital Corpus of the European Parliament](#). Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014, pp. 3164-3171.
- Küçük Dilek, Guillaume Jacquet, Ralf Steinberger (2014). **Named entity recognition on Turkish Tweets**. Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014, pp. 450-454. ([Read online](#))
- Jacquet Guillaume, Maud Ehrmann, Ralf Steinberger (2014). **Clustering of multi-word named entity variants: Multilingual evaluation**. Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014, pp. 2548-2553. ([Read online](#))
- Balahur Alexandra, Marco Turchi, Ralf Steinberger, José M. Perea-Ortega, Guillaume Jacquet, Dilek Küçük, Vanni Zavarella, Adil El Ghali (2014). **Resource creation and evaluation for multilingual sentiment analysis in social media texts**. Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC), Reykjavik, Iceland, 26-31 May 2014, pp. 4265-4269. ([Read online](#))
- Küçük Dilek & Ralf Steinberger (2014). **Experiments to Improve Named Entity Recognition on Turkish Tweets**. Proceedings of the EACL'2014 workshop *Language Analysis in Social Media (LASM)*, pp. 71-78. Gothenburg, Sweden, 27 April 2014. ([Read online](#))

- Zavarella Vanni, Hristo Tanev, Ralf Steinberger & Erik van der Goot (2014). **An ontology-based approach to social media mining for crisis management**. Proceedings of the Workshop on Social Media and Linked Data for Emergency Response (SMILE'2014). Anissaras, Crete, Greece, 25-29 May 2014. ([Read online](#))
- Zavarella Vanni, Dilek Küçük, Hristo Tanev & Ali Hürriyetoğlu (2014). **Event extraction for Balkan languages.** 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL'2014), pp. 65-68. ([Read online](#))
- El Ghali Adil & El Ghali Kaoutar (2014). **Combiner espaces sémantiques, structure et contraintes.** Actes de l'atelier Défi Fouille de Textes (DEFT'2014). ([Read online](#))
- Flavia Riccardo, Mika Shigematsu, Catherine Chow, Jens Linge, Brian Doherty, Maria Grazia Dente, Silvia Declich, Mike Barker, Philippe Barboza, Laetitia Vaillant, Alastair Donachie, Abla Mawudeku, Ray Arthur & Jason McKnight (2014). **Interfacing a biosurveillance portal and an international network of institutional analysts to detect biological threats.** Biosecurity and Bioterrorism-Biodefense Strategy Practice and Science 12:6, pp. 325-336. ([Read online](#))
- Barboza Philippe, Laetitia Vaillant, David Hartley, Noele Nelson, Abla Mawudeku, Larry Madoff, Jens Linge, Nigel Collier, John Brownstein & Pascal Astagneau (2014). **Factors Influencing Performance of Internet-Based Biosurveillance Systems Used in Epidemic Intelligence for Early Detection of Infectious Diseases Outbreaks.** PLOS One 9:3, e90536. ([Read online](#))
- Consoli Sergio, Diego Reforgiato Recupero & Vanni Zavarella (2014). **A survey on tidal analysis and forecasting methods for tsunami detection**. Science of Tsunami Hazards, Vol. 33 no. 1 pp. 1-56.

2013

- ➔ Steinberger Ralf, Maud Ehrmann, Júlia Pajzs, Mohamed Ebrahim, Josef Steinberger & Marco Turchi (2013). [Multilingual media monitoring and text analysis – Challenges for highly inflected languages](#). In: Ivan Habernal & Václav Matoušek (eds). Text, Speech and Dialogue. 16th International Conference, TSD 2013, Pilsen, Czech Republic, September 2013, Proceedings. Springer Lecture Notes in Artificial Intelligence LNAI 8082, pp. 22-33.
- Steinberger Ralf (2013). **Multilingual and cross-lingual news analysis in the Europe Media Monitor (EMM)**. In: Mihai Lupu, Evangelos Kanoulas & Fernando Loizides (eds.): *Multidisciplinary Information Retrieval.* 6th Information Retrieval Facility Conference (IRFC'2013), Limassol, Cyprus. Springer Lecture Notes in Computer Science, Vol. 8201, pp. 1-4. ([Purchase online](#))
- Kabadjov Mijail, Josef Steinberger & Ralf Steinberger (2013). **Multilingual Statistical News Summarization**. In: Thierry Poibeau, Horacio Saggion, Jakub Piskorski & Roman Yangarber (eds), Multi-source, Multilingual Information Extraction and Summarization, pp. 229-252. ISBN: 978-3-642-28569-1, DOI: 10.1007/978-3-642-28569-1_11, Springer, Berlin & Heidelberg, Germany. ([Purchase online](#))
- ➔ Ehrmann Maud, Leonida della Rocca, Ralf Steinberger & Hristo Tanev (2013). [Acronym recognition and processing in 22 languages](#)**.** Proceedings of the 9th Conference *Recent Advances in Natural Language Processing* (RANLP), pp. 237-244. Hissar, Bulgaria, 7-13 September 2013.
- Turchi Marco, Domenico Perrotta, Andrea Cerioli & Marco Riani (2013). **Robustness Issues in Text Mining**. In: Rudolf Kruse, Michael R. Berthold, Christian Moewes, María Ángeles Gil, Przemyslaw Grzegorzewski, Olgierd Hryniewicz (eds). Synergies of Soft Computing and Statistics for Intelligent Data Analysis. Berlin Heidelberg (Germany): Springer Verlag; 2013. p. 263-272. ([Purchase online](#))
- Piskorski Jakub & Maud Ehrmann. **On Named Entity Recognition in Targeted Twitter Streams in Polish**. In: Proceedings of the 4th Biennial Workshop on Balto-Slavic Natural Language Processing (BSNLP), collocated with ACL 2013, pp. 84-93.
- Piskorski Jakub, Lidia Pivovarova, Hristo Tanev & Roman Yangarber (eds). **Proceedings of the 4th Biennial Workshop on Balto-Slavic Natural Language Processing** (BSNLP 2013). Held at ACL 2013. Sofia, Bulgaria, 8-9 August 2013. Association for Computational Linguistics, 2013.
- Piskorski Jakub, Hristo Tanev & Alexandra Balahur. **Exploiting Twitter for Border Security-Related Intelligence Gathering**. In IEEE Proceedings of the 3rd European Intelligence and Security Informatics Conference (EISIC 2013), pp. 239-246, Uppsala, Sweden, 2013.

- Atkinson Martin, Jakub Piskorski, Hristo Tanev, Roman Yangarber & Vanni Zavarella. **Techniques for Multilingual Security-related Event Extraction from Online News**. In: Przepiórkowski Adam et al. *Computational Linguistics Applications*, pp. 163-186. Springer-Verlag, Berlin, 2013. (Purchase online)
- Balahur Alexandra, Jesús Hermida, Andrés Montoyo & Rafael Muñoz (2013). **Detecting implicit expressions of affect in text using EmotiNet and its extensions**. Data & Knowledge Engineering 88, pp. 113-125. Elsevier.
- Balahur Alexandra (2013). **Sentiment analysis in social media texts**. Proceedings of the 4th workshop on Computational approaches to subjectivity, sentiment and social media analysis at ACL'2013, pp. 120-128.
- Balahur Alexandra & Hristo Tanev (2013). **Detecting event-related links and sentiments from social media texts**. Proceedings of the Conference of the Association for Computational Linguistics (ACL'2013).
- Balahur Alexandra, Jesús Hermida & Hristo Tanev (2013): **Detecting implicit emotion expressions from text using ontological resources and lexical learning**. In A. Altamari et al. (eds). New Trends of Research in Ontologies and Lexical Resources. Theory and Applications of Natural Language Processing. Springer Verlag, Berlin Heidelberg.
- Balahur Alexandra & José-Manuel Perea-Ortega (2013). **Experiments using varying sizes and machine translated data for sentiment analysis in Twitter**. Proceedings of the TASS workshop at SEPLN 2013. IV Congreso Español de Informática, pp. 17-20.
- Van der Goot Erik, Hristo Tanev & Jens Linge (2013). **Combining twitter and media reports on public health events in MedISys**. Proceedings of the 22nd international conference on World Wide Web companion, pp. 703-718. International World Wide Web Conferences Steering Committee, 2013.
- Tanev Hristo & Josef Steinberger (2013). **Semi-automatic acquisition of lexical resources and grammars for event extraction in Bulgarian and Czech**. Proceedings of the 4th Biennial International Workshop on Balto-Slavic Natural Language Processing, held at ACL'2013, pp. 110-118.
- Barboza P, Vaillant L, Mawudeku A, Nelson NP, Hartley DM, Madoff LC, Linge JP, Collier N, Brownstein JS, Yangarber R, Astagneau P (2013). **Early Alerting Reporting Project Of The Global Health Security Initiative. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events**. PLoS One. 2013;8(3):e57252. DOI: 10.1371/journal.pone.0057252. Epub 2013 Mar 5.
- Hartley DM, Nelson NP, Arthur RR, Barboza P, Collier N, Lightfoot N, Linge JP, van der Goot E, Mawudeku A, Madoff LC, Vaillant L, Walters R, Yangarber R, Mantero J, Corley CD &Brownstein JS (2013). **An overview of Internet biosurveillance**. Clin Microbiol Infect. 2013 May 24. DOI: 10.1111/1469-0691.12273.
- Zavarella Vanni (2013). **FSS-TimEx for TempEval-3: Extracting Temporal Information from Text.** Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 58-63.
- Cojazzi Giacomo, Erik van der Goot, Marco Verile, Erik Wolfart, Marcy Rutan Fowler, Yana Feldman, William Hammond, John Schweighardt & Matthew Ferguson (2013). **Collection And Analysis Of Open Source News For Information Awareness And Early Warning in Nuclear Safeguards.** ESARDA Bulletin 50, pp. 94-105.

