UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Time-aware online reputation analysis

Peetz, M.-H.

Publication date
2015
Document Version
Final published version

Link to publication

Citation for published version (APA):
Peetz, M.-H. (2015). *Time-aware online reputation analysis*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:14 Nov 2026

Bibliography

[1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In *WSDM '08*, pages 183–194, 2008. (Cited on page 15.)

[2] G. Ainslie and N. Haslam. *Choice over time*, chapter Hyperbolic discounting. Russell Sage Foundation, 1992. (Cited on page 130.)

[3] O. Alonso, J. Strötgen, R. Baeza-Yates, and M. Gertz. Temporal information retrieval: Challenges and opportunities. In *TWAW '11*, pages 1–8, 2011. (Cited on page 19.)

[4] G. Amati, G. Amodeo, M. Bianchi, G. Marcone, C. Gaibisso, A. Celi, C. Nicola, and M. Flammini. FUB, IASI-CNR, UNIVAQ at TREC 2011. *TREC 2011*, 2011. (Cited on page 132.)

[5] E. Amigó, J. Artiles, J. Gonzalo, D. Spina, B. Liu, and A. Corujo. WePS-3 evaluation campaign: Overview of the online reputation management task. In *CLEF '10 (Online Working Notes/Labs/Workshop)*, 2010. (Cited on pages 2, 21, 51, 65, and 141.)

[6] E. Amigó, A. Corujo, J. Gonzalo, E. Meij, and M. de Rijke. Overview of RepLab 2012: Evaluating online reputation management systems. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on pages 2, 15, 16, 21, 25, 32, 51, 58, 63, 65, 66, 72, 73, 77, 93, and 141.)

[7] E. Amigó, J. Carrillo de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Martín, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2013: Evaluating online reputation monitoring systems. In *CLEF '13*, pages 333–352. Springer, 2013. (Cited on pages 2, 5, 15, 16, 21, 27, 32, 51, 63, 72, 74, 83, 93, 141, and 148.)

[8] E. Amigó, J. Gonzalo, and F. Verdejo. A general evaluation measure for document organization tasks. In *SIGIR '13*, pages 643–652, July 2013. (Cited on pages 149 and 150.)

[9] E. Amigó, J. Carrillo-de Albornoz, I. Chugur, A. Corujo, J. Gonzalo, E. Meij, M. de Rijke, and D. Spina. Overview of RepLab 2014: Author profiling and reputation dimensions for online reputation management. In *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, volume 8685, pages 307–322. Springer, 2014. (Cited on pages 32 and 92.)

[10] G. Amodeo, G. Amati, and G. Gambosi. On relevance, time and query expansion. In *CIKM '11*, pages 1973–1976, 2011. (Cited on page 21.)

[11] S. Aral. What would Ashton do—and does it matter? In *Harvard Business Review*, 2013. (Cited on pages 11, 60, and 62.)

[12] S. Aral and D. Walker. Identifying influential and susceptible members of social networks. *Science*, 337 (6092):337–341, 2012. (Cited on pages 10 and 62.)

[13] J. Atserias, G. Attardi, M. Simi, and H. Zaragoza. Active learning for building a corpus of questions for parsing. In *LREC '10*, 2010. (Cited on page 145.)

[14] M. Atvesson. Organization: From substance to image? *Organization Studies*, 11(3):373–394, 1990. (Cited on page 12.)

[15] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, and S. Vigna. Four degrees of separation. In *WebSci '12*, pages 33–42. ACM, 2012. (Cited on page 10.)

[16] P. Bailey, N. Craswell, I. Soboroff, P. Thomas, A. P. de Vries, and E. Yilmaz. Relevance assessment: Are judges exchangeable and does it matter? In *SIGIR '08*, pages 667–674, 2008. (Cited on page 34.)

[17] R. Bakhshandeh, M. Samadi, Z. Azimifar, and J. Schaeffer. Degrees of separation in social networks. In *SOCS '11*, 2011. (Cited on page 10.)

[18] A. Balahur and H. Tanev. Detecting entity-related events and sentiments from tweets using multilingual resources. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[19] A. Balahur, R. Steinberger, M. A. Kabadjov, V. Zavarella, E. V. der Goot, M. Halkia, B. Pouliquen, and J. Belyaeva. Sentiment analysis in the news. In *LREC '10*, 2010. (Cited on page 67.)

[20] S. Balbo. EMA: automatic analysis mechanism for the ergonomic evaluation of user interfaces. *CSIRO-Division of Information Technology–Sydney, Technical Report*, 96:44, 1996. (Cited on page 34.)

[21] K. Balog, W. Weerkamp, and M. de Rijke. A few examples go a long way: constructing query models from elaborate query formulations. In *SIGIR '08*, pages 371–378, 2008. (Cited on page 18.)

[22] K. Balog, M. de Rijke, R. Franz, M.-H. Peetz, B. Brinkman, I. Johgi, and M. Hirschel. Sahara: Discovering entity-topic associations in online news. In *ISWC '09*. Springer, Springer, 10/2009 2009. (Cited on page 8.)

[23] K. Balog, M. Bron, and M. de Rijke. Category-based query modeling for entity search. In *ECIR '10*, pages 319–331, 2010. (Cited on page 18.)

[24] D. Barnlund. A Transactional Model of Communication. *Foundations of communication theory*, pages 23–45, 1970. (Cited on page 70.)

[25] R. Bekkerman, A. Mccallum, G. Huang, and Others. Automatic Categorization of Email into Folders:

Benchmark Experiments on Enron and SRI Corpora. Technical report, Center for Intelligent Information Retrieval, 2004. (Cited on pages 26 and 74.)

[26] K. Berberich and S. Bedathur. Temporal diversification of search results. In *TAIA '13*, 2013. (Cited on page 21.)

[27] K. Berberich, S. Bedathur, O. Alonso, and G. Weikum. A language modeling approach for temporal information needs. In *ECIR '10*, pages 13–25. Springer, 2010. (Cited on page 21.)

[28] M. Berman. In search of decay in verbal short term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 35:317–333, 2009. (Cited on page 127.)

[29] M. S. Bernstein, E. Bakshy, M. Burke, and B. Karrer. Quantifying the invisible audience in social networks. In *CHI '13*, pages 21–30, 2013. (Cited on page 10.)

[30] A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavaldà. New ensemble methods for evolving data streams. In *KDD '09*, pages 139–148, 2009. (Cited on page 143.)

[31] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022, 2003. (Cited on page 110.)

[32] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011. (Cited on page 3.)

[33] A. A. Bolourian, Y. Moshfeghi, and C. J. V. Rijsbergen. Quantification of Topic Propagation Using Percolation Theory: A Study of the ICWSM Network. In *ICWSM '09*. AAAI Press, 2009. (Cited on page 11.)

[34] N. Booth and J. A. Matic. Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, 16(3):184–191, 2011. (Cited on page 62.)

[35] J. Borges and M. Levene. Mining association rules in hypertext databases. In *KDD '99*, pages 149–153, 1999. (Cited on page 34.)

[36] J. Borges and M. Levene. Data mining of user navigation patterns. In B. Masand and M. Spiliopoulou, editors, *Web Usage Analysis and User Profiling*, volume 1836, pages 92–112. Springer, 2000. (Cited on page 34.)

[37] M. Bron, K. Balog, and M. de Rijke. Ranking related entities: Components and analyses. In *CIKM '10*, Toronto, 2010. ACM. (Cited on page 93.)

[38] M. Bron, E. Meij, M.-H. Peetz, E. Tsagkias, and M. de Rijke. Team commit at trec 2011. In *TREC 2011*. NIST, 02/2012 2012. (Cited on pages 7 and 8.)

