

References Cited

- [1] Salman Ahmad, Alexis Battle, Zahan Malkani, and Sepander Kamvar. The jabberwocky programming environment for structured social computing. In *ACM User Interface Software and Technology (UIST)*, pages 53–64, 2011.
- [2] Saleema Amershi, James Fogarty, Ashish Kapoor, and Desney Tan. Overview based example selection in end user interactive concept learning. In *ACM User Interface Software and Technology (UIST)*, pages 247–256, 2009.
- [3] Saleema Amershi, James Fogarty, Ashish Kapoor, and Desney Tan. Examining multiple potential models in end-user interactive concept learning. In *ACM Human Factors in Computing Systems (CHI)*, pages 1357–1360, 2010.
- [4] Saleema Amershi, James Fogarty, and Daniel Weld. Regroup: Interactive machine learning for on-demand group creation in social networks. In *ACM Human Factors in Computing Systems (CHI)*, pages 21–30, 2012.
- [5] Saleema Amershi, Bongshin Lee, Ashish Kapoor, Ratul Mahajan, and Blaine Christian. Cuet: Human-guided fast and accurate network alarm triage. In *ACM Human Factors in Computing Systems (CHI)*, pages 157–166, 2011.
- [6] Apache lucene. <http://lucene.apache.org/>.
- [7] Alan R Aronson and François-Michel Lang. An overview of metamap: historical perspective and recent advances. *Journal of the American Medical Informatics Association*, 17(3):229–236, 2010.
- [8] David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. Gender in twitter: Styles, stances, and social networks. *CoRR*, abs/1210.4567, 2012.
- [9] Michele Banko and Eric Brill. Scaling to very very large corpora for natural language disambiguation. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, ACL ’01, pages 26–33, 2001.
- [10] Lee Becker, George Erhart, David Skiba, and Valentine Matula. Avaya: Sentiment analysis on twitter with self-training and polarity lexicon expansion. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 333–340, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
- [11] Steven Bedrick, Russell Beckley, Brian Roark, and Richard Sproat. Robust kaomoji detection in twitter. In *Proceedings of the Second Workshop on Language in Social Media*, pages 56–64, Montréal, Canada, June 2012. Association for Computational Linguistics.
- [12] Steven Bird, Edward Loper, and Ewan Klein. *Natural Language Processing with Python*. OReilly Media, 2009.
- [13] William Black, Rob Procter, Steven Gray, and Sophia Ananiadou. A data and analysis resource for an experiment in text mining a collection of micro-blogs on a political topic. In *LREC*, 2012.
- [14] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet allocation. *J of Machine Learning Research*, 3(1):993–1022, 2003.
- [15] Tabitha Bonilla and Justin Grimmer. Elevated threat levels and decreased expectations: How democracy handles terrorist threats. *Poetics*, 41(6):650 – 669, 2013.
- [16] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D3: Data-driven documents. In *InfoVis*, pages 2301–2309, 2011.