2012

- ➔ Steinberger Ralf (2012). A survey of methods to ease the development of highly multilingual Text Mining applications. Language Resources and Evaluation Journal, Springer, Volume 46, Issue 2, pp. 155-176 (DOI 10.1007/s10579-011-9165-9).
- ➔ Turchi Marco, Martin Atkinson, Alastair Wilcox, Brett Crawley, Stefano Bucci, Ralf Steinberger & Erik van der Goot (2012). ONTS: "OPTIMA" News Translation System. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL), pp. 25–30, Avignon, France, April 23 - 27 2012.
- ➔ Ebrahim Mohamed, Maud Ehrmann, Marco Turchi & Ralf Steinberger (2012). Multi-label EuroVoc classification for Eastern and Southern EU Languages. In: Cristina Vertan & Walther v. Hahn: Multilingual processing in Eastern and Southern EU languages - Low-resourced technologies and translation, pp. 370-394. Cambridge Scholars Publishing, Cambridge, UK.

➔ Steinberger Ralf, Andreas Eisele, Szymon Klocek, Spyridon Pilos & Patrick Schlüter (2012). DGT-TM: A freely Available Translation Memory in 22 Languages. Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012) pp. 454-459, Istanbul, 21-27 May 2012.

➔ Steinberger Ralf, Mohamed Ebrahim & Marco Turchi (2012). JRC EuroVoc Indexer JEX - A freely available multi-label categorisation tool. Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012), pp. 798-805, Istanbul, 21-27 May 2012.

• Steinberger Ralf (2012). **Het nut van de enorme meertalige tekstcollecties van de EU om taaltechnologische oplossingen voor alle EU-talen te bouwen** (English: Exploiting the EU's enormous multilingual text collections to build Language Technology solutions for all EU languages). In: Journal DIXIT - Tijdschrift over Taal- en Spraaktechnologie, Issue 'Big Data en TST', December 2012, pp. 30-31, Nijmegen, Netherlands. (Read online)

• Tanev Hristo, Maud Ehrmann, Jakub Piskorski & Vanni Zavarella (2012). **Enhancing Event Descriptions through Twitter Mining**. In: AAAI Publications, Sixth International AAAI Conference on Weblogs and Social Media, pp 587-590. Dublin, June 2012. (Read online)

• Balahur-Dobrescu Alexandra & Hristo Tanev (2012). **Detecting Entity-Related Events and Sentiments from Tweets Using Multilingual Resources**. In: CLEF 2012: Conference and Labs of the Evaluation Forum Information Access Evaluation meets Multilinguality, Multimodality, and Visual Analytics; 17 September 2012; Rome (Italy). CLEF online working notes.

• Balahur Alexandra & Marco Turchi (2012). **Multilingual Sentiment Analysis using Machine Translation?**. Proceedings of the Association for Computational Linguistics: 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 52-60, Jeju, Republic of Korea, 12 July 2012. (Read online)

• Zavarella Vanni, Jakub Piskorski, Ana Esteves & Stefano Bucci. **Refining Border Security News Event Geotagging through Deployment of Lexico-Semantic Patterns**. In IEEE Proceedings of the European Intelligence and Security Informatics Conference (EISIC 2012) Odense, Denmark, 2012.

• Geierhos Michaela &Mohamed Ebrahim (2012). **Customer Interaction Management Goes Social : Getting Business Processes plugged in Social Networks**. In: Ajith Abraham, Aboul-Ella Hassanien (eds): Computational Social Networks: Tools, Perspectives and Applications. London (UK): Springer Verlag; 2012. p. 367-389.

• Al-Jumaily Harith, Paloma Martínez, José Martínez-Fernández & Erik van der Goot (2012). **A real time Named Entity Recognition system for Arabic text mining**. Language Resources and Evaluation, Springer, Volume 46, Issue 4, pp. 543-563. (Purchase online)

• Balahur Alexandra, Mijail Kabadjov, Josef Steinberger, Ralf Steinberger & Andrés Montoyo (2012). **Challenges and solutions in the opinion summarization of user-generated content**. Journal of Intelligent Information Systems (JIIS) 39 (2); 2012, pp. 375-398, Springer. DOI: 10.1007/s10844-011-0194-z. (purchase online)

➔ Steinberger Josef, Mohamed Ebrahim, Maud Ehrmann, Ali Hürriyetoğlu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vázquez, & Vanni Zavarella (2012). Creating Sentiment Dictionaries via Triangulation. Journal Decision Support Systems 53 (2012), pp. 689-694, Elsevier, Amsterdam.

• Steinberger Josef & Marco Turchi (2012). **Machine Translation for Multilingual Summary Content Evaluation**. Proceedings of the Workshop on Evaluation Metrics and System Comparison for Automatic Summarization, pp. 19-27. NAACL, Montreal, Canada, 3-8 June 2012. (Read online)

• Turchi Marco, Tijl De Bie & Nello Cristianini (2012). **An intelligent web agent that autonomously learns ow to translate**. Web Intelligence and Agent Systems (WIAS) 10 (2), 2012, pp. 165-178.

• Turchi Marco, Lucia Specia & Josef Steinberger (2012). **Relevance Ranking for Translated Texts**. Proceedings of 16th Annual Conference of the European Association for Machine Translation, Trento, Italy 28-30 May 2012.

• Turchi Marco, Cyril Goutte and Nello Cristianini (2012). **Learning Machine Translation from In-domain and Out-of-domain Data**. Proceedings of 16th Annual Conference of the European Association for Machine Translation, Trento, Italy, 28-30 May 2012.

• Turchi Marco, Tijl De Bie, Cyril Goutte & Nello Cristianini (2012). **Learning to translate: a statistical and computational analysis**. Advances in Artificial Intelligence. Hindawi Publishing Corporation, New York, USA.

- Bouvier d'Yvoire M, Bremer S, Casati S, Ceridono M, Coecke S, Corvi R, Eskes C, Gribaldo L, Griesinger C, Knaut H, Linge JP, Roi A, Zuang V. (2012). **ECVAM and new technologies for toxicity testing**. Adv Exp Med Biol. 2012;745:154-80.
- Linge Jens, Marco Verile, Hristo Tanev, Vanni Zavarella, Flavio Fuart, Erik van der Goot (2012). **Media monitoring of public health threats with MedISys**. Living in Surveillance Societies. ([Read online](#))
- Steinberger Ralf (2012). **Cross-lingual similarity calculation for plagiarism detection and more - Tools and Resources**. In: Pamela Forner, Jussi Karlgren & Christa Womser-Hacker (eds): CLEF 2012 Evaluation Labs and Workshop. Abstracts - Working Notes Papers, p.81. FBK Press, Trento, Italy. ([Read online](#))
- Mantero Jas, Edit Szegedi, Lara Payne Hallstrom, Evelyn Depoortere, Bernard Kaic, Lucille Blumberg, Jens Linge & Denis Coulombier (2012). **Enhanced epidemic intelligence using a web-based screening system during the 2010 FIFA World cup in South Africa.** Eurosurveillance 19:18, pp. 1-9.