[39] M. Bron, J. van Gorp, F. Nack, M. de Rijke, A. Vishneuski, and S. de Leeuw. A subjunctive exploratory search interface to support media studies researchers. In *SIGIR '12*, pages 425–434, 2012. (Cited on page 34.)

[40] M. Bron, J. Van Gorp, and M. de Rijke. Media studies research in the data-driven age: How research questions evolve. *Journal of the American Society for Information Science and Technology*, 2015. To appear. (Cited on page 31.)

[41] J. Bullas. Blogging Statistics, Facts, and Figures, 2012. URL `http://www.jeffbullas.com/2012/08/02/blogging-statistics-facts-and-figures-in-2012-infographic/`. (Cited on page 2.)

[42] N. Bunkley. Hyundai apologizes for U.K. ad that depicts suicide attempt, 2013. URL `http://adage.com/article/news/hyundai-apologizes-u-k-ad-depicts-suicide-attempt/241119/`. (Cited on page 11.)

[43] M. Burghardt. Usability recommendations for annotation tools. In *LAW '12*, pages 104–112, 2012. (Cited on page 35.)

[44] J. Carrillo-de Albornoz, I. Chugur, and E. Amigó. Using an emotion-based model and sentiment analysis techniques to classify polarity for reputation. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[45] C. Carroll. *The Handbook of Communication and Corporate Reputation*. Handbooks in Communication and Media. Wiley, 2013. (Cited on page 12.)

[46] S. Carter, W. Weerkamp, and E. Tsagkias. Microblog language identification: Overcoming the limitations of short, unedited and idiomatic text. *Language Resources and Evaluation*, 47(1):195–215, 2013. (Cited on pages 24, 69, and 73.)

[47] L. Caspari. Der #aufschrei und seine Folgen. *Zeit Online*, January 2014. URL `http://www.zeit.de/politik/deutschland/2014-01/sexismus-debatte-folgen`. (Cited on page 2.)

[48] A. Castellanos, J. Cigarrán, and A. García-Serrano. Modelling techniques for Twitter contents: A step beyond classification based approaches. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2013. (Cited on page 15.)

[49] N. V. Chawla. Data mining for imbalanced datasets: An overview. In *Data Mining and Knowledge Discovery Handbook*. Springer, 2010. (Cited on page 73.)

[50] J. M. Chenlo, J. Atserias, C. Rodriguez, and R. Blanco. FBM-Yahoo! at RepLab 2012. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[51] A. G. Chessa and J. M. Murre. A memory model for internet hits after media exposure. *Physica A Statistical Mechanics and its Applications*, 2004. (Cited on pages 125, 127, and 128.)

[52] A. G. Chessa and J. M. Murre. Modelling memory processes and internet response times: Weibull or power-law? *Physica A Statistical Mechanics and its Applications*, 2006. (Cited on pages 125, 127, 128, and 138.)

[53] J. Choi and W. B. Croft. Temporal models for microblogs. In *CIKM '12*, pages 2491–2494, 2012. (Cited on page 20.)

[54] W. Chu, M. Zinkevich, L. Li, A. Thomas, and B. Tseng. Unbiased online active learning in data streams. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 195–203, 2011. (Cited on page 143.)

[55] K. G. Coffman and A. M. Odlyzko. The size and growth rate of the internet. *First Monday*, 3(10):l–25, 1998. (Cited on page 16.)

[56] G. M. D. Corso, A. Gullí, and F. Romani. Ranking a stream of news. In *WWW '05*, 2005. (Cited on page 20.)

[57] A. Corujo. Meet the user. Keynote Speech at RepLab 2012, 2012. (Cited on pages 2, 15, 16, 64, 66, 69, 73, 74, and 88.)

[58] J.-V. Cossu, B. Bigot, L. Bonnefoy, M. Morchid, X. Bost, G. Senay, R. Dufour, V. Bouvier, J.-M. Torres-Moreno, , and M. El-Beze. LIA@RepLab 2013. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2013. (Cited on pages 15 and 22.)

[59] T. M. Cover and P. E. Hart. Nearest neighbour pattern classification. In *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, volume 13, pages 21–27, 1967. (Cited on page 101.)

[60] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan, N. Aswani, I. Roberts, G. Gorrell, A. Funk, A. Roberts, D. Damljanovic, T. Heitz, M. A. Greenwood, H. Saggion, J. Petrak, Y. Li, and W. Peters. *Text Processing with GATE (Version 6)*. Gateway Press, 2011. URL `http://tinyurl.com/gatebook`. (Cited on pages 34 and 35.)

[61] W. Dakka, L. Gravano, and P. G. Ipeirotis. Answering general time-sensitive queries. In *CIKM '08*, pages 1437–1438, 2008. (Cited on page 21.)

[62] W. Dakka, L. Gravano, and P. G. Ipeirotis. Answering general time-sensitive queries. *IEEE Transactions on Knowledge and Data Engineering*, 24(2):220–235, 2012. (Cited on pages 21, 23, 29, 94, and 166.)

[63] danah boyd. Making sense of teen life: Strategies for capturing ethnographic data in a networked era. *Digital Research Confidential: The Secrets of Studying Behavior Online*, in press. (Cited on pages 31 and 33.)

[64] G. Davies. *Corporate Reputation and Competitiveness*. Psychology Press, 2003. (Cited on page 13.)

[65] J. C. de Albornoz, E. Amigó, D. Spina, and J. Gonzalo. ORMA: A semi-automatic tool for online reputation monitoring in twitter. In *ECIR '14*, pages 742–745, 2014. (Cited on page 15.)

[66] F. Diaz and D. Metzler. Improving the estimation of relevance models using large external corpora. In *SIGIR '06*, pages 154–161, 2006. (Cited on pages 19 and 93.)

[67] G. R. Dowling. Managing your corporate image. *Industrial Marketing Management*, 15(2):109–115, 1986. (Cited on page 12.)

[68] L. L. Downey and D. M. Tice. A usability case study using TREC and ZPRISE. *Information Processing & Management*, 35(5):589 – 603, 1999. (Cited on pages 34 and 35.)

[69] M. Duggan and A. Smith. Social media update 2013, 2013. URL `http://www.pewinternet.org/2013/12/30/demographics-of-key-social-networking-platforms/`. (Cited on pages 31 and 33.)

[70] P. Dyer. Blogs influence consumer spending more than social networks, 2013. URL `http://www.pamorama.net/2013/03/14/blogs-influence-consumer-spending-more-than-social-networks/#.UUIWiVfp_Xt`. (Cited on page 2.)

[71] H. Ebbinghaus. *Memory: a contribution to experimental psychology*. Teachers College, Columbia University, 1913. (Cited on page 126.)

[72] M. Efron. Information search and retrieval in microblogs. *Journal of the American Society for Information Science and Technology*, 62(6):996–1008, 2011. (Cited on page 20.)

[73] M. Efron. Query-specific recency ranking: Survival analysis for improved microblog retrieval. In *TAIA '12*, 2012. (Cited on pages 20 and 137.)