- [17] Michael Brooks, Katie Kuksenok, Megan K. Torkildson, Daniel Perry, John J. Robinson, Taylor J. Scott, Ona Anicello, Ariana Zukowski, Paul Harris, and Cecilia R. Aragon. Statistical affect detection in collaborative chat. In *ACM Computer-Supported Cooperative Work (CSCW)*, pages 317–328, 2013.
- [18] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, and K. D. Mandl. Surveillance sans frontieres: Internet-based emerging infectious disease intelligence and the healthmap project. *PLoS Medicine*, 2008.
- [19] Nan Cao, Jimeng Sun, Yu-Ru Lin, D. Gotz, Shixia Liu, and Huamin Qu. FacetAtlas: Multifaceted visualization for rich text corpora. In *InfoVis*, pages 1172–1181, 2010.
- [20] S. K. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction*. Erlbaum, 1983.
- [21] Alexandra Carmichael. Crowdsourced Health Confirms Infertility-Asthma Finding, September 2009.
- [22] H. A. Carneiro and E. Mylonakis. Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical infectious diseases*, 2009.
- [23] K. Chen, A. Kannan, Y. Yano, J. M. Hellerstein, and T. S. Parikh. Shreddr: pipelined paper digitization for low-resource organizations. In *ACM Computing for Development (DEV)*, 2012.
- [24] Lydia B. Chilton, Greg Little, Darren Edge, Daniel S. Weld, and James A. Landay. Cascade: Crowdsourcing taxonomy creation. In *ACM Human Factors in Computing Systems (CHI)*, pages 1999–2008, 2013.
- [25] Jason Chuang, Sonal Gupta, Christopher D. Manning, and Jeffrey Heer. Topic model diagnostics: Assessing domain relevance via topical alignment. In *ICML*, 2013.
- [26] Jason Chuang, Christopher D. Manning, and Jeffrey Heer. Termite: Visualization techniques for assessing textual topic models. In *AVI*, pages 74–77, 2012.
- [27] Jason Chuang, Christopher D. Manning, and Jeffrey Heer. ”without the clutter of unimportant words”: Descriptive keyphrases for text visualization. *ACM Trans. on Computer-Human Interaction*, 19:1–29, 2012.
- [28] Jason Chuang, Daniel Ramage, Christopher D. Manning, and Jeffrey Heer. Interpretation and trust: Designing model-driven visualizations for text analysis. In *CHI*, pages 443–452, 2012.
- [29] P.J. Crossno, D.M. Dunlavy, and T.M. Shedd. LSAView: A tool for visual exploration of latent semantic modeling. In *VAST*, pages 83–90, 2009.
- [30] P. Dai, Mausam, and D. S. Weld. Artificial intelligence for artificial, artificial intelligence. In *AAAI*, 2011.
- [31] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. A computational approach to politeness with application to social factors. In *ACL*, 2013.
- [32] Steven Dow, Anand Kulkarni, Scott Klemmer, and Björn Hartmann. Shepherding the crowd yields better work. In *ACM Computer Supported Cooperative Work (CSCW)*, pages 1013–1022, 2012.
- [33] G. Druck, G. Mann, and A. McCallum. Learning from labeled features using generalized expectation criteria. In *Proc. ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 595–602. ACM Press, 2008.
- [34] G. Druck and A. McCallum. Toward interactive training and evaluation. In *Proc. ICKM*, pages 947–956, 2011.
- [35] G. Druck, B. Settles, and A. McCallum. Active learning by labeling features. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 81–90. ACL Press, 2009.
- [36] Adnan Duric and Fei Song. Feature selection for sentiment analysis based on content and syntax models. *Decis. Support Syst.*, 53(4):704–711, November 2012.

- [37] Jacob Eisenstein. Phonological factors in social media writing. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 11–19, Atlanta, Georgia, June 2013. Association for Computational Linguistics.
- [38] James Fogarty, Desney Tan, Ashish Kapoor, and Simon Winder. Cueflik: Interactive concept learning in image search. In *ACM Human Factors in Computing Systems (CHI)*, pages 29–38, 2008.
- [39] K. Ganchev, J. Graça, J. Gillenwater, and B. Taskar. Posterior Regularization for Structured Latent Variable Models. *Journal of Machine Learning Research*, 2010.
- [40] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *Proceedings of the Twenty-eight International Conference on Machine Learning, ICML*, 2011.
- [41] Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. Identifying sarcasm in twitter: A closer look. In *ACL-HLT*, pages 581–586, 2011.
- [42] Spence Green, Jeffrey Heer, and Christopher D. Manning. The efficacy of human post-editing for language translation. In *ACM Human Factors in Computing Systems (CHI)*, 2013.
- [43] Brynjar Gretarsson, John O’Donovan, Svetlin Bostandjiev, Tobias Höllerer, Arthur Asuncion, David Newman, and Padhraic Smyth. TopicNets: Visual analysis of large text corpora with topic modeling. *ACM Trans on Intelligent Systems and Technology*, 3(2):1–26, 2012.
- [44] Thomas L. Griffiths and Mark Steyvers. Finding scientific topics. *PNAS: Proceedings of the National Academy of Sciences*, 101(1):5228–5235, 2004.
- [45] Justin Grimmer and Brandon M Stewart. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3):267–297, 2013.
- [46] A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *Intelligent Systems, IEEE*, 24(2):8–12, 2009.
- [47] David Hall, Daniel Jurafsky, and Christopher D. Manning. Studying the history of ideas using topic models. In *EMNLP*, pages 363–371, 2008.
- [48] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The WEKA data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.
- [49] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.
- [50] Marti A Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics-Volume 2*, pages 539–545. Association for Computational Linguistics, 1992.
- [51] Jeffrey Heer and Maneesh Agrawala. Design considerations for collaborative visual analytics. *Information Visualization Journal*, 7:49–62, 2008.
- [52] Jeffrey Heer and Michael Bostock. Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In *ACM Human Factors in Computing Systems (CHI)*, pages 203–212, 2010.
- [53] Jeffrey Heer, Stuart K. Card, and James A. Landay. prefuse: A toolkit for interactive information visualization. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 421–430, New York, NY, USA, 2005. ACM Press.
- [54] Jeffrey Heer and Adam Perer. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. In *IEEE Visual Analytics Science & Technology (VAST)*, 2011.