2011

- ➔ Steinberger Ralf, Sylvia Ombuya, Mijail Kabadjov, Bruno Pouliquen, Leonida Della Rocca, Jenya Belyaeva, Monica De Paola & Erik van der Goot (2011). [Expanding a multilingual media monitoring and information extraction tool to a new language: Swahili](#). Language Resources and Evaluation Journal, Volume 45, Issue 3, pp. 311-330.
- ➔ Steinberger Ralf, Bruno Pouliquen, Mijail Kabadjov & Erik van der Goot (2011). [JRC-Names: A freely available, highly multilingual named entity resource](#). Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011), pp. 104-110. Hissar, Bulgaria, 12-14 September 2011.
- Turchi Marco & Maud Ehrmann (2011). **Knowledge Expansion of a Statistical Machine Translation System using Morphological Resources**. Proceedings of the 12th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'2011). Tokyo, Japan, 20-26 February 2011. ([Read online](#))
- Steinberger Josef, Hristo Tanev, Mijail Kabadjov & Ralf Steinberger (2011). **Aspect-Driven News Summarization**. In: International Journal of Computational Linguistics and Applications 2 (1-2), pp 301-317, ISSN: 0976-0962,Bahri Publications. 2011. ([Purchase online](#))
- Atkinson Martin, Jakub Piskorski, Erik van der Goot & Roman Yangarber (2011). **Multilingual Real-Time Event Extraction for Border Security Intelligence Gathering**. In: U. Kock Wiil (ed.) Counterterrorism and Open Source Intelligence. Springer Lecture Notes in Social Networks, Vol. 2, 1st Edition, 2011, ISBN: 978-3-7091-0387-6, pp 355-390.
- Atkinson Martin & Jakub Piskorski (2011). **Frontex Real-time News Event Extraction Framework**. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD-2011), pp. 749-752. San Diego, USA, 2011.
- Jakub Piskorski, Hristo Tanev, Martin Atkinson, Erik van der Goot & Vanni Zavarella (2011). **Online News Event Extraction for Global Crisis Surveillance**. Transactions on Computational Collective Intelligence. Springer Lecture Notes in Computer Science LNCS 6910/2011, pp. 182-212. ([purchase online](#))
- Ehrmann Maud, Marco Turchi & Ralf Steinberger (2011). Building a Multilingual Named Entity-Annotated Corpus Using Annotation Projection. Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011), 118-124. Hissar, Bulgaria, 12-14 September 2011. ([Read online](#))
- Steinberger Josef, Mijail Kabadjov, Ralf Steinberger, Hristo Tanev, Marco Turchi & Vanni Zavarella (2011). **Towards language-independent news summarization**. Proceedings of the Text Analysis Conference 2011 (TAC'2011). National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, USA, 14-15 November 2011.
- ➔ Steinberger Josef, Polina Lenkova, Mijail Kabadjov, Ralf Steinberger & Erik van der Goot (2011). [Multilingual Entity-Centered Sentiment Analysis Evaluated by Parallel Corpora](#). Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011), pp. 770-775. Hissar, Bulgaria, 12-14 September 2011.
- ➔ Steinberger Josef, Jenya Belyaeva, Jonathan Crawley, Leonida Della Rocca, Mohamed Ebrahim, Maud Ehrmann, Mijail Kabadjov, Ralf Steinberger & Erik van der Goot (2011). **Highly Multilingual Coreference Resolution Exploiting a Mature Entity Repository**. Proceedings of the 8th International Conference Recent

Advances in Natural Language Processing (RANLP'2011), pp. 254-260. Hissar, Bulgaria, 12-14 September 2011.

➔ Piskorski Jakub, Jenya Belayeva & Martin Atkinson (2011). [On Refining Real-time News Event Extraction through Deployment of Cross-lingual Information Fusion Techniques](). IEEE Proceedings of the European Intelligence and Security Informatics Conference (EISIC) 2011, pp. 38-45. Athens, Greece.

- Piskorski Jakub, Jenya Belyaeva & Martin Atkinson (2011). **Exploring the usefulness of cross-lingual information fusion for refining real-time news event extraction**. Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011), pp. 210-217. Hissar, Bulgaria, 12-14 September 2011.

- Turchi Marco, Vanni Zavarella & Hristo Tanev (2011). **Pattern Learning for Event Extraction using Monolingual Statistical Machine Translation**. Proceedings of the 8th International Conference Recent Advances in Natural Language Processing (RANLP'2011), pp 371-377. Hissar, Bulgaria, 12-14 September 2011. ([Read online]())

- Steinberger Ralf (2011). **Combining various text analysis tools for multilingual media monitoring**. In: Hamburg Working Paper in Multilingualism 96-2011. In: Hanna Hedeland, Thomas Schmidt, Kai Wörner (eds.). Multilingual Resources and Multilingual Applications. Proceedings of the Conference of the German Society for Computational Linguistics and Language Technology (GSCL'2011), pp. 25-30. Hamburg, Germany, 28-30 September 2011.

- Steinberger Josef, Polina Lenkova, Mohamed Ebrahim, Maud Ehrmann, Silvia Vázquez, Ali Hürriyetoğlu, Mijail Kabadjov, Ralf Steinberger, Hristo Tanev & Vanni Zavarella (2011). **Creating Sentiment Dictionaries via Triangulation**. Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, WASSA, held at the ACL-HLT Conference, pp. 28-36. Portland, Oregon, USA, 24 June 2011.

- Linge J, Belyaeva J. (2011). **Customization of MedISys for the monitoring of food and feed hazards**. European Commission, PUBSY JRC63973.

- Linge, J.P., Mantero, J. Fuart, F., Belyaeva, J., Atkinson, M., van der Goot, E. (2011). **Tracking Media Reports on the Shiga toxin-producing Escherichia coli O104:H4 outbreak in Germany**. In: Malaga. P. Kostkova, M. Szomszor, and D. Fowler (eds.), Proceedings of eHealth conference (eHealth 2011), LNICST 91, pp. 178–185, 2012.

- Mantero J, Belyaeva E (2011). **How to maximise event-based surveillance web-systems: the example of ECDC/JRC collaboration to improve the performance of MedISys**. Ed: Jens Linge. EUR 24763 EN. Luxembourg (Luxembourg): Publications Office of the European Union; 2011. PUBSY JRC63805.

- Denecke K, Kirchner G, Dolog P, Smrz P, Linge J, Backfried G, Dreesman J. (2011). **Event-driven architecture for health event detection from multiple sources**. Stud Health Technol Inform. 169. pp. 160-164.

- Tomeh N., M. Turchi, G. Wisniewski, A. Allauzen & F. Yvon (2011). **How Good Are Your Phrases? Assessing Phrase Quality with Single Class Classification**. Proceedings of IWSLT 2011: International Workshop on Spoken Language Translation, San Francisco, USA December 8-9, 2011.

- Flaounas I., O. Ali, M. Turchi, T. Snowsill, F. Nicart, T. De Bie & N. Cristianini (2011). **NOAM: News Outlets Analysis and Monitoring System**. Proceedings of the ACM SIGMOD/PODS Conference, Athens, Greece June 12-16, 2011.

- N. Tomeh, M. Turchi, G. Wisniewski, A. Allauzen and F. Yvon (2011). **How Good Are Your Phrases? Assessing Phrase Quality with Single Class Classification**. Proceedings of IWSLT 2011: International Workshop on Spoken Language Translation, San Francisco, USA December 8-9, 2011.

2010

- Tanev Hristo, Bruno Pouliquen, Vanni Zavarella & Ralf Steinberger (2010). **Automatic Expansion of a Social Network Using Sentiment Analysis**. In: Nasrullah Memon, Jennifer Jie Xu, David Hicks & Hsinchun Chen (eds). *Annals of Information Systems, Volume 12. Special Issue on Data Mining for Social Network Data, pp. 9-29.* Springer Science and Business Media (DOI 10.1007/978-1-4419-6287-4_2). ([Purchase online]())

➔ Turchi Marco, Josef Steinberger, Mijail Kabadjov & Ralf Steinberger (2010). [Using Parallel Corpora for Multilingual (Multi-Document) Summarisation Evaluation](). Multilingual and Multimodal Information Access Evaluation. Springer Lecture Notes for Computer Science, LNCS 6360/2010, pp. 52-63.