[74] M. Efron and G. Golovchinsky. Estimation methods for ranking recent information. In *SIGIR '11*, pages 495–504, 2011. (Cited on pages 17, 20, 23, 29, 94, 101, 103, and 125.)
[75] M. Efron, P. Organisciak, and K. Fenlon. Improving retrieval of short texts through document expansion. In *SIGIR '12*, pages 911–920, 2012. (Cited on pages 20, 125, and 166.)
[76] S. Eliot and J. Rose. *A Companion to the History of the Book*, volume 98. John Wiley & Sons, 2009. (Cited on page 16.)
[77] G. Eryiğit. ITU treebank annotation tool. In *LAW '07*, Prague, 24-30 June 2007. (Cited on pages 34 and 35.)
[78] Facebook. Facebook reports fourth quarter and full year 2013 results, 2014. URL `http://investor.fb.com/releasedetail.cfm?ReleaseID=821954#sthash.SZWPNhKh.dpuf`. (Cited on page 2.)
[79] M. A. Faust. The use of social media and the impact of support on the well-being of adult cystic fibrosis patients. Technical report, University of South Carolina - Columbia, 2014. Thesis (MA). (Cited on pages 31 and 33.)
[80] R. L. Figueroa, Q. Zeng-Treitler, L. H. Ngo, S. Goryachev, and E. P. Wiechmann. Active learning for clinical text classification: Is it better than random sampling? *Journal of the American Medical Informatics Association*, 19(5):809–816, 2012. (Cited on page 145.)
[81] J. Filgueiras and S. Amir. POPSTAR at RepLab 2013: Polarity for reputation classification. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2013. (Cited on page 15.)
[82] M. Fishbein and I. Ajzen. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Addison-Wesley, 1975. (Cited on page 12.)
[83] C. Fombrun. The RepTrak system. Presented at the 10th Anniversary Conference on Reputation, Image, Identity and Competitiveness, May 2006. (Cited on page 13.)
[84] C. J. Fombrun. Indices of corporate reputation: An analysis of media rankings and social monitors' ratings. *Corporate Reputation Review*, 1(4):327–340, 1998. (Cited on page 12.)
[85] C. J. Fombrun and C. B. Van Riel. *Fame & fortune: How successful companies build winning reputations*. FT Press, 2004. (Cited on pages 1, 10, 12, and 13.)
[86] C. J. Fombrun, N. A. Gardberg, and J. M. Sever. The reputation quotient: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management*, 7(4):241–255, 2000. (Cited on page 13.)
[87] J. R. Frank, M. Kleiman-Weiner, D. A. Roberts, F. Niu, C. Zhang, C. Ré, and I. Soboroff. Building an entity-centric stream filtering test collection for trec 2012. Technical report, DTIC Document, 2012. (Cited on page 21.)
[88] J. R. Frank, S. J. Bauer, M. KleimanAWeine, D. A. Roberts, N. Tripuraneni, C. Zhang, C. Re, E. M. Voorhees, and I. Soboroff. Evaluating stream filtering for entity profile updates for TREC 2013. In *TREC 2013*, 2013. (Cited on page 21.)
[89] L. Gaines-Ross. Leveraging corporate equity. *Corporate Reputation Review*, 1:51–56, 1998. (Cited on page 13.)
[90] E. Garfield. A tribute to Calvin N. Mooers, a pioneer of information retrieval. *The Scientist*, 11(6):9, 1997. (Cited on page 16.)
[91] P. Gillin. *The New Influencers: A Marketer's Guide to the New Social Media*. Quill Driver Books, Sanger, CA, 2007. (Cited on page 1.)
[92] D. Graus, M.-H. Peetz, D. Odijk, O. de Rooij, and M. de Rijke. yourHistory – Semantic linking for a personalized timeline of historic events. In *LinkedUp Veni Competition on Linked and Open Data for Education*, 2014. (Cited on page 8.)
[93] M. A. Greenwood, N. Aswani, and K. Bontcheva. Reputation profiling with gate. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)
[94] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *WWW '04*, pages 491–501. ACM, 2004. (Cited on page 146.)
[95] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi. Faking Sandy: Characterizing and identifying fake images on twitter during hurricane Sandy. In *WWW '13 Companion*, pages 729–736, 2013. (Cited on page 51.)
[96] D. Gupta and K. Berberich. Identifying time intervals of interest to queries. In *CIKM '14*, 2014. (Cited on page 21.)
[97] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explorations Newsletter*, 11(1):10–18, Nov. 2009. (Cited on page 75.)
[98] J. D. Hamilton. *Time-Series Analysis*. Princeton University Press, 1 edition, Jan. 1994. (Cited on page 101.)
[99] V. Hangya and R. Farkas. Filtering and polarity detection for reputation management on tweets. In

CLEF '13 (Online Working Notes/Labs/Workshop), 2013. (Cited on pages 15 and 22.)

[100] D. Harman. Overview of the first text retrieval conference (TREC-1). In *TREC 1992*, pages 1–20, 1992. (Cited on page 16.)

[101] A. Heathcote, S. Brown, and D. J. Mewhort. The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7(2):185–207, 2000. (Cited on page 126.)

[102] R. Hertwig, S. M. Herzog, L. J. Schooler, and T. Reimer. Fluency heuristic: a model of how the mind exploits a by-product of information retrieval. *Journal of experimental psychology. Learning, memory, and cognition*, 2008. (Cited on page 125.)

[103] B. Hilligoss and S. Y. Rieh. Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, 44(4):1467–1484, 2008. (Cited on page 35.)

[104] B. Hjørland. The foundation of the concept of relevance. *Journal of the American Society for Information Science and Technology*, 61(2):217–237, 2010. (Cited on page 19.)

[105] D. L. Hoffmann and M. Fodor. Can you measure the ROI of your social media marketing? *MIT Sloan Management Review*, 52(10):40–49, 2010. (Cited on page 2.)

[106] K. Hofmann and W. Weerkamp. Content extraction for information retrieval in blogs and intranets. Technical report, University of Amsterdam, 2008. (Cited on page 24.)

[107] K. Hofmann, K. Balog, T. Bogers, and M. de Rijke. Contextual factors for finding similar experts. *Journal of the American Society for information Science and Technology*, 61(5):994–1014, 2010. (Cited on pages 31 and 62.)

[108] D. Hopkins and G. King. A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, 2010. (Cited on pages 14 and 32.)

[109] M. Hu and B. Liu. Mining and summarizing customer reviews. In *SIGKDD '04*, 2004. (Cited on page 73.)

[110] R. Hu. *Active learning for text classification*. PhD thesis, Dublin Institute of Technology, 2011. (Cited on page 143.)

[111] B. A. Huberman, D. M. Romero, and F. Wu. Social networks that matter: Twitter under the microscope. *arXiv preprint arXiv:0812.1045*, 2008. (Cited on page 10.)

[112] B. Huurnink, L. Hollink, W. van den Heuvel, and M. de Rijke. Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the American society for information science and technology*, 61(6):1180–1197, 2010. (Cited on page 31.)

[113] D. Ienco, A. Bifet, I. Žliobaitė, and B. Pfahringer. Clustering based active learning for evolving data streams. In *Discovery Science*, volume 8140, pages 79–93. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-40897-7_6. (Cited on page 143.)

[114] M. Y. Ivory and M. A. Hearst. The state of the art in automating usability evaluation of user interfaces. *ACM Computing Surveys*, 33(4):470–516, Dec. 2001. (Cited on page 34.)

[115] A. Iyengar, M. Squillante, and L. Zhang. Analysis and characterization of largescale web server access patterns and performance. *WWW '99*, 2(1-2):85–100, 1999. (Cited on page 34.)

[116] N. A. Jaleel, J. Allan, W. B. Croft, F. Diaz, L. S. Larkey, X. Li, M. D. Smucker, and C. Wade. UMass at TREC 2004: Novelty and hard. In *TREC 2004*, 2004. (Cited on page 18.)

[117] B. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 2009. (Cited on pages 2, 10, 31, 33, and 63.)

[118] A. Java, P. Kolari, T. Finin, A. Joshi, and J. Martineau. The BlogVox opinion retrieval system. In *TREC 2006*, 2006. (Cited on page 19.)

[119] A. Java, X. Song, T. Finin, and B. Tseng. Why We Twitter: Understanding Microblogging Usage and Communities. In *9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, pages 56–65, 2007. (Cited on pages 10, 31, and 33.)