- [55] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 1029–1038, New York, USA, 2007. ACM Press.
- [56] Alfred Inselberg and Bernard Dimsdale. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In *IEEE Visualization*, pages 361–378, 1990.
- [57] Panos Ipeirotis. Mechanical turk: The demographics, 2008. <http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html>.
- [58] Clement Jonquet, Nigam H Shah, and Mark A Musen. The open biomedical annotator. *Summit on translational bioinformatics*, 2009:56, 2009.
- [59] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Wrangler: Interactive visual specification of data transformation scripts. In *ACM Human Factors in Computing Systems (CHI)*, 2011.
- [60] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Enterprise data analysis and visualization: An interview study. In *IEEE Visual Analytics Science & Technology (VAST)*, 2012.
- [61] Sean Kandel, Ravi Parikh, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. Profiler: Integrated statistical analysis and visualization for data quality assessment. In *Advanced Visual Interfaces*, 2012.
- [62] Hyunmo Kang, Lise Getoor, Ben Shneiderman, Mustafa Bilgic, and Louis Licamele. Interactive entity resolution in relational data: A visual analytic tool and its evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 14(5):999–1014, September 2008.
- [63] D. R. Karger, S. Oh, and D. Shah. Iterative learning for reliable crowdsourcing systems. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1953–1961, 2011.
- [64] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th International Conference on Computational Linguistics, COLING '04*, 2004.
- [65] Gary King, Jennifer Pan, and Margaret E. Roberts. How censorship in china allows government criticism but silences collective expression. *American Political Science Review*, 107:1–18, 2013.
- [66] A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with mechanical turk. In *ACM Human Factors in Computing Systems (CHI)*, pages 453–456, 2008.
- [67] Aniket Kittur, Susheel Khamkar, Paul André, and Robert Kraut. Crowdweaver: Visually managing complex crowd work. In *ACM Computer Supported Cooperative Work (CSCW)*, pages 1033–1036, 2012.
- [68] Aniket Kittur, Boris Smus, Susheel Khamkar, and Robert E. Kraut. Crowdforge: Crowdsourcing complex work. In *ACM User Interface Software and Technology (UIST)*, pages 43–52, 2011.
- [69] Nicholas Kong, Jeffrey Heer, and Maneesh Agrawala. Perceptual guidelines for creating rectangular treemaps. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2010.
- [70] Zornitsa Kozareva, Eduard H. Hovy, and Ellen Riloff. Learning and evaluating the content and structure of a term taxonomy. In *AAAI Spring Symposium: Learning by Reading and Learning to Read*, pages 50–57, 2009.
- [71] Anand Kulkarni, Matthew Can, and Björn Hartmann. Collaboratively crowdsourcing workflows with turkomatic. In *ACM Computer Supported Cooperative Work (CSCW)*, pages 1003–1012, 2012.
- [72] Greg Little, Lydia B. Chilton, Max Goldman, and Robert C. Miller. TurKit: Human computation algorithms on mechanical turk, 2010.