- Linge Jens, Ralf Steinberger, Flavio Fuart, Stefano Bucci, Jenya Belyaeva, Monica Gemo, Delilah Al-Khudhairy, Roman Yangarber & Erik van der Goot (2010). **MediSys: Medical Information System**. In: Eleana Asimakopoulou & Nik Bessis (eds). Advanced ICTs for Disaster Management and Threat Detection: Collaborative and Distributed Frameworks, pp. 131-142. IGI Global. (Purchase online)
- Steinberger Josef, Marco Turchi, Mijail Kabadjov, Nello Cristianini & Ralf Steinberger (2010). **Wrapping up a Summary: from Representation to Generation**. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL'2010), pp. 382-386. Uppsala, Sweden, 11-16 July. (Read online)
- Steinberger Ralf (2010). **Challenges and Methods for Multilingual Text Mining**. In: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010). Valletta, Malta, 19-21 May 2010. (Read online)
- Specia Lucia, Dhwaj Raj & Marco Turchi (2010). **Machine Translation evaluation versus quality estimation**. Machine Translation Journal Vol. 24, Nb. 1, pp. 39-50. (Purchase online)
- Flaounas Ilias, Marco Turchi, Omar Ali, Nick Fyson, Tijl De Bie, Nick Mosdell, Justin Lewis & Nello Cristianini (2010). **The structure of the EU mediasphere**. PLoS ONE 5(12): e14243. doi:10.1371/journal.pone.0014243. (Read online)
- Kabadjov Mijail, Martin Atkinson, Josef Steinberger, Ralf Steinberger & Erik van der Goot (2010). **NewsGist: A Multilingual Statistical News Summarizer**. In: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'2010). Barcelona, Spain, 20-24 September 2010. In: José Luis Balcázar, Francesco Bonchi, Aristides Gionis and Michèle Sebag (eds): Lecture Notes in Computer Science, Vol. 6323, pp. 591-594. Springer. (Purchase Online)
- Ehrmann Maud & Marco Turchi (2010). **Building Multilingual Named Entity Annotated Corpora Exploiting Parallel Corpora**. Proceedings of the Workshop on Annotation and Exploitation of Parallel Corpora (AEPC), pp. 24-33, Tartu, Estonia, 2 December 2010.
- ➔ Balahur Alexandra, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, Bruno Pouliquen & Jenya Belyaeva (2010). Sentiment Analysis in the News. In: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010), pp. 2216-2220. Valletta, Malta, 19-21 May 2010.
- Zaghouani Wajdi, Bruno Pouliquen, Mohamed Ebrahim & Ralf Steinberger (2010). Adapting a resource-light highly multilingual Named Entity Recognition system to Arabic. In: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'2010), pp. 563-567. Valletta, Malta, 19-21 May 2010. (Read online)
- Kabadjov Mijail, Josef Steinberger, Ralf Steinberger, Massimo Poesio & Bruno Pouliquen (2010). **Enhancing N-Gram-based Summary Evaluation Using Information Content and a Taxonomy**. In: Proceedings of the 32nd European Conference on Information Retrieval (ECIR'2010). Milton Keynes, UK, 28-31 March 2010. In: C. Gurrin et al. (eds): Lecture Notes in Computer Science, Vol. 5993, pp. 662-666. Springer. (Purchase Online)
- Zavarella Vanni, Hristo Tanev, Jens Linge, Jakub Piskorski, Martin Atkinson & Ralf Steinberger (2010). **Exploiting Multilingual Grammars and Machine Learning Techniques to Build an Event Extraction System for Portuguese**. In: Proceedings of the International Conference on Computational Processing of Portuguese Language (PROPOR'2010), Porto Alegre, Brazil, 27-30 April 2010. Springer Lecture Notes for Artificial Intelligence, Vol. 6001, pp. 21-24. Springer. (Purchase Online)
- Balahur Alexandra, Mijail Kabadjov & Josef Steinberger (2010). **Exploiting Higher-level Semantic Information for the Opinion-oriented Summarization of Blogs**. International Journal of Computational Linguistics and Applications (IJCLA), Vol., No. 1-2, Jan-Dec 2010, pp. 45-59.
- Tanev Hristo, Mijail Kabadjov & Monica Gemo (2010). **Learning Event Semantics from Online News**. In: Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'2010). Iaşi, Romania, 21-27 March 2010. International Journal of Computational Linguistics and Applications (IJCLA), Vol., No. 1-2, Jan-Dec 2010, pp. 27-43.
- Piskorski Jakub, Martin Atkinson, Jenya Belyaeva, Vanni Zavarella, Silja Huttunen, Roman Yangarber (2010). **Real-Time Text Mining in Multilingual News for the Creation of a Pre-frontier Intelligence Picture**. In Proceedings of the 16th Conference on Knowledge Discovery and Data Mining (KDD-2010) ACM SIGKDD Workshop on Intelligence and Security Informatics, Washington DC, USA.

- Atkinson Martin, Jenya Belayeva, Vanni Zavarella, Jakub Piskorski, S. Huttunen, A. Vihavainen, Roman Yangarber (2010). **News Mining for Border Security Intelligence**. In IEEE ISI-2010: Intelligence and Security Informatics, Vancouver, BC, Canada.
- Atkinson M, Keim D, Schaefer M, Franz W, Leitner-Fischer F, Zintgraf F. (2010). **DYNEVI - DYnamic News Entity VIsualization**. In: J. Kohlhammer, D. Keim (eds). Proceedings of the International Symposium on Visual Analytics Science and Technology. Golsar (Germany): The Eurographics Association. pp. 69-74.
- Atkinson Martin, Jakub Piskorski, Hristo Tanev, Erik van der Goot, Roman Yangarber, Vanni Zavarella (2010). **Automated Event Extraction in the Domain of Border Security**. In Proceedings of MINUCS-2009: Mining User-Generated Content for Security at the UCMedia-2009: ICST Conference on User-Centric Media, Venice, Italy.
- Al Khudhairy D, De Roo A, Annunziato A, Van Der Goot E, De Groeve T, Linge J. (2010). **EU Crisis Management: uitagingen en oplossingen**. Nationale Veiligheid en crisisbeheersing 9 (6); 2010. pp. 12-14.
- Rortais Agnès, Jenya Belyaeva, Monica Gemo, Erik van der Goot, Jens P. Linge (2010). **MedISys: An early-warning system for the detection of (re-)emerging food- and feed-borne hazards**. Food Research International Volume 43, Issue 5, June 2010, Pages 1553–1556.
- Hartley D, Nelson N, Walters R, Arthur R, Yangarber R, Madoff L, Linge J, Mawudeku A, Collier N, Brownstein J, Thinus G, Lightfoot N. (2010). **Landscape of international event-based biosurveillance**. Emerg Health Threats J. 2010;3:e3. Epub 2010 Feb 19.
- Krstajic, M.; Bak, P.; Oelke, D..; Atkinson, M.; Keim, D.A. (2010). **Applied Visual Exploration on Real-Time News Feeds Using Polarity and Geo-Spatial Analysis**. Web Information Systems and Technologies WEBIST 2010, Valencia, 7-10 April 2010.
- Krstajic, M.; Hansmanns, F.; Stoffel, A.; Atkinson, M.; Keim, D.A. (2010). **Processing online news streams for large-scale semantic analysis**. 26th International Conference on Data Engineering (ICDE) Workshops, pp.215-220, 1-6 March 2010.

2009

- ➔ Pouliquen Bruno & Ralf Steinberger (2009). Automatic Construction of Multilingual Name Dictionaries. In: Cyril Goutte, Nicola Cancedda, Marc Dymetman & George Foster (eds.): Learning Machine Translation. pp. 59-78. MIT Press - Advances in Neural Information Processing Systems Series (NIPS).
- ➔ Steinberger Ralf & Bruno Pouliquen (2009). Cross-lingual Named Entity Recognition. In: Satoshi Sekine & Elisabete Ranchhod (eds.): Named Entities - Recognition, Classification and Use, Benjamins Current Topics, Volume 19, pp. 137-164. John Benjamins Publishing Company. ISBN 978-90-272-8922 3.
- ➔ Steinberger Ralf, Bruno Pouliquen & Erik van der Goot (2009). An Introduction to the Europe Media Monitor Family of Applications. In: Fredric Gey, Noriko Kando & Jussi Karlgren (eds.): Information Access in a Multilingual World - Proceedings of the SIGIR 2009 Workshop (SIGIR-CLIR'2009), pp. 1-8. Boston, USA. 23 July 2009.
- ➔ Tanev Hristo, Vanni Zavarella, Jens Linge, Mijail Kabadjov, Jakub Piskorski, Martin Atkinson & Ralf Steinberger (2009). Exploiting Machine Learning Techniques to Build an Event Extraction System for Portuguese and Spanish. In: linguaMÁTICA Journal:2, pp. 55-66.
- Steinberger Josef, Mijail Kabadjov, Bruno Pouliquen, Ralf Steinberger & Massimo Poesio (2009). **WB-JRC-UT's Participation in TAC 2009: Update Summarization and AESOP Tasks**. In: Proceedings of the Text Analysis Conference 2009 (TAC'2009). National Institute of Standards and Technology, Gaithersburg, Maryland USA, 16-17 November 2009.
- Koehn Philipp, Alexandra Birch & Ralf Steinberger (2009). **462 Machine Translation Systems for Europe**. In: Laurie Gerber, Pierre Isabelle, Roland Kuhn, Nick Bemish, Mike Dillinger, Marie-Josée Goulet (eds.): Proceedings of the Twelfth Machine Translation Summit (MT-Summit XII), pages 65-72. Ottawa, Canada, 26-30 August 2009. (Read online)
- Steinberger Ralf (2009). **Preface**. In: Tadić Marco, Bojana Dalbelo Bašić, Marie-Francine Moens (eds.): Technologies for the Processing and Retrieval of Semi-Structured Documents - Experience from the CADIAL Project, pp. vii-ix. Croatian Language Technologies Society, Zagreb, Croatia. (Table-of-Contents)