[120] P. Jean. Greenpeace puts dove in their sights and fires off viral "onslaught(er)" video, April 2008. URL `http://www.digitaljournal.com/article/253725`. (Cited on page 11.)

[121] H. Jeong and H. Lee. Using feature selection metrics for polarity analysis in replab 2012. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[122] V. Jijkoun, M. de Rijke, and W. Weerkamp. Generating focused topic-specific sentiment lexicons. In *ACL '10*, pages 585–594, 2010. (Cited on page 14.)

[123] G. H. Jones, B. H. Jones, and P. Little. Reputation as Reservoir: Buffering Against Loss in Times of Economic crisis. *Corporate Reputation Review*, 3(1):29, 2000. (Cited on pages 1 and 12.)

[124] K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972. (Cited on page 17.)

[125] R. Jones and F. Diaz. Temporal profiles of queries. *ACM Transactions on Information Systems*, 25, 2007. (Cited on pages 20, 159, and 166.)
[126] J. Kamps. Improving retrieval effectiveness by reranking documents based on controlled vocabulary. In *ECIR '04*, pages 283–295, 2004. (Cited on page 93.)
[127] R. Kaptein. Learning to Analyze Relevancy and Polarity of Tweets. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)
[128] J. Karlgren, M. Sahlgren, F. Olsson, F. Espinoza, and O. Hamfors. Profiling reputation of corporate entities in semantic space. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)
[129] M. Keikha, S. Gerani, and F. Crestani. Time-based relevance models. In *SIGIR '11*, pages 1087–1088, 2011. (Cited on pages 20, 21, and 94.)
[130] M. Keikha, S. Gerani, and F. Crestani. Temper: a temporal relevance feedback method. In *ECIR '11*, 2011. (Cited on pages 20 and 94.)
[131] G. A. Kelly. *The Psychology of Personal Constructs*. Norton, 1955. (Cited on page 12.)
[132] A. Khodaei and O. Alonso. Temporally-aware signals for social search. In *TAIA '12*, 2012. (Cited on page 20.)
[133] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999. (Cited on page 16.)
[134] A. Kulkarni, J. Teevan, K. M. Svore, and S. T. Dumais. Understanding temporal query dynamics. In *WSDM '11*, 2011. (Cited on page 106.)
[135] K. Kurniawati, G. Shanks, and N. Bekmamedova. The business impact of social media analytics. In *ECIS '13*, 2013. (Cited on page 13.)
[136] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW '10*, pages 591–600, 2010. (Cited on page 10.)
[137] V. Lampos and N. Cristianini. Tracking the flu pandemic by monitoring the social web. In *CIP '10*, pages 411–416, June 2010. doi: 10.1109/CIP.2010.5604088. (Cited on page 3.)
[138] D. Laniado and P. Mika. Making sense of Twitter. In *ISWC '10*, pages 470–485. Springer Berlin Heidelberg, 2010. (Cited on page 51.)
[139] V. Lavrenko and W. B. Croft. Relevance based language models. In *SIGIR '01*, pages 120–127, 2001. (Cited on pages 18, 103, and 105.)
[140] A. Lenhart, K. Purcell, A. Smith, and K. Zickuhr. Social media & mobile internet use among teens and young adults. *Pew Internet & American Life Project*, 2010. (Cited on pages 31 and 33.)
[141] C. Lewis. *Using the "thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center, 1982. (Cited on page 34.)
[142] C. Lewis and J. Rieman. *Task-centered User Interface Design: A Practical Introduction*. University of Colorado, Boulder, Department of Computer Science, 1993. (Cited on page 34.)
[143] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330 – 342, 2008. (Cited on page 10.)
[144] X. Li and W. B. Croft. Time-based language models. In *CIKM '03*, 2003. (Cited on pages 19, 20, 23, 29, 100, 101, 103, 105, 125, 127, 129, 133, 140, and 166.)
[145] J. Lin and M. Efron. Overview of the TREC-2013 microblog track. In *TREC 2013*, 2013. (Cited on page 21.)
[146] J. Lin and M. Efron. Temporal relevance profiles for tweet search. In *TAIA '13*. Citeseer, 2013. (Cited on page 21.)
[147] B. Liu. *Sentiment Analysis and Opinion Mining*. Morgan and Claypool, 2012. (Cited on pages 14 and 67.)
[148] B. Liu, M. Hu, and J. Cheng. Opinion observer: Analyzing and comparing opinions on the web. In *WWW '05*, 2005. (Cited on page 73.)
[149] W. Liu and T. Wang. Active learning for online spam filtering. In *AIRS '08*, pages 555–560, 2008. (Cited on page 143.)
[150] M. Lopatka and M.-H. Peetz. Vibration sensitive keystroke analysis. In *Benelearn '09*, pages 75–80, 2009. (Cited on page 8.)
[151] D. E. Losada and L. Azzopardi. An analysis on document length retrieval trends in language modeling smoothing. *Information Retrieval*, 11(2):109–138, 2008. (Cited on page 17.)
[152] G. Lotan, E. Graeff, M. Ananny, D. Gaffney, I. Pearce, and danah boyd. The Arab Spring—The revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. *International Journal of Communication*, 5(0), 2011. (Cited on pages 2, 3, 11, 31, and 33.)
[153] C. Macdonald and I. Ounis. The TREC Blogs06 collection: Creating and analyzing a blog test collection.