- [73] Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28(4):587–604, 2002.
- [74] Yucheng Low, Danny Bickson, Joseph Gonzalez, Carlos Guestrin, Aapo Kyrola, and Joseph M. Hellerstein. Distributed GraphLab: A framework for machine learning and data mining in the cloud. *Proc. VLDB Endow.*, 5(8):716–727, April 2012.
- [75] Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, and Joseph M. Hellerstein. Graphlab: A new parallel framework for machine learning. In *Uncertainty in Artificial Intelligence (UAI)*, July 2010.
- [76] Edwin Lughofer. Hybrid active learning for reducing the annotation effort of operators in classification systems. *Pattern Recognition*, 45(2):884–896, 2012. Uses unsupervised methods. Does not incorporate domain expert input.
- [77] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Association for Computational Linguistics (ACL)*, 2011.
- [78] D. L. MacLean and J. Heer. Identifying medical terms in patient-authored text: a crowdsourcing-based approach. *JAMIA*, 2013.
- [79] Patricia Yancey Martin and Barry A. Turner. Grounded theory and organizational research. *The Journal of Applied Behavioral Science*, 22(2):141–157, 1986.
- [80] Daniel A. McFarland, Daniel Ramage, Jason Chuang, Jeffrey Heer, Christopher D. Manning, and Dan Jurafsky. Differentiating language usage through topic models. *Poetics*, 41:607–625, 2013.
- [81] P. Melville and V. Sindhwani. Active dual supervision: Reducing the cost of annotating examples and features. In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, pages 49–57. Association for Computational Linguistics, 2009.
- [82] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems (NIPS)*, 2013.
- [83] Thomas Mühlbacher and Harald Piringer. A partition-based framework for building and validating regression models. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):1962–1971, 2013.
- [84] Arjun Mukherjee, Bing Liu, and Natalie Glance. Spotting fake reviewer groups in consumer reviews. In *WWW*, pages 191–200, 2012.
- [85] Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. Semeval-2013 task 2: Sentiment analysis in twitter. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 312–320, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
- [86] R. M. Neal and G. E. Hinton. A new view of the EM algorithm that justifies incremental, sparse and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer, 1998.
- [87] David Newman, Youn Noh, Edmund Talley, Sarvnaz Karimi, and Timothy Baldwin. Evaluating topic models for digital libraries. In *JCDL*, pages 215–224, 2010.
- [88] Alexandros Ntoulas, Marc Najork, Mark Manasse, and Dennis Fetterly. Detecting spam web pages through content analysis. In *WWW*, pages 83–92, 2006.

- [89] Fredrik Olsson. *Bootstrapping Named Entity Annotation by means of Active Machine Learning A Method for Creating Corpora*. PhD thesis, University of Gothenburg, 2008.
- [90] Fredrik Olsson. A literature survey of active machine learning in the context of natural language processing. Sics report, Swedish Institute of Computer Science, 2009.
- [91] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL*, pages 115–124, 2005.
- [92] Chintan Patel, Kaustubh Supekar, Yugyung Lee, and E. K. Park. Ontokhoj: A semantic web portal for ontology searching, ranking and classification. In *In Proc. 5th ACM Int. Workshop on Web Information and Data Management*, pages 58–61, 2003.
- [93] Kayur Patel, Naomi Bancroft, Steven M. Drucker, James Fogarty, Andrew J. Ko, and James Landay. Gestalt: Integrated support for implementation and analysis in machine learning. In *UIST*, pages 37–46, 2010.
- [94] Pew Internet & American life project: Health online 2013. <http://pewinternet.org/Reports/2013/Health-online/Summary-of-Findings.aspx>. [Online; accessed 2-April-2013].
- [95] Wanda Pratt and Meliha Yetisgen-Yildiz. A study of biomedical concept identification: Metamap vs. people. In *AMIA Annual Symposium Proceedings*, volume 2003, page 529. American Medical Informatics Association, 2003.
- [96] H. Raghavan, O. Madani, and R. Jones. Active learning with feedback on both features and instances. *Journal of Machine Learning Research*, 7:1655–1686, 2006. Boosting of features, not prior or regularization.
- [97] Daniel Ramage, David Hall, Ramesh Nallapati, and Christopher D. Manning. Labeled LDA: A supervised topic model for credit attribution in multi-label corpora. In *EMNLP*, pages 248–256, 2009.
- [98] Daniel Ramage, Evan Rosen, Jason Chuang, Christopher D. Manning, and Daniel A. McFarland. Topic modeling for the social sciences. In *NIPS Workshop on Applications of Topic Models*, 2009.
- [99] Reuters-21578 data set. <http://www.daviddlewis.com/resources/testcollections/reuters21578/>. [Online; accessed 4-January-2014].
- [100] E. Ringger, M. Carmen, R. Haertel, K. Seppi, D. Lonsdale, P. McClanahan, J. Carroll, and N. Ellison. Assessing the costs of machine-assisted corpus annotation through a user study. In *Proc. International Conference on Language Resources and Evaluation (LREC)*. European Language Resources Association, 2008.
- [101] Molly Roberts, Brandon Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Bethany Albertson, Shana Gadarian, and David Rand. Structural topic models for open ended survey responses. *American Journal of Political Science*, Forthcoming.
- [102] Jeffrey Rzeszutarski and Aniket Kittur. Crowdscape: Interactively visualizing user behavior and output. In *ACM User Interface Software and Technology*, pages 55–62, 2012.
- [103] Tyler Schnoebelen. Do you smile with your nose? stylistic variation in twitter emoticons. *University of Pennsylvania Working Papers in Linguistics*, 18, 2012.
- [104] H. Andrew Schwartz, Johannes C. Eichstaedt, Margaret L. Kern, Lukasz Dziurzynski, Stephanie M. Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin E. P. Seligman, and Lyle H. Ungar. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, 8(9), 09 2013.