- Atkinson Martin & Erik Van der Goot (2009). **Near Real Time Information Mining in Multilingual News**. Proceedings of the 18th International World Wide Web Conference (WWW'2009), pp. 1153-1154. Madrid, 20-24 April 2009. (Read online)
- Linge Jens, Ralf Steinberger, Thomas Weber, Roman Yangarber, Erik van der Goot, Delilah Al Khudhairy & Nikolaos Stilianakis (2009). **Internet Surveillance Systems for Early Alerting of Health Threats**. EuroSurveillance Vol. 14, Issue 13. Stockholm, 2 April 2009. (Read online)
- Balahur-Dobrescu Alexandra, Mijail Kabadjov, Josef Steinberger, Ralf Steinberger & Andrés Montoyo (2009). **Summarizing Opinions in Blog Threads**. Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC), pp. 606-613, Hong Kong, 3-5 December 2009.
- Steinberger Ralf (2009). **Linking News Content Across Languages**. In: Kristiina Jokinen & Eckhard Bick (eds.) NEALT Proceedings Series Vol.4 - Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA'2009), p. 4-5, Odense, Denmark, 14-16 May 2009.
- ➔ Balahur-Dobrescu Alexandra & Ralf Steinberger (2009). Rethinking sentiment analysis in the news: from theory to practice and back. 'Workshop on Opinion Mining and Sentiment Analysis' (WOMSA), held at the 2009 CAEPIA-TTIA 13th Conference of the Spanish Association for Artificial Intelligence, pp. 1-12. Sevilla, Spain, 13.11.2009.
- Balahur Alexandra, Ester Boldrini, Andrés Montoyo & Patricio Martínez-Barco (2009). **Opinion and Generic Question Answering Systems: a Performance Analysis**. Proceedings of the joint conference ACL-IJCNLP. Singapore, 2-7 August 2009.
- Balahur Alexandra, Elena Lloret, Ester Boldrini, Andrés Montoyo, Manuel Palomar & Patricio Martínez-Barco (2009). **Summarizing Threads in Blogs Using Opinion Polarity**. Proceedings of the RANLP Workshop 'Events in Emerging Text Types' (RANLP-eETTs), pp. 5-14. Borovets, Bulgaria, 17-18 September 2009.
- Kabadjov Mijail, Josef Steinberger, Bruno Pouliquen, Ralf Steinberger & Massimo Poesio (2009). **Multilingual Statistical News Summarisation: Preliminary Experiments with English**. Proceedings of the workshop 'Intelligent Analysis and Processing of Web News Content' (IAPWNC). Milano, Italy, 15.09.2009.
- Kabadjov Mijail, Alexandra Balahur-Dobrescu & Ester Boldrini (2009). **Sentiment Intensity: Is It a Good Summary Indicator?**. Proceedings of the 4th Language Technology Conference LTC, pp. 380-384. Poznan, Poland, 6-8.11.2009.
- ➔ Balahur Alexandra, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen & Mijail Kabadjov (2009). Opinion Mining on Newspaper Quotations. Proceedings of the workshop 'Intelligent Analysis and Processing of Web News Content' (IAPWNC), held at the 2009 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 523-526. Milano, Italy, 15.09.2009.
- Yangarber Roman, Peter von Etter & Ralf Steinberger (2009). **Automatic Epidemiological Surveillance from On-line News in MedISys and PULS**. Proceedings of the International Meeting on Emerging Diseases and Surveillance (IMED'2009), Vienna, 13-16 February 2009.
- Piskorski J., B. Watson, A. Yli–Jyrä (editors) (2009). **Post-proceedings of the Workshop on Finite-State Methods and Natural Language Processing 2008**. Book in the book series 'Frontiers in Artificial Intelligence and Applications', IOS Press, Amsterdam, The Netherlands, 2009.
- ➔ Piskorski J., K. Wieloch, M. Sydow (2009). On knowledge-poor methods for person name matching and lemmatization for highly inflectional languages. In: F. Lazarinis, J. Vilares, J. Tait, E. N. Efthimiadis (eds) Journal of Information Retrieval, Special Issue on non-English Web Retrieval, Springer, The Netherlands.
- Piskorski Jakub (2009). **Exploring Curvature-based Topic Development Analysis for Detecting Event Reporting Boundaries**. In: M. Marciniak, A. Mykowiecka (eds): Aspects of Natural Language Processing. Springer, Lecture Notes in Computer Science, Vol. 5070, pp. pp 311-331. (Purchase Online)
- ➔ Piskorski J., M. Sydow, K. Wieloch (2009). Comparison of String Distance Metrics for Lemmatisation of Named Entities in Polish. In: Z. Vetulani, H. Uszkoreit (eds): Selected and Extended Papers from the 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'2007), Poznan, Poland. Springer, Lecture Notes in Artificial Intelligence, Vol. 5603, pp. 413-427.
- Norguet Jean-Pierre, Esteban Zimányi & Ralf Steinberger (2009). **Semantic analysis of web site audience by integrating web usage mining and web content mining**. In I-Hsien Ting (editor): Web Mining Applications in E-commerce and E-services, Vol. 172/2009, pp. 65-80, Springer Verlag book series Studies in Computational Intelligence Series, Berlin/Heidelberg. (Purchase online)

2008

➔ Pouliquen Bruno, Hristo Tanev & Martin Atkinson (2008). Extracting and Learning Social Networks out of Multilingual News. Proceedings of the social networks and application tools workshop (SocNet-08) pp. 13-16. Skalica, Slovakia, 19-21 September 2008.

➔ Steinberger Ralf, Bruno Pouliquen & Camelia Ignat (2008). Using language-independent rules to achieve high multilinguality in Text Mining. In: Fogelman-Soulié Françoise, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.): Mining Massive Data Sets for Security. pp. 217-240. IOS Press, Amsterdam, The Netherlands.

➔ Steinberger Ralf, Flavio Fuart, Erik van der Goot, Clive Best, Peter von Etter & Roman Yangarber (2008). Text Mining from the Web for Medical Intelligence. In: Fogelman-Soulié Françoise, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.): Mining Massive Data Sets for Security. pp. 295-310. IOS Press, Amsterdam, The Netherlands.

• Fogelman-Soulié Françoise, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.) (2008): **Mining Massive Data Sets for Security**. IOS Press, Amsterdam, The Netherlands. (Purchase online)

• Tanev Hristo & Bernardo Magnini (2008). **Weakly supervised approaches for ontology population**. In: Paul Buitelaar & Philipp Cimiano (eds.): Ontology learning and population: Bridging the Gap between Text and Knowledge. IOS Press, Amsterdam, The Netherlands. Frontiers in Artificial Intelligence and Applications, Volume 167. (Purchase online)

• Tanev Hristo & Pinar Özden Wennerberg (2008). **Learning to Populate an Ontology of Politically Motivated Violent Events**. In: Domenico Perrotta, Jakub Piskorski, Françoise Soulié-Fogelman & Ralf Steinberger (eds.): Mining Massive Data Sets for Security. IOS Press, Amsterdam, The Netherlands.

• Pouliquen Bruno (2008). **Similarity of Names Across Scripts: Edit Distance Using Learned Costs of N-Grams**. In: Bengt Nordström & Aarne Ranta (eds.): *Advances in Natural Language Processing - 6th International Conference* (GoTal'2008), Lecture Notes in Artificial Intelligence 5221, Gothenburg, Sweden, 25-27 August 2008 Proceedings, pp. 405-416. (Purchase online)

• Piskorski Jakub, Hristo Tanev, Martin Atkinson & Erik van der Goot (2008). **Cluster-Centric Approach to News Event Extraction** . In: Proceedings of the International Conference on Multimedia & Network Information Systems Wroclaw, Poland, 18-19.09.2008. Publication by IOS Press, Amsterdam, The Netherlands.

• Zavarella Vanni, Hristo Tanev & Jakub Piskorski (2008). **Event Extraction for Italian Using a Cascade of Finite-State Grammars**. In: Proceedings of the International Workshop Finite-State Methods and Natural language Processing (FSMNLP'2008). Ispra, Italy, 11-12.09.2008.

• Steinberger Ralf, Flavio Fuart, Bruno Pouliquen & Erik van der Goot (2008). **MedISys: A Multilingual Media Monitoring Tool for Medical Intelligence and Early Warning**. In: Proceedings of the *International Disaster and Risk Conference* (IDRC'2008), pp. 612-614, Davos, Switzerland.