Technical Report TR-2006-224, U. Glasgow, 2006. (Cited on pages 18 and 24.)
[154] W. G. Mangold and D. J. Faulds. Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4):357–365, 2009. (Cited on pages 2 and 63.)
[155] C. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. (Cited on pages 16, 17, 69, 103, 105, and 128.)
[156] B. Martins, H. Manguinhas, and J. Borbinha. Extracting and exploring the geo-temporal semantics of textual resources. In *ICSC '08*, pages 1–9, 2008. (Cited on page 19.)
[157] K. Massoudi, E. Tsagkias, M. de Rijke, and W. Weerkamp. Incorporating query expansion and quality indicators in searching microblog posts. In *ECIR '11*, pages 362–367, 2011. (Cited on pages 20, 101, 125, 127, and 131.)
[158] G. Meck. Ein Werbespot verärgert die Banken-Branche, May 2013. URL `www.faz.net/-gqi-79884`. (Cited on page 2.)
[159] M. Meeter, J. M. J. Murre, and S. M. J. Janssen. Remembering the news: modeling retention data from a study with 14,000 participants. *Memory & Cognition*, 33(5):793–810, 2005. (Cited on pages 125, 126, 131, 132, 138, and 140.)
[160] E. Meij and M. de Rijke. Supervised query modeling using Wikipedia. In *SIGIR '10*, 2010. (Cited on page 18.)
[161] E. Meij, D. Trieschnigg, M. de Rijke, and W. Kraaij. Conceptual language models for domain-specific retrieval. *Information Processing and Management*, 46(4):448–469, 2010. (Cited on page 93.)
[162] D. Meister and D. Sullivan. *Evaluation of User Reactions to a Prototype On-line Information Retrieval System*. Number v. 918 in NASA contractor report. ERIC, 1967. (Cited on page 34.)
[163] D. Metzler, C. Cai, and E. Hovy. Structured event retrieval over microblog archives. In *NAACL HLT '12*, pages 646–655, Stroudsburg, PA, USA, 2012. (Cited on page 20.)
[164] S. Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967. (Cited on page 10.)
[165] G. Mishne and M. de Rijke. Moodviews: Tools for blog mood analysis. In *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, pages 153–154, 2006. (Cited on page 3.)
[166] G. Mishne and M. de Rijke. A study of blog search. In *Proceedings 28th European Conference on Information Retrieval (ECIR 2006)*, pages 289–301. Springer, April 2006. (Cited on page 31.)
[167] A. Moniz and F. de Jong. Sentiment analysis and the impact of employee satisfaction on firm earnings. In *Advances in Information Retrieval*, pages 519–527. Springer, 2014. (Cited on page 14.)
[168] A. Mosquera, J. Fernandez, J. M. Gomez, P. Martnez-Barco, and P. Moreda. DLSI-Volvam at RepLab 2013: Polarity classification on Twitter data. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2013. (Cited on page 15.)
[169] C. Müller and M. Strube. Multi-level annotation of linguistic data with MMAX2. In S. Braun, K. Kohn, and J. Mukherjee, editors, *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*, pages 197–214. Peter Lang, Frankfurt a.M., Germany, 2006. (Cited on pages 34 and 35.)
[170] M. Naaman, J. Boase, and C.-H. Lai. Is it really about me?: Message content in social awareness streams. In *CSCW '10*, pages 189–192, 2010. (Cited on pages 3, 10, 31, and 33.)
[171] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi. Bad news travel fast: A content-based analysis of interestingness on twitter. In *WebSci '11*, 2011. (Cited on pages 64 and 69.)
[172] S. J. Newell and R. E. Goldsmith. The development of a scale to measure perceived corporate credibility. *Journal of Business Research*, 52(3):235–247, 2001. (Cited on page 13.)
[173] R. E. Nisbett and T. D. Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231, 1977. (Cited on page 34.)
[174] D. Odijk, O. de Rooij, M.-H. Peetz, T. Pieters, M. de Rijke, and S. Snelders. Semantic Document Selection. Historical Research on Collections that Span Multiple Centuries. In *TPDL '12*, 2012. (Cited on pages 8, 19, and 166.)
[175] S. Orgad. How can researchers make sense of the issues involved in collecting and interpreting online and offline data? In *Internet Inquiry: Conversations About Method*. Sage, 2009. (Cited on page 33.)
[176] I. Ounis, M. de Rijke, C. Macdonald, G. Mishne, and I. Soboroff. Overview of the TREC-2006 blog track. In *TREC 2006*, 2006. (Cited on pages 14, 18, and 24.)
[177] I. Ounis, C. Macdonald, J. Lin, and I. Soboroff. Overview of the TREC 2011 microblog track. In *TREC 2011*, 2011. (Cited on pages 20, 65, and 146.)
[178] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *WWW '99*, 1999. (Cited on page 16.)
[179] S. Pan, Y. Zhang, and X. Li. Dynamic classifier ensemble for positive unlabeled text stream classification. *Knowledge and Information Systems*, 33(2):267–287, 2012. (Cited on page 143.)
[180] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends® in Information*

Retrieval, 2008. (Cited on pages 14, 68, and 89.)
[181] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *EMNLP '02*, pages 79–86, 2002. (Cited on page 14.)
[182] N. Park and K. M. Lee. Effects of online news forum on corporate reputation. *Public Relations Review*, 33(3):346 – 348, 2007. (Cited on page 65.)
[183] M.-H. Peetz and M. de Rijke. Cognitive temporal document priors. In *ECIR'13*, 2013. (Cited on pages 8 and 84.)
[184] M.-H. Peetz and M. Marx. Tree patterns with full text search. In *WebDB '10*, pages 15:1–15:6. ACM, 2010. (Cited on page 8.)
[185] M.-H. Peetz, M. de Rijke, and A. Schuth. From sentiment to reputation. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on pages 8 and 15.)
[186] M.-H. Peetz, E. Meij, and M. de Rijke. Opengeist: Insight in the stream of page views on wikipedia. In *TAIA '12*, 08/2012 2012. (Cited on page 8.)
[187] M.-H. Peetz, E. Meij, M. de Rijke, and W. Weerkamp. Adaptive temporal query modeling. In *ECIR 2012*, pages 455–458, 2012. (Cited on page 8.)
[188] M.-H. Peetz, E. Meij, and M. de Rijke. Using temporal bursts for query modeling. *Information Retrieval Journal*, 17(1):74–108, July 2013. (Cited on pages 8, 21, and 146.)
[189] M.-H. Peetz, D. Spina, J. Gonzalo, and M. de Rijke. Towards an Active Learning System for Company Name Disambiguation in Microblog Streams. In *CLEF '13 (Online Working Notes/Labs/Workshop)*. CLEF, 2013. (Cited on page 8.)
[190] M.-H. Peetz, M. de Rijke, and R. Kaptein. Estimating reputation polarity on microblog posts. *Under submission*, 2014. (Cited on pages 8, 55, and 62.)
[191] M.-H. Peetz, J. van Gorp, M. de Rijke, M. Bron, R. Berendsen, and W. van Dolen. Social media analysis at work: Outcomes from a multi-method observational study. *Under submission*, 2015. (Cited on page 8.)
[192] J. Pei, J. Han, B. Mortazavi-asl, and H. Zhu. Mining access patterns efficiently from web logs. In T. Terano, H. Liu, and A. Chen, editors, *Knowledge Discovery and Data Mining. Current Issues and New Applications*, volume 1805, pages 396–407. Springer, 2000. (Cited on page 34.)
[193] V. Pérez-Rosas, C. Banea, and R. Mihalcea. Learning Sentiment Lexicons in Spanish. In *LREC '12*, 2012. (Cited on page 73.)
[194] F. Perez-Tellez, D. Pinto, J. Cardiff, and P. Rosso. On the difficulty of clustering microblog texts for online reputation management. In *WASSA '11*, 2011. (Cited on page 21.)
[195] E. Pilkington. Unsold H&M clothes found in rubbish bags as homeless face winter chill, January 2010. URL `http://www.theguardian.com/world/2010/jan/07/h-m-wal-mart-clothes-found`. (Cited on pages 12 and 141.)
[196] B. Piwowarski, A. Trotman, and M. Lalmas. Sound and complete relevance assessment for XML retrieval. *ACM Transactions on Information Systems*, 27(1):1–37, Dec. 2008. (Cited on pages 34 and 35.)
[197] T. Poiesz. The image concept: Its place in consumer psychology and its potential for other psychological areas. In *24th International Congress of Psychology*, 1988. (Cited on page 12.)
[198] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In *SIGIR '98*, pages 275–281, 1998. (Cited on pages 17, 93, 103, and 105.)
[199] S. Porter and A. R. Birt. Is traumatic memory special? A comparison of traumatic memory characteristics with memory for other emotional life experiences. *Applied Cognitive Psychology*, 2001. (Cited on page 125.)
[200] A. Pruyn. Imago: Een analytische benadering van het begrip en de implicaties daarvan voor onderzoek. In C. van Riel, editor, *Corporate communication*, pages 1–5. Bohn Stafleu Van Loghum, 1994. (Cited on page 12.)
[201] J. Pustejovsky, J. M. Castaño, R. Ingria, R. Sauri, R. J. Gaizauskas, A. Setzer, G. Katz, and D. R. Radev. Timeml: Robust specification of event and temporal expressions in text. In *New Directions in Question Answering*, pages 28–34, 2003. (Cited on page 105.)
[202] Y. Qiu and H.-P. Frei. Concept based query expansion. In *SIGIR '93*, pages 160–169, 1993. (Cited on page 18.)
[203] J. K. Rea. The role of social networking sites in the lives of military spouses. Technical report, NCSU, 2014. Thesis (M.Sc.). (Cited on pages 31 and 33.)
[204] Z. Ren, M.-H. Peetz, S. Liang, W. van Dolen, and M. de Rijke. Hierarchical multi-label classification of social text streams. In *SIGIR '14*, July 2014. (Cited on pages 8 and 51.)
[205] E. Riloff, J. Wiebe, and T. Wilson. Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In *CoNLL '03*, 2003. (Cited on page 14.)