- [105] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2634–2643, 2013.
- [106] B. Settles. Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1467–1478. ACL Press, 2011.
- [107] B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1069–1078. ACL Press, 2008.
- [108] B. Settles, M. Craven, and S. Ray. Multiple-instance active learning. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1289–1296. MIT Press, 2008.
- [109] B. Settles and X. Zhu. Behavioral factors in interactive training of text classifiers. In *Proc. North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT)*, pages 563–567. ACL Press, 2012.
- [110] Burr Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009.
- [111] Ben Shneiderman and Catherine Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)*, pages 1–7, 2006.
- [112] V. Sindhwani, P. Melville, and R.D. Lawrence. Uncertainty sampling and transductive experimental design for active dual supervision. In *Proc. ICML*, 2009.
- [113] R. Snow, B. OConnor, D. Jurafsky, and A. Ng. Cheap and fast—but is it good? evaluating non-expert annotations for natural language tasks. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 254–263. ACL Press, 2008.
- [114] B. Snyder and R. Barzilay. Multiple aspect ranking using the good grief algorithm. In *Proc. HLT-NAACL*, 2007.
- [115] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, 2013.
- [116] Stanford corenlp: A suite of core nlp tools. <http://nlp.stanford.edu/software/corenlp.shtml>.
- [117] J. Stasko, C. Görg, Zhicheng Liu, and K. Singhal. Jigsaw: Supporting investigative analysis through interactive visualization. In *VAST*, pages 131–138, 2007.
- [118] Philip Stone. Thematic text analysis: new agendas for analyzing text content. In Carl Roberts, editor, *Text Analysis for the Social Sciences*. Lawrence Erlbaum Associates, 1997.
- [119] Carlo Strapparava and Rada Mihalcea. Semeval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [120] Justin Talbot, Bongshin Lee, Ashish Kapoor, and Desney S. Tan. Ensemblematrix: Interactive visualization to support machine learning with multiple classifiers. In *ACM Human Factors in Computing Systems (CHI)*, CHI '09, pages 1283–1292, New York, NY, USA, 2009. ACM.
- [121] Edmund M. Talley, David Newman, David Mimno, Bruce W. Herr, Hanna M. Wallach, Gully A. P. C. Burns, A. G. Miriam Leenders, and Andrew McCallum. Database of NIH grants using machine-learned categories and graphical clustering. *Nature Methods*, 8(6):443–444, 2011.

- [122] Yla R. Tausczik and James W. Pennebaker. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1):24–54, 2010.
- [123] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proc. WWW*, pages 111–120, 2008.
- [124] John W. Tukey and Martin B. Wilk. Data analysis and statistics: An expository overview. In Lyle V. Jones, editor, *The Collected Works of John W. Tukey Volume IV: Philosophy and Principles of Data Analysis, 1965-1986*. Wadsworth & Brooks/Cole, 1966.
- [125] A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welp. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM*, 2010.
- [126] Twenty newsgroups data set. <http://qwone.com/~jason/20Newsgroups/>. [Online; accessed 4-January-2014].
- [127] L. J. P. van der Maaten and G.E. Hinton. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 2008.
- [128] Y. Wang, R. E. Kraut, and J. M. Levine. To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *ACM Computer-Supported Cooperative Work (CSCW)*, 2012.
- [129] Furu Wei, Shixia Liu, Yangqiu Song, Shimei Pan, Michelle X. Zhou, Weihong Qian, Lei Shi, Li Tan, and Qiang Zhang. TIARA: a visual exploratory text analytic system. In *KDD*, pages 153–162, 2010.
- [130] R. W. White, N. P. Tatonetti, N. H. Shah, R. B. Altman, and E. Horvitz. Web-scale pharmacovigilance: listening to signals from the crowd. *JAMIA*, 2013.
- [131] Casey Whitelaw, Navendu Garg, and Shlomo Argamon. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management, CIKM '05*, 2005.
- [132] P. Wicks, T. E. Vaughan, M. P. Massagli, and J. Heywood. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. *Nature Biotechnology*, 2011.
- [133] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. Strategies for crowdsourcing social data analysis. In *ACM Human Factors in Computing Systems (CHI)*, 2012.
- [134] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. Commentspace: Structured support for collaborative visual analysis. In *ACM Human Factors in Computing Systems (CHI)*, 2011.