• Best Clive, Jakub Piskorski, Bruno Pouliquen, Ralf Steinberger & Hristo Tanev (2008). **Automatic Event Extraction for the Security Domain**. In: Intelligence and Security Informatics - Techniques and Applications, Volume 135/2008, pp. 17-43, Studies in Computational Intelligence Series, Springer, Heidelberg/New York. (Purchase online)

➔ Pouliquen Bruno, Ralf Steinberger & Olivier Deguernel (2008). Story tracking: linking similar news over time and across languages. In Proceedings of the 2nd workshop Multi-source Multilingual Information Extraction and Summarization (MMIES'2008) held at CoLing'2008. Manchester, UK, 23 August 2008.

• Atkinson Martin, Jakub Piskorski, Bruno Pouliquen, Ralf Steinberger, Hristo Tanev & Vanni Zavarella (2008). **Online-monitoring of security-related events**. In Proceedings of the 22nd International Conference on Computational Linguistics (CoLing'2008). Manchester, UK, 18-22 August 2008.

• Piskorski Jakub, Karol Wieloch, Mariusz Pikula & Marcin Sydow (2008). **Towards person name matching for highly inflective languages**. Proceedings of the WWW'2008 workshop on Natural Language Processing Challenges in the Information Explosion Era (NLPIX 2008). Beijing, April 2008.

➔ Tanev Hristo, Jakub Piskorski & Martin Atkinson (2008). Real-time News Event Extraction for Global Crisis Monitoring. In V. Sugumaran, M. Spiliopoulou, E. Kapetanios (editors) Proceedings of 13th International Conference on Applications of Natural Language to Information Systems (NLDB 2008), Lecture Notes in Computer Science, Cool. 5039, 24-27 June, London, UK.

- Yangarber Roman, Peter von Etter & Ralf Steinberger (2008). **Content Collection and Analysis in the Domain of Epidemiology**. In Proceedings of the 1st international MIE'2008 workshop on describing medical web resources (DRMed), held at the 21st International Congress of the European Federation for Medical Informatics. Göteborg, Sweden, 27 May 2008.
- Piskorski Jakub, Marcin Sydow & Dawid Weiss (2008). **Exploring Linguistic Features for Web Spam Detection: A Preliminary Study**. Proceedings of the WWW'2008 workshop on Adversarial Information Retrieval on the Web (AIRWEB 2008). Beijing, April 2008.
- Kübler Sandra, Jakub Piskorski & Adam Przepiórkowski (2008). **Proceedings of the LREC'2008 workshop on Partial Parsing: Between Chunking and Parsing (PAPA 2008)**. Marrakech, Morocco, 1 June, 2008.

2007

- ➔ Steinberger Ralf & Bruno Pouliquen (2007). [Cross-lingual Named Entity Recognition](). In: Satoshi Sekine & Elisabete Ranchhod (eds.), Journal Linguisticae Investigationes, Special Issue on Named Entity Recognition and Categorisation, LI 30:1, pp. 135-162. John Benjamins Publishing Company. ISSN 0378-4169.
- ➔ Piskorski Jakub (2007). [ExPRESS – Extraction Pattern Recognition Engine and Specification Suite](). In: Proceedings of the International Workshop Finite-State Methods and Natural language Processing 2007 (FSMNLP'2007). Potsdam, Germany, 14-16.09.2007.
- Tanev Hristo (2007). **Unsupervised Learning of Social Networks from a Multiple-Source News Corpus**. Proceedings of the Workshop *Multi-source Multilingual Information Extraction and Summarization* (MMIES'2007) held at RANLP'2007, pp. 33-40. Borovets, Bulgaria, 26 September 2007.
- Pouliquen Bruno, Ralf Steinberger, Jenya Belyaeva (2007). **Multilingual multi-document continuously updated social networks**. Proceedings of the Workshop *Multi-source Multilingual Information Extraction and Summarization* (MMIES'2007) held at RANLP'2007, pp. 25-32. Borovets, Bulgaria, 26 September 2007.
- Piskorski Jakub (2007). **On Some Aspects of Implementing a Pattern Engine based on Regular Expressions over Feature Structures**. In: Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'2007). Borovets, Bulgaria, 27-29.09.2007.
- Pouliquen Bruno & Ralf Steinberger (2007). **Acquisition and Use of Multilingual Name Dictionaries**. pp. 1-10. Proceedings of the Workshop *Acquisition and Management of Multilingual Lexicons* (AMML'2007) held at RANLP'2007. Borovets, Bulgaria, 26 September 2007.
- Piskorski Jakub & Marcin Sydow (2007). **Usability of String Distance Metrics for Name Matching Tasks in Polish**. In: Proceedings of the 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, (LTC'2007), Poznań, Poland, 5-7.10.2007.
- Mykowiecka Agnieszka, Anna Kupść, Małgorzata Marciniak & Jakub Piskorski (2007). **Resources for Information Extraction from Polish texts**. In: Proceedings of the 3rd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, (LTC'2007), Poznań, Poland, 5-7.10.2007.
- ➔ Pouliquen Bruno, Ralf Steinberger & Clive Best (2007). [Automatic Detection of Quotations in Multilingual News](). In: Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP'2007), pp. 487-492. Borovets, Bulgaria, 27-29.09.2007.
- Yangarber Roman, Clive Best, Peter von Etter, Flavio Fuart, David Horby & Ralf Steinberger (2007). **Combining Information about Epidemic Threats from Multiple Sources**. Proceedings of the Workshop *Multi-source Multilingual Information Extraction and Summarization* (MMIES'2007) held at RANLP'2007, pp. 41-48. Borovets, Bulgaria, 26 September 2007.
- Özden Wennerberg Pinar (2007). **Analyzing Social Networks in Online News Articles**. In: Norbert Gronau & Claudia Müller (eds.): Analyse sozialer Netzwerke und Social Software -Grundlagen und Anwendungsbeispiele, pp. 157-184. GITO-Verlag - Expertenwissen für die industrielle Praxis, Berlin.
- Piskorski Jakub, Hristo Tanev, Bruno Pouliquen & Ralf Steinberger (eds.) (2007). **Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007** (BSNLP'2007) - Special Theme: Information Extraction and Enabling Technologies. Held at the 45th Annual Meeting of the Association for Computational Linguistics (ACL'2007). Prague, Czech Republic, 29 June 2007.
- Piskorski Jakub, Marcin Sydow & Anna Kupść (2007). **Lemmatization of Polish Person Names**. In: Proceedings of the ACL Workshop on Balto-Slavonic Natural Language Processing 2007 - Special Theme:

Information Extraction and Enabling Technologies (BSNLP'2007). Held at ACL'2007. Prague, Czech Republic, 29 June 2007.

- Özden Wennerberg Pinar, Hristo Tanev, Jakub Piskorski & Clive Best (2007). **Ontology-based Analysis of Violent Events**. In: Proceedings of Intelligence and Security Informatics (ISI'2007). New Brunswick, New Jersey, USA, 23-24 May 2007.

- Piskorski Jakub, Hristo Tanev & Pinar Özden Wennerberg (2007). **Extracting Violent Events from On-line News for Ontology Population**. 10th International Conference on Business Information Systems (BIS'2007). Poznan, Poland, 25-27 April 2007. Lecture Notes in Computer Science, LNCS 4439, pages 287-300. Springer-Verlag, Berlin, Heidelberg, New York.

- Piskorski Jakub & Marcin Sydow (2007). **String Distance Metrics for Reference Matching and Search Query Correction**. 10th International Conference on Business Information Systems (BIS'2007). Poznan, Poland, 25-27 April 2007. Lecture Notes in Computer Science, LNCS 4439, pages 353-365. Springer-Verlag, Berlin, Heidelberg, New York.

- Ignat Camelia & François Rousselot. **Représentation de textes a l'aide d'étiquettes sémantiques dans le cadre de la classification automatique***. Romanian Review of Linguistics, VOL. LI, 2006, Issues 3-4, Ed. Romanian Academy, July-December.

- Versino Cristina, Camelia Ignat, Louis-Victor Bril (2007). **Open Source Information for Export Control.** Proceedings of the 29th ESARDA Symposium on Safeguards and Nuclear Material Management. page 1-8, OPOCE (publ.), Luxembourg, 2007.

2006

- Steinberger Ralf, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaž Erjavec, Dan Tufiş, Dániel Varga (2006). [The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages](#). Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006), pp. 2142-2147. Genoa, Italy, 24-26 May 2006.