[206] S. E. Robertson and I. Soboroff. The TREC 2002 filtering track report. In *TREC 2002*, 2002. (Cited on page 21.)

[207] J. J. Rocchio. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART Retrieval System – Experiments in Automatic Document Processing*, pages 313–323. Prentice Hall, Englewood Cliffs, NJ, 1971. (Cited on page 93.)

[208] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman. Influence and Passivity in Social Media. In *WWW '11*, pages 113–114, 2011. (Cited on page 11.)

[209] D. M. Romero, B. Meeder, and J. Kleinberg. Differences in the Mechanics of Information Diffusion across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In *WWW '11 Companion*, pages 695–704, 2011. (Cited on page 11.)

[210] D. C. Rubin, S. Hinton, and A. Wenzel. The precise time course of retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1999. (Cited on pages 126, 129, and 132.)

[211] S. J. Russell, P. Norvig, J. F. Candy, J. M. Malik, and D. D. Edwards. *Artificial intelligence: A modern approach*. Prentice-Hall, Upper Saddle River, NJ, USA, 1996. (Cited on page 72.)

[212] J. Saias. In search of reputation assessment: experiences with polarity classi cation in RepLab 2013. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2013. (Cited on page 15.)

[213] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes Twitter users: Real-time event detection by social sensors. In *WWW '10*, pages 851–860, 2010. (Cited on pages 2 and 3.)

[214] P. Saleiro, L. Rei, A. Pasquali, C. Soares, J. Teixeira, F. Pinto, M. Nozari, C. Felix, and P. Strecht. POPSTAR at RepLab 2013: Name ambiguity resolution on Twitter. In *CLEF '13 (Online Working Notes/Labs/Workshop)*. CLEF, 2013. (Cited on pages 22 and 149.)

[215] G. Salton, A. Wong, and C. S. Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975. (Cited on page 17.)

[216] C. Sanchez-Sanchez, H. Jimenez-Salazar, and W. A. Luna-Ramirez. UAMCLyR at Replab2013: Monitoring task. In *CLEF '13 (Online Working Notes/Labs/Workshop)*. CLEF, 2013. (Cited on page 22.)

[217] M. Sanderson and W. B. Croft. The history of information retrieval research. *Proceedings of the IEEE*, 100(Special Centennial Issue):1444–1451, 2012. (Cited on page 16.)

[218] M. Sassano. An empirical study of active learning with support vector machines for japanese word segmentation. In *ACL '02*, pages 505–512, 2002. (Cited on page 143.)

[219] L. J. Schooler and J. R. Anderson. The role of process in the rational analysis of memory. *Cognitive Psychology*, 1997. (Cited on page 126.)

[220] D. Sculley. Online active learning methods for fast label-efficient spam filtering. In *CEAS '07*, 2007. (Cited on page 143.)

[221] K. Seki, Y. Kino, S. Sato, and K. Uehara. TREC 2007 Blog Track Experiments at Kobe University. In *TREC 2007*, 2007. (Cited on page 20.)

[222] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009. (Cited on pages 143 and 145.)

[223] D. A. Shamma, L. Kennedy, and E. F. Churchill. Peaks and persistence: Modeling the shape of microblog conversations. In *CSCW '11*, pages 355–358, 2011. (Cited on page 10.)

[224] C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois, 1949. (Cited on page 67.)

[225] D. Shen, J. Zhang, J. Su, G. Zhou, and C.-L. Tan. Multi-criteria-based active learning for named entity recognition. In *ACL '04*, 2004. (Cited on page 143.)

[226] A. C. Siochi and R. W. Ehrich. Computer analysis of user interfaces based on repetition in transcripts of user sessions. *ACM Transactions on Information Systems*, 9(4):309–335, Oct. 1991. (Cited on page 34.)

[227] A. N. Smith, E. Fischer, and C. Yongjian. How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? *Journal of Interactive Marketing*, 26(2):102 – 113, 2012. (Cited on page 10.)

[228] I. Soboroff, I. Ounis, J. Lin, and I. Soboroff. Overview of the TREC 2012 microblog track. In *TREC 2012*, 2012. (Cited on pages 20 and 65.)

[229] D. Spina. *Entity-Based Filtering and Topic Detection for Online Reputation Monitoring in Twitter*. PhD thesis, UNED, 2014. (Cited on pages 142, 143, 148, and 149.)

[230] D. Spina, J. Carrillo de Albornoz, T. Martín, E. Amigó, J. Gonzalo, and F. Giner. UNED Online Reputation Monitoring Team at RepLab 2013. In *CLEF '13 (Online Working Notes/Labs/Workshop)*. CLEF, 2013. (Cited on pages 4, 15, 21, 22, 93, 141, and 162.)

[231] D. Spina, J. Gonzalo, and E. Amigó. Discovering filter keywords for company name disambiguation in Twitter. *Expert Systems with Applications*, 40(12):4986–5003, 2013. (Cited on page 150.)

[232] D. Spina, J. Gonzalo, and E. Amigó. Learning similarity functions for topic detection in online reputa-

tion monitoring. In *SIGIR '14*, pages 527–536, 2014. (Cited on pages 51 and 62.)
[233] B. St. Jean, S. Y. Rieh, J. Y. Yang, and Y.-M. Kim. How content contributors assess and establish credibility on the web. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–11, 2011. (Cited on page 35.)
[234] D. Stacks. *A Practioner's Guide to Public Relations Research, Measurement and Evaluation*. Business Expert Press, 2010. (Cited on pages 1 and 13.)
[235] W. Stephenson. *The study of behavior; Q-technique and its methodology.* University of Chicago Press, Chicago, IL, 1953. (Cited on page 12.)
[236] A. Strauss and J. Corbin. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE, 1998. (Cited on page 42.)
[237] K. Subbian and P. Melville. Supervised rank aggregation for predicting influencers in twitter. In *Passat '11 and Socialcom '11*, pages 661–665, 2011. (Cited on page 62.)
[238] M. R. Subramani and B. Rajagopalan. Knowledge-sharing and influence in online social networks via viral marketing. *Communications of the ACM*, 46(12):300–307, 2003. (Cited on page 62.)
[239] J. Teevan, D. Ramage, and M. R. Morris. #TwitterSearch: A comparison of microblog search and web search. In *WSDM '11*, pages 35–44, 2011. (Cited on page 20.)
[240] M. Thelwall, B. K., and P. G. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012. (Cited on pages 14, 15, 67, and 68.)
[241] T. Tomlinson, D. Huber, C. Riethb, and E. Davelaarc. An interference account of cue-independent forgetting in the no-think paradigm. *Proceedings of the National Academy of Sciences*, 2009. (Cited on page 127.)
[242] S. Tong and D. Koller. Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, 2:45–66, Mar. 2002. (Cited on page 145.)
[243] A. Topalian. Corporate identity: Beyond the visual overstatements. *International Journal of Advertising*, 3(1), 1984. (Cited on page 12.)
[244] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, pages 425–443, 1969. (Cited on page 10.)
[245] M. Tsagkias and K. Balog. The university of amsterdam at WePS3. In *CLEF '10 (Online Working Notes/Labs/Workshop)*, September 2010. (Cited on page 21.)
[246] M. Tsagkias, W. Weerkamp, and M. Rijke. News comments: Exploring, modeling, and online prediction. In *Advances in Information Retrieval*, volume 5993, pages 191–203. Springer, 2010. (Cited on page 95.)
[247] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, 29(4):402–418, 2011. (Cited on pages 2 and 3.)
[248] @Twitter. New tweets per second record, and how!, 2013. URL `https://blog.twitter.com/2013/new-tweets-per-second-record-and-how`. (Cited on page 2.)
[249] Twitter. FAQs about verified accounts, October 2014. URL `https://support.twitter.com/articles/119135-faqs-about-verified-accounts`. (Cited on page 69.)
[250] J. Van Dijck. *The culture of connectivity: A critical history of social media*. Oxford University Press, 2013. (Cited on page 1.)
[251] J. van Dijk, T. Boeschoten, S. ten Tije, and L. van de Wijngaert. De weg naar Haren: De rol van jongeren, sociale media, massamedia en autoriteiten bij de mobilisatie voor Project X Haren : Deelrapport 2, 2013. (Cited on pages 2 and 11.)
[252] G. van Noort and L. Willemsen. Online damage control: The effects of proactive versus reactive webcare interventions in consumer-generated and brand-generated platforms. *Journal of Interactive Marketing*, 2011. (Cited on page 14.)
[253] C. Van Riel. *Principles of Corporate Communication*. Prentice Hall PTR, 1995. (Cited on page 2.)
[254] C. B. M. van Riel and C. J. Fombrun. *Essentials of corporate communication*. Routledge, 2007. (Cited on pages 1, 2, 12, 14, and 63.)
[255] M. W. Van Someren, Y. F. Barnard, J. A. Sandberg, et al. *The think aloud method: A practical guide to modelling cognitive processes*, volume 2. Academic Press London, 1994. (Cited on pages 34 and 42.)
[256] G. Velayathan and S. Yamada. Can we find common rules of browsing behavior? In E. Amitay, C. G. Murray, and J. Teevan, editors, *Query Log Analysis: Social And Technological Challenges.*, May 2007. (Cited on page 34.)
[257] Z. Vendler. Verbs and times. *The Philosophical Review*, 66(2), 1957. (Cited on pages 106 and 122.)
[258] M. Verhagen and J. Pustejovsky. Temporal processing with the TARSQI toolkit. In *22nd International*

