

Selected References

Online Materials

Tidy Text Mining with R by Julia Silge and David Robinson <http://tidytextmining.com/>

Introduction to Text Mining (Books)

Miner, Gary. *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press, 2012.

Jockers, Matthew. *Text Analysis with R for Students of Literature*. Springer, 2014.

Articles that use TM or provide important insight/critiques.

- Bail, Christopher A. "The cultural environment: Measuring culture with big data." *Theory and Society* 43, no. 3-4 (2014): 465-482.
- Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L. Boyd-Graber, and David M. Blei. "Reading tea leaves: How humans interpret topic models." In *Advances in neural information processing systems*, pp. 288-296. 2009.
- DiMaggio, Paul, Manish Nag, and David Blei. "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding." *Poetics* 41, no. 6 (2013): 570-606.
- Fligstein, Neil, Jonah S. Brundage, and Michael Schultz. "Why the Federal Reserve Failed to See the Financial Crisis of 2008: The Role of "Macroeconomics" as a Sense making and Cultural Frame." (2014).
- Gentzkow, Matthew, and Jesse M. Shapiro. "What drives media slant? Evidence from US daily newspapers." *Econometrica* 78, no. 1 (2010): 35-71.
- Grimmer, Justin. "Appropriators not position takers: The distorting effects of electoral incentives on congressional representation." *American Journal of Political Science* 57, no. 3 (2013): 624-642.
- Hargittai, Eszter. "Is bigger always better? Potential biases of big data derived from social network sites." *The ANNALS of the American Academy of Political and Social Science* 659, no. 1 (2015): 63-76.
- Ignatow, Gabe. "Theoretical foundations for digital text analysis." *Journal for the Theory of Social Behaviour* (2015).
- Jacobi, Carina, Wouter van Atteveldt, and Kasper Welbers. "Quantitative analysis of large amounts of journalistic texts using topic modelling." *Digital Journalism* 4, no. 1 (2016): 89-106.
- Jelveh, Zubin, Bruce Kogut, and Suresh Naidu. "Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics." In *EMNLP*, pp. 1804-1809. 2014.
- Lee, Monica, and John Levi Martin. "Coding, counting and cultural cartography." *American Journal of Cultural Sociology* 3, no. 1 (2015): 1-33.
- Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. "Meme-tracking and the dynamics of the news cycle." In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 497-506. ACM, 2009.
- Loughran, Tim, and Bill McDonald. "When is a Liability not a Liability?." *Journal of Finance*, forthcoming (2009).
- Mohr, John W., Robin Wagner-Pacifci, Ronald L. Breiger, and Petko Bogdanov. "Graphing the grammar of motives in National Security Strategies: Cultural interpretation, automated text analysis and the drama of global politics." *Poetics* 41, no. 6 (2013): 670-700.
- Mohr, John W., and Petko Bogdanov. "Introduction—Topic models: What they are and why they matter." *Poetics* 41, no. 6 (2013): 545-569.

- Niculae, Vlad, Srijan Kumar, Jordan Boyd-Graber, and Cristian Danescu-Niculescu-Mizil. "Linguistic Harbingers of Betrayal: A Case Study on an Online Strategy Game." *arXiv preprint arXiv:1506.04744* (2015).
- Resnik, Philip, Anderson Garron, and Rebecca Resnik. "Using topic modeling to improve prediction of neuroticism and depression." In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language*, pp. 1348-1353. Association for Computational Linguistics, 2013.
- Saavedra, Serguei, Kathleen Hagerty, and Brian Uzzi. "Synchronicity, instant messaging, and performance among financial traders." *Proceedings of the National Academy of Sciences* 108, no. 13 (2011): 5296-5301.
- Yu, Dian, Yulia Tyshchuk, Heng Ji, and W. A. Wallace. "Detecting deceptive groups using conversations and network analysis." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL*, pp. 26-31. 2015.

Latent Dirichlet Allocation References

- Blei D.M., Ng A.Y., Jordan M.I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Phan X.H., Nguyen L.M., Horguchi S. (2008). Learning to Classify Short and Sparse Text & Web with Hidden Topics from Large-scale Data Collections. In *Proceedings of the 17th International World Wide Web Conference (WWW 2008)*, pages 91–100, Beijing, China.
- Lu, B., Ott, M., Cardie, C., Tsou, B.K. (2011). Multi-aspect Sentiment Analysis with Topic Models. In *Proceedings of the 2011 IEEE 11th International Conference on Data Mining Workshops*, pages 81–88.

Multiverse Analysis

- Steege, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. "Increasing Transparency Through a Multiverse Analysis." *Perspectives on Psychological Science* 11, no. 5 (2016): 702-712.