- Daciuk Jan & Jakub Piskorski (2006). **Gazetteer Compression Technique based on Substructure Recognition**. In: Intelligent information processing and web mining, Proceedings of the International Conference on Intelligent Information Systems (IIS'2006), Ustroń, Poland. Publication in the Springer Verlag series: Advances in Soft Computing, 2006

- Piskorski Jakub & Marcin Sydow (2006). **Experiments on classification of Polish Newspaper Articles.**. Journal article in Archives of Control Sciences, Special issue on Human Language Technologies as a challenge for Computer Science and Linguistics Part II, special editor: Z. Vetulani, Volume 15, 2006

- Piskorski Jakub & Marcin Sydow (2006). **Fine-tuning N-gram-based Text Classifier for Highly Inflective Languages.**. In Proceedings of ICAISC 2006, Zakopane, Poland. Publication in: Challenging Problems of Computer Science, Artificial Intelligence and Soft Computing, editors: A. Cader, L. Rutkowski, R. Tadeusiewicz, J. Zurada, Academic Publishing House EXIT, Polish Neural Society, IEEE Computational Intelligence Society - Poland Chapter, Warsaw, 2006

- Pouliquen Bruno, Marco Kimler, Ralf Steinberger, Camelia Ignat, Tamara Oellinger, Ken Blackler, Flavio Fuart, Wajdi Zaghouani, Anna Widiger, Ann-Charlotte Forslund, Clive Best (2006). [Geocoding multilingual texts: Recognition, Disambiguation and Visualisation](#). Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006), pp. 53-58. Genoa, Italy, 24-26 May 2006.

- Abramowicz Witold, Agata Filipowska, Jakub Piskorski, Krzysztof Wecel & Karol Wieloch (2006). **Linguistic Suite for Polish Cadastral System.**. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006), pp. 53-58. Genoa, Italy, 24-26 May 2006.

- Oellinger Tamara & Pinar Özden Wennerberg (2006). **Ontology based modeling and visualization of social networks for the web**. Workshop 'Formation of Social Networks in Social Software Applications' at 36. Jahrestagung der Gesellschaft für Informatik (Informatik'2006), Dresden, Germany, 6.10.2006.

- Yun-Chuang Chiao, Olivier Kraif, Dominique Laurent, Thi Minh Huyen Nguyen, Nasredine Semmar, François Stuck, Jean Véronis, Wajdi Zaghouani (2006). **Evaluation of multilingual text alignment systems: the ARCADE II project**. Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006). pp. 1975-1978. Paris: Hermès-Lavoisier. Genoa, Italy, 24-26 May 2006.

- Norguet Jean-Pierre, Esteban Zimányi & Ralf Steinberger (2006). **Semantic analysis of web site audience**. 21st Annual ACM Symposium on Applied Computing (ACM SAC'2006), Dijon, France, 23-27.04.2006. Pages 525-529.
- Žižka Jan, Jiří Hroza, Bruno Pouliquen, Camelia Ignat & Ralf Steinberger (2006). **The selection of electronic text documents supported by only positive examples**. Proceedings of the 8th International Conference on the Statistical Analysis of Textual Data (JADT'2006). Besançon, 19-21 April 2006.
- Pouliquen Bruno, Ralf Steinberger, Camelia Ignat & Tamara Oellinger (2006). **Building and displaying name relations using automatic unsupervised analysis of newspaper articles**. Proceedings of the 8th International Conference on the Statistical Analysis of Textual Data (JADT'2006). Besançon, 19-21 April 2006.
- Ignat Camelia & François Rousselot (2006). **Un algorithme de génération de profil de document et son évaluation dans le contexte de la classification thématique**. Proceedings of the 8th International Conference on the Statistical Analysis of Textual Data (JADT'2006). Besançon, 19-21 April 2006.
- Best Clive, Bruno Pouliquen, Ralf Steinberger, Erik van der Goot, Ken Blackler, Flavio Fuart, Tamara Oellinger & Camelia Ignat (2006). **Towards automatic event tracking**. In: Sharad Mehrota, Daniel Zeng, Hsinchun Chen, Bhavani Thuraisingham & Fei-Yue Wang (Eds.): *Intelligence and Security Informatics* - Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI'2006), San Diego, California, USA, 23-24.05.2006. Springer Lecture Notes in Computer Science, LNCS 3975, pp. 26-34. Springer-Verlag, Berlin Heidelberg, New York. ISBN: 978-3-540-34478-0.
- Norguet Jean-Pierre, Esteban Zimányi & Ralf Steinberger (2006). **Improving web sites with web usage mining, web content mining, and semantic analysis**. In: Jirí Wiedermann, Gerard Tel, Jaroslav Pokorný, Mária Bieliková, Július Štuller (Eds.): SOFSEM 2006: Theory and Practice of Computer Science. 32nd Conference on Current Trends in Theory and Practice of Computer Science, Merin, Czech Republic, 21.-27.01.2006. Proceedings. Lecture Notes in Computer Science, LNCS 3831, pages 430-439. ISBN: 978-3-540-31198-0. Springer-Verlag, Berlin, Heidelberg, New York.

2005

- → Steinberger Ralf, Bruno Pouliquen, Camelia Ignat (2005). [Navigating multilingual news collections using automatically extracted information](#). Journal of Computing and Information Technology - CIT 13, 2005, 4, 257-264.
- → Pouliquen Bruno, Ralf Steinberger, Camelia Ignat, Irina Temnikova, Anna Widiger, Wajdi Zaghouani & Jan Žižka (2005). [Multilingual person name recognition and transliteration](#). Journal CORELA - Cognition, Représentation, Langage. Numéros spéciaux, Le traitement lexicographique des noms propres. ISSN 1638-5748.
- Erjavec Tomaž, Camelia Ignat, Bruno Pouliquen & Ralf Steinberger (2005). **Massive multilingual corpus compilation: Acquis Communautaire and totale**. Journal Archives of Control Sciences, Volume *15(LI), 2005, No. 4, pages 529-540*.
- Steinberger Ralf, Bruno Pouliquen, Camelia Ignat (2005). Navigating multilingual news collections using automatically extracted information. In: Vesna Lužar-Stiffler & Vesna Hljuz Dobric (Eds.): Proceedings of the 27th International Conference 'Information Technology Interfaces' (ITI'2005), pp. 27-34. Cavtat / Dubrovnik, Croatia, June 20-23, 2005.
- Montejo-Ráez Arturo, L. Alfonso Ureña-López & Ralf Steinberger (2005). **Text categorisation using bibliographic records: beyond document content**. Journal Procesamiento del Lenguaje Natural (PLN), núm. 35 (2005), pp. 119-126. Proceedings of the 21st Conference of the Spanish Society for Natural Language Processing (SEPLN'2005). Granada, Spain, 14-16 September 2005.
- → Ignat Camelia, Bruno Pouliquen, Ralf Steinberger & Tomaž Erjavec (2005). [A tool set for the quick and efficient exploration of large document collections](#). Proceedings of the Symposium on Safeguards and Nuclear Material Management. 27th Annual Meeting of the European SAfeguards Research and Development Association (ESARDA-2005). London, UK, 10-12 June 2005.
- Erjavec Tomaž, Camelia Ignat, Bruno Pouliquen & Ralf Steinberger (2005). **Massive multilingual corpus compilation; Acquis Communautaire and totale**. In: 2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (L&T'05). Poznań, Poland, 21-23 April 2005.

- Pouliquen Bruno, Ralf Steinberger, Camelia Ignat, Irina Temnikova, Wajdi Zaghouani & Jan Žižka (2005). **Detection of person names and their translations in multilingual news**. Colloque Traitement lexicographique des noms propres, Tours, 24 March 2005.
- Best Clive, Erik van der Goot, Ken Blackler, Teofilo Garcia, David Horby, Ralf Steinberger and Bruno Pouliquen (2005). **Mapping World Events**. In: Peter van Oosterom, Siyka Zlatanova & Elfriede M. Fendel (eds.) *Geo-information for Disaster Management*. pp. 683-696. Springer. ISBN: 3-540-24988-5.
- Pouliquen Bruno, Franck Le Duff, Denis Delamarre, Marc Cuggia, Fleur Mougin, Pierre Le Beux (2005). **Managing educational resource in medicine: system design and integration**. In: International Journal of Medical Informatics V. 74, pages 201-207.