Conference on on Computational Linguistics: Demonstration Papers, pages 189–192, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics. (Cited on page 19.)

[259] T. M. Verhallen. Psychologisch marktonderzoek. Inaugural lecture, October 1988. (Cited on page 12.)

[260] J. Villena-Román, S. Lana-Serrano, C. Moreno, J. García-Morera, and J. C. G. Cristóbal. DAEDALUS at RepLab 2012: Polarity classification and filtering on twitter data. In *CLEF '12 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[261] I. Žliobaitė, A. Bifet, B. Pfahringer, and G. Holmes. Active learning with evolving streaming data. In *ECML '11*, pages 597–612, 2011. (Cited on page 143.)

[262] M. Wall. 'blogs of war': Weblogs as news. *Journalism*, 6(2):153–172, 2005. (Cited on page 2.)

[263] K. E. Warner. Tobacco industry response to public health concern: A content analysis of cigarette ads. *Health Education & Behavior*, 12(1):115–127, 1985. (Cited on page 2.)

[264] W. Weerkamp. *Finding People and their Utterances in Social Media*. PhD thesis, University of Amsterdam, 2011. (Cited on page 19.)

[265] W. Weerkamp and M. de Rijke. Credibility improves topical blog post retrieval. In *HLT '08*, pages 923–931, Columbus, Ohio, 2008. (Cited on pages 20, 102, and 119.)

[266] W. Weerkamp and M. de Rijke. Credibility-inspired ranking for blog post retrieval. *Information Retrieval Journal*, 15(3–4):243–277, 2012. (Cited on pages 15, 69, and 97.)

[267] W. Weerkamp, K. Balog, and M. de Rijke. A generative blog post retrieval model that uses query expansion based on external collections. In *ACL/AFNLP '09*, pages 1057–1065, 2009. (Cited on page 19.)

[268] W. Weerkamp, K. Balog, and M. de Rijke. Exploiting external collections for query expansion. *ACM Transactions on the Web*, 6(4):18, 2012. (Cited on page 19.)

[269] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In *WSDM '10*, pages 261–270. ACM, 2010. (Cited on page 62.)

[270] S. Whiting, I. A. Klampanos, and J. M. Jose. Temporal pseudo-relevance feedback in microblog retrieval. In *Advances in Information Retrieval*, pages 522–526. Springer, 2012. (Cited on page 20.)

[271] T. D. Wickens. Measuring the time course of retention. *On human memory: Evolution, progress, and reflections on the 30th anniversary of the Atkinson–Shiffrin model*, 1999. (Cited on pages 126 and 129.)

[272] K.-P. Wiedmann, N. Hennigs, and S. Langner. Spreading the word of fashion: Identifying social influencers in fashion marketing. *Journal of Global Fashion Marketing*, 1(3):142–153, 2010. (Cited on page 62.)

[273] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT '05*, 2005. (Cited on page 14.)

[274] P. G. Wojahn, C. M. Neuwirth, and B. Bullock. Effects of interfaces for annotation on communication in a collaborative task. In *CHI '98*, pages 456–463, 1998. (Cited on page 35.)

[275] Wordpress. Wordpress sites in the world, 2013. URL `http://en.wordpress.com/stats/`. (Cited on page 2.)

[276] Z. Xu, K. Yu, V. Tresp, X. Xu, and J. Wang. Representative sampling for text classification using support vector machines. In *ECIR '03*, pages 393–407. Springer, 2003. (Cited on page 143.)

[277] Z. Xu, R. Akella, and Y. Zhang. Incorporating diversity and density in active learning for relevance feedback. In *ECIR '07*, pages 246–257. Springer, 2007. (Cited on page 143.)

[278] C. Yang, S. Bhattacharya, and P. Srinivasan. Lexical and machine learning approaches toward online reputation management. In *CLEF '13 (Online Working Notes/Labs/Workshop)*, 2012. (Cited on page 15.)

[279] S. R. Yerva, Z. Miklós, and K. Aberer. It was easy, when apples and blackberries were only fruits. In M. Braschler, D. Harman, and E. Pianta, editors, *CLEF '10 (Online Working Notes/Labs/Workshop)*, 2010. (Cited on pages 21 and 141.)

[280] H. Zaragoza. Some of the problems and applications of opinion analysis. Invited Talk at SIGIR '13 Industrial Track (Dublin), July 2013. (Cited on pages 15 and 32.)

[281] C. Zhai and J. Lafferty. Model-based feedback in the language modeling approach to information retrieval. In *CIKM '01*, pages 403–410, 2001. (Cited on page 93.)

[282] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. *ACM Transactions on Information Systems*, 22(2):179–214, 2004. (Cited on page 17.)

[283] W. Zhang and C. Yu. UIC at TREC 2006 blog track. In *TREC 2006*, 2006. (Cited on page 19.)

[284] D. Zhao and M. B. Rosson. How and why people Twitter: The role that micro-blogging plays in informal communication at work. In *GROUP '09*, pages 243–252, 2009. (Cited on page 10.)

[285] J. Zhu, H. Wang, and B. Tsou. A density-based re-ranking technique for active learning for data annotations. In *Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy*, volume 5459, pages 1–10. Springer Berlin Heidelberg, 2009. (Cited on page 143.)

[286] X. Zhu, P. Zhang, X. Lin, and Y. Shi. Active learning from data streams. In *ICDM '07*, pages 757–762, 2007. (Cited on page 143.)

[287] X. Zhu, P. Zhang, X. Lin, and Y. Shi. Active learning from stream data using optimal weight classifier ensemble. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(6):1607–1621, Dec 2010. (Cited on page 143.)

[288] A. Zubiaga, D. Spina, E. Amigó, and J. Gonzalo. Towards real-time summarization of scheduled events from twitter streams. In *HT '2012*, pages 319–320. ACM, 2012. (Cited on page 146.)