2004

- Pouliquen Bruno, Ralf Steinberger & Camelia Ignat (2004). Automatic Linking of Similar Texts Across Languages. In: N. Nicolov, K. Bontcheva, G. Angelova & R. Mitkov (eds.): Current Issues in Linguistic Theory 260 - Recent Advances in Natural Language Processing III. Selected Papers from RANLP'2003. John Benjamins Publishers, Amsterdam.
- Steinberger Ralf, Bruno Pouliquen & Camelia Ignat (2004). Providing cross-lingual information access with knowledge-poor methods. In: Andrej Brodnik, Matjaž Gams & Ian Munro (eds.): Informatica. An international Journal of Computing and Informatics. Vol. 28-4, pp. 415-423. Special Issue 'Information Society in 2004'. ISSN: 0350-5596. The Slovene Society Informatica, Ljubljana, Slovenia.
- Montejo-Ráez Arturo & Ralf Steinberger (2004). **Why keywording matters**. In. *High Energy Physics Libraries Webzine*, Issue 10, December 2004.
- Steinberger Ralf, Bruno Pouliquen & Camelia Ignat (2004). Exploiting Multilingual Nomenclatures and Language-Independent Text Features as an Interlingua for Cross-lingual Text Analysis Applications. In: Information Society 2004 (IS'2004) - Proceedings B of the 7th International Multiconference - Language Technologies, pages 2-12. Ljubljana, Slovenia, 13-14 October 2004.
- Montejo-Ráez Arturo, Luís Alfonso Ureña-López & Ralf Steinberger (2004). **Adaptive selection of base classifiers in one-against-all learning for large multi-labeled collections**. In: J.L. Vicedo, P. Martínez-Barco, R. Muñoz et al. (eds). Advances in Natural Language Processing: 4th International Conference, España for Natural Language Processing (EsTAL'2004), Proceedings. Lecture Notes in Computer Science, LNCS 3230, pages 1-12. Alicante, Spain, 20-22 October 2004.
- Pouliquen Bruno, Ralf Steinberger, Camelia Ignat, Emilia Käsper & Irina Temnikova (2004). Multilingual and cross-lingual news topic tracking. In: Proceedings of the 20th International Conference on Computational Linguistics (CoLing'2004), Vol. II, pages 959-965. Geneva, Switzerland, 23-27 August 2004.
- Pouliquen Bruno, Ralf Steinberger, Camelia Ignat & Tom de Groeve (2004). Geographical Information Recognition and Visualisation in Texts Written in Various Languages. In: Proceedings of the 19th Annual ACM Symposium on Applied Computing (SAC'2004), Special Track on Information Access and Retrieval (SAC-IAR), vol. 2, pp. 1051-1058. Nicosia, Cyprus, 14 - 17 March 2004.

2003

- Pouliquen Bruno, Ralf Steinberger & Camelia Ignat (2003). Automatic Identification of Document Translations in Large Multilingual Document Collections. In: Proceedings of the International Conference *Recent Advances in Natural Language Processing* (RANLP'2003). Borovets, Bulgaria, 10 - 12 September 2003.
- Ignat Camelia, Bruno Pouliquen, António Ribeiro & Ralf Steinberger (2003). Extending an Information Extraction Tool Set to Central and Eastern European Languages. In: Proceedings of the Workshop *Information Extraction for Slavonic and other Central and Eastern European Languages* (IESL'2003), held at RANLP'2003. Borovets, Bulgaria, 8 - 9 September 2003.
- Pouliquen Bruno, Steinberger Ralf, Camelia Ignat (2003). Automatic Annotation of Multilingual Text Collections with a Conceptual Thesaurus. In: Proceedings of the Workshop *Ontologies and Information Extraction* at the Summer School *The Semantic Web and Language Technology - Its Potential and Practicalities* (EUROLAN'2003). Bucharest, Romania, 28 July - 8 August 2003.

- Steinberger Ralf, Bruno Pouliquen, Stefan Scheer & António Ribeiro (2003). **Continuous Multi-Source Information Gathering and Classification**. In: Proceedings of the International Conference on Computational Intelligence for Modeling, Control and Automation (CIMCA'2003). Vienna, Austria, 12-14 February 2003.

2002

- Happe André, Bruno Pouliquen, Anita Burgun, Marc Cuggia, Pierre Le Beux (2002). **Combining voice recognition and automatic indexing of medical reports**. Proceedings of XVIIth International Congress of the European Federation for Medical Informatics (MIE 2002). Studies in health technology and informatics, 2002, 90: 382-7. Budapest, 25-29 August 2002.
- Mary Vincent, Bruno Pouliquen, Franck Le Duff, Stefan J. Darmoni, Alain Segui, Pierre Le Beux (2002). **Automatic conceptual indexing of French Pharmaceutical theses**. Proceedings of XVIIth International Congress of the European Federation for Medical Informatics (MIE 2002). Studies in health technology and informatics, 2002, 90. Budapest, 25-29 August 2002.
- ➔ Steinberger Ralf, Bruno Pouliquen & Johan Hagman (2002). [Cross-lingual Document Similarity Calculation Using the Multilingual Thesaurus Eurovoc](#). In: A. Gelbukh (ed.) Computational Linguistics and Intelligent Text Processing, Third International Conference, CICLing'2002. Springer Lecture Notes in Computer Science, LNCS 2276, pp. 415-424. Mexico-City, Mexico, 17-23 February 2002. Springer-Verlag, Berlin Heidelberg. ISSN: 0302-9743.
- Hagman Johan (2002). **Zooming in on some Components of a System for Gathering, Analyzing, and Visualizing Multilingual Data**. In: A. Morin & P. Sébillot (eds.): 6th International Conference on the Statistical Analysis of Textual Data, JADT'2002, vol. 1, pp 347-348. St. Malo, France, 13-15 March 2002.
- Pouliquen Bruno, Denis Delamarre & Pierre Le Beux (2002). **Indexation de textes médicaux par extraction de concepts, et ses utilisations**. In: A. Morin & P. Sébillot (eds.): 6th International Conference on the Statistical Analysis of Textual Data, JADT'2002, vol. 2, pp 617-628. St. Malo, France, 13-15 March 2002.

2001 and earlier

- ➔ Steinberger Ralf (2001). [Cross-lingual Keyword Assignment](#). Proceedings of the XVII Conference of the Spanish Society for Natural Language Processing (SEPLN'2001). Procesamiento del Lenguaje Natural, Revista N° 27, pp. 273-280. Jaén, Spain, 12-14 September 2001. ISSN 1135-5948.
- Steinberger Ralf, Johan Hagman & Stefan Scheer (2000). **Using Thesauri for Information Extraction and for the Visualisation of Multilingual Document Collections**. Proceedings of the Workshop on Ontologies and Lexical Knowledge Bases (OntoLex'2000), pp. 130-141. Sozopol, Bulgaria, September 2000.
- Hagman Johan, Domenico Perrotta, Ralf Steinberger & Aristide Varfis (2000). **Document Classification and Visualisation to Support the Investigation of Suspected Fraud**. Working Notes of the Workshop on Machine Learning and Textual Information Access (MLTIA) at the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'2000), 12 pages. Lyon, September 2000.
- Hagman Johan, Ralf Steinberger, Domenico Perrotta & Aristide Varfis (1999). **Approaches to Document Classification and Visualisation**. Working Notes of the Workshop on Text Mining at the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99), Stockholm, August 1999. JRC reference number: ORA 60278.

Selected slides, presentations, videos, reports, history

- ➔ [Video lectures on the *Europe Media Monitor* EMM by Erik van der Goot](#) (2010, 2012).
- ➔ [Video lectures on the *Europe Media Monitor* EMM by Ralf Steinberger](#) (2006, 2008, 2009, 2011).
- ➔ [EMM HISTORY](#): Scheer Stefan, Ralf Steinberger, Giovanni Valerio & Paul Henshaw (2000). [A Methodology to Retrieve, to Manage, to Classify and to Query Open Source Information](#) - Results of the OSILIA Project. JRC

Technical Note No. I.01.016, 35 pages. December 2000. ([OSILIA was the precursor system of the Europe Media Monitor EMM](#), see also [slides on the EMM precursor OSILIA: Automatic gathering of newspaper articles on internet abuse from the internet](#), JRC, 20.12.2000)

➔ Steinberger Ralf & Bruno Pouliquen (2008). [NewsExplorer - combining various text analysis tools to allow multilingual news linking and exploration](#). Notes for the lecture held at the SORIA Summer School *Cursos de Tecnologías Lingüísticas: Técnicas de extracción y visualización de información: aplicación en la construcción de portales especializados*, Fundación Duques de Soria, 7-11 July 2008, Soria, Spain.

➔ **Keynote talk at the 7th international Conference on Language Resources and Evaluation** ([LREC'2010](#)). [Challenges and solutions for multilingual text mining](#). Valletta, Malta, 21.05.2010.

- Piskorski Jakub (2008). **CORLEONE - Core Linguistic Entity Online Extraction**. Technical Report EN 23393, Joint Research Centre of the European Commission, Ispra, Italy.
- Balahur-Dobrescu Alexandra (2011). **Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types**. Ph.D. Thesis, University of Alicante. ([Read online](#))
- Kimler Marco (2004). **Geo-Coding: Recognition of geographical references in unstructured text, and their visualisation**. Diploma thesis submitted to the University of Applied Sciences in Hof, Germany, in August 2004. 85 pages.
- Pedersen Jane & Ralf Steinberger (2002). **Evaluation of Multilingual Name Recognition Software - Thing Finder (TM) 2.2**. JRC Technical Note No. I.02.120, 29 pages. December 2002.
- Steinberger Ralf (1999). **Language Engineering Technologies and Their Use for TF-UCLAF**. A report on the JRC's institutional activities in 1998. JRC Technical Note No. I.99.83, April 1999.