Samenvatting

Sociale media zijn een integraal onderdeel van de maatschappij. Met alomtegenwoordige mobiele apparaten kunnen ervaringen direct gedeeld worden. Deze ervaringen kunnen over merken of andere entiteiten gaan. Voor analisten van sociale media kan een collectie van berichten die een merk noemen dienen als een vergrootglas voor de dominante mening over een merk. Daarom is de globale inschatting van de reputatie van een merk meer en meer gebaseerd op de aggregatie van de polariteit van reputatie in sociale media berichten. Het bepalen van deze polariteit wordt momenteel met de hand gedaan, echter met de dramatische toename van sociale mediagebruik is dit niet langer haalbaar.

Dit proefschrift beoogt het proces van het inschatten van de reputatie van een merk te faciliteren en te automatiseren. Wij motiveren dit met gebruikersstudies onder experts in sociale media analyse. We analyseren drie datasets: een vragenlijst, log data afkomstig uit een applicatie voor handmatige annotatie en videobeelden van experts die het think-aloud protocol volgen. De beslissing bij annotaties blijkt het meest te worden beïnvloed door de online en offline autoriteit van de gebruiker die een bericht deelt. Deze online en offline autoriteit is daarom een sterke indicatie voor de polariteit van reputatie. Daarnaast geven experts aan dat zij automatisering van zoek- en filtertaken verwelkomen. Voor deze taken, als ook voor meerdere indicatoren, blijkt achtergrond informatie essentieel. Gebaseerd op de indicatoren die worden gebruikt voor handmatige annotatie ontwikkelen wij algoritmes voor het automatisch inschatten van de polariteit van reputatie. In tegenstelling tot eerdere statische evaluatie scenarios volgen wij een dynamisch scenario, dat de dagelijks werkstroom van sociale media analisten nabootst. Onze algoritmes zijn succesvol, omdat we onderscheid maken tussen reputatie en sentiment.

De motivatie voor het tweede deel van dit proefschrift is de wens van analisten om het zoeken in en filteren van nieuwe media te automatiseren. Wij beschrijven twee verbeteringen aan bestaande zoekalgoritmes. De eerste verbetering is gebaseerd op het identificeren van plotselinge uitbarstingen (bursts) met behulp van tijdreeksanalyse over pseudo-relevante documenten. We nemen een steekproef uit de termen in deze bursts voor het modelleren van queries. Dit verbetert de effectiviteit van zoekalgoritmes in nieuws en blog corpora. Ten tweede is nieuwheid (recency) een belangrijk aspect van relevantie in sociale media. Geïnspireerd door de herinneringsmodellen uit de cognitieve wetenschap stellen wij nieuwe a priori-kansen (priors) voor de relevantie van documenten voor. Wij laten zien dat priors gebaseerd op deze herinneringsmodellen effectiever, efficiënter en plausibeler zijn dan de veelgebruikte temporele priors. Achtergrondinformatie is essentieel voor het filteren van informatie. Daarnaast zijn onderwerpen rondom een entiteit dynamisch. Voor consistent sterke resultaten van filteralgoritmes is daarom de expertise van sociale media analisten vereist. De filteralgoritmes die worden gepresenteerd in dit proefschrift zijn daarom gebaseerd op active learning: wanneer een document niet met hoge zekerheid kan worden geclassificeerd wordt een handmatige annotatie van de analist gevraagd. Met behulp van intuïties over bursts en cognitieve priors voor het nemen van een steekproef uit documenten die lastig te classificeren zijn hebben we erg weinig hulp van analisten nodig om hoge effectiviteit te bereiken.

We concluderen dat veel aspecten van de annotatie van reputatie geautomatiseerd kunnen wordenspecifiek met behulp van tijdreeksanalyse, herinneringsmodellen en beperkte hulp van experts in de analyse van sociale media.

Zusammenfassung

In den letzten Jahren sind soziale Medien unter anderem durch die Mobilisierung der Technologien allgegenwärting geworden. Unmittelbares Teilen von Meinungen über Produkte ist ein Schatz für Meinungsforscher: sie können nun, ohne die sonst üblichen Fragebögen, direkt aus Äusserungen von Stakeholdern Meinungen folgern und aggregieren. Bisher war dies zumeist manuelle Arbeit; mit den stets wachsenden Datenmengen ist dies nicht mehr machbar.

Diese Dissertation erzielt eine Automatisierung der Extraktion der Reputation von Entitäten aus sozialen Medien. Zunächst analysieren wir die Indikatoren die bei der Annotation von Reputation verwendet werden. Die Daten für diese Analyse basieren auf Fragebögen, Logdaten einer Annotationsbenutzeroberfläche und Videos (der Think-Aloud Methode folgend) von Annotatoren: Social Media Analysten, also Experten. Hauptresultat dieser Studie ist, dass sowohl die Stellung (online und offline) des Autors einer Äusserung als auch die Reichweite derselben die Entscheidungsfindung bei der Annotation erheblich beeinflussen. Desweiteren wünschen sich die Analysten eine Automatisierung der Informationsbeschaffung (information retrieval) und -filterung. Hintergrundinformationen sind sowohl für die Informationsbeschaffung und -filterung, als auch für die genaue Bewertung der Indikatoren essentiell.

Ergänzend entwickeln wir Algorithmen, die auf den Indikatoren der Analysten basieren. Frühere Algorithmen wurden mit statischen Simulationsdaten evaluiert. Basierend auf ebendiesen Daten entwerfen wir ein neues Szenario, das die dynamische Entwicklung der Themen um eine Entität simuliert. Unsere Algorithmen sind erfolgreich, da sie zwischen Reputation und Sentiment unterscheiden.

Motiviert durch das Bedürfnis der Analysten, beschäftigt sich der zweite Teil dieser Dissertation mit der Automatisierung von Informationsbeschaffung und -filterung. Zunächst präsentieren wir Methoden um plötzliche Ausbrüche (bursts) von bestimmten Themen in bestehende Algorithmen zu integrieren. Diese plötzlichen Ausbrüche werden mittels Zeitreihenanalysen pseudo-relevanter Dokumente identifiziert. Die ursprünglichen Suchanfragen werden dann mithilfe der Terme aus den Ausbrüchen neu modelliert. Gute Resultate auf Zeitungs- und Blogartikeln unterstützen unsere Herangehensweise.

Nun widmen wir uns der Bevorzugung von neueren Dokumenten, da grade in sozialen Medien alte Dokumente zu Irrelevanz tendieren. Inspiriert durch Erinnerungsmodelle der Kognitiven Psychologie, zeigen wir dass eine A-priori-Wahrscheinlichkeit basierend auf diesen Modellen besser funktioniert als arbiträre Wahrscheinlichkeitsmodelle.

Da Analysten Hintergrundwissen für die Informationsfilterung verwenden und die Themenbereiche um eine Entität sich stetig verändern, wird ihre Expertise bei kontinuierlicher und alltäglicher Nutzung der Algorithmen benötigt. Die in dieser Dissertation verwendeten Algorithmen basieren auf aktiven Lernalgorithmen: die Algorithmen fragen Experten bei Unsicherheit in der Klassifikation um Hilfe. Mit Hilfe von Zeitreihenanalysemethoden und den A-priori-Wahrscheinlichkeiten der vorherigen Dokumenten werden nur sehr wenige manuelle Hilfestellungen (Annotationen) durch Analysten für hohe Leistungsmessgrößen benötigt.

Das Fazit dieser Dissertation ist, dass sich viele manuelle Prozesse der Annotation von Reputation automatisieren lassen—insbesondere mit Hilfe von Zeitreihenanalysen, Erinnerungsmodellen, und wenigen Hilfestellungen der Analysten.