

Bibliography

Bioinformatics Papers

*

May 15, 2017

References

- [1] Naoki Abe and Hiroshi Mamitsuka. Predicting protein secondary structure using stochastic tree grammars. *ML*, 29(2-3):275–301, November 1997.
- [2] Syed Toufeeq Ahmed, Deepthi Chidambaram, Hasan Davulcu, and Chitta Baral. InTex: A syntactic role driven protein-protein interaction extractor. In *ISMB-WS2005A*, pages 54–61, 2005.
- [3] Hisham Al-Mubaid and Rajit K Singh. A new text mining approach for finding protein-to-disease associations. *American Journal of Biochemistry and Biotechnology*, 1(3):145–152, 2005.
- [4] Blaise T. Alako, Antoine Veldhoven, Sjozef van Baal, Rob Jelier, Stefan Verhoeven, Ton Rullmann, Jan Polman, and Guido Jenster. Copub mapper: mining medline based on search term co-publication. *BMC Bioinformatics*, 6(1), March 2005.
- [5] Reka Albert. Scale-free networks in cell biology. *J Cell Sci*, 118(21):4947–4957, 2005.
- [6] Beatrice Alex, Barry Haddow, and Claire Grover. Recognising nested named entities in biomedical text. In *BioNLP workshop at ACL 2007*, pages 65–72. ACL, 2007.
- [7] M. A. Andrade and P. Bork. Automated extraction of information in molecular biology. *FEBS Lett*, 476(1-2):12–17, June 2000.
- [8] M. A. Andrade and A. Valencia. Automatic annotation for biological sequences by extraction of keywords from MEDLINE abstracts - development of a prototype system. In *ISMB1997*, pages 25–32, 1997.
- [9] Arnon Arazi, Eshel Ben-Jacob, and Uri Yechiali. Bridging genetic networks and queueing theory. *Physica A*, 332:585–616, 2004.

- [10] Teresa K. Attwood, Paul Bradley, Walter Daelemans, Luc Dehaspe, Frederique Durant, Melanie Hilario, Jee-Hyub Kim, Alex L. Mitchell, Johann Petrak, Violaine Pillet, Alexander K. Seewald, Kristof Van Belleghem, Anne-Lise Veuthey, and Marc Zehnder. BioMinT: A database curator's assistant for biomedical text processing. In *ISMB-WS2004A*, 2004.
- [11] G. D. Bader and C.W.V. Hogue. Analyzing yeast protein-protein interaction data obtained from different sources. *Nature Biotechnology*, 20:991–997, October 2002.
- [12] Timothy L. Bailey and Charles Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Russ B. Altman, Douglas L. Brutlag, Peter D. Karp, Richard H. Lathrop, and David B. Searls, editors, *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology (ISMB '94)*, pages 28–36, Stanford, California, USA, August 14–17, 1994. AAAI.
- [13] Pierre Baldi, Søren Brunak, Yves Chauvin, Jacob Engelbrecht, and Anders Krogh. Periodic sequence patterns in human exons. In *ISMB1995*, pages 30–38, 1995.
- [14] Nuno Bandeira, Dekel Tsur, Ari Frank, and Pavel A. Pevzner. Protein identification by spectral networks analysis. *Proceedings of the National Academy of Sciences*, 104(15):6140–6145, 2007.
- [15] Carol A. Bean and Milton Corn. Extramural funding opportunities in bioinformatics from the National Library of Medicine. *JASIST*, 56(5):551–556, March 2005.
- [16] Asa Ben-Hur and William Stafford Noble. Kernel methods for predicting protein-protein interactions. In *ISMB2005*, pages 38–46, 2005.
- [17] C. Blaschke and A. Valencia. The potential use of SUISEKI as a protein interaction discovery tool. *Genome Inform Ser Workshop Genome Inform.*, 12:123–134, 2001.
- [18] C. Blaschke and A. Valencia. The frame-based module of the SUISEKI information extraction system. *IEEE Intelligent Systems*, 17:14–20, 2002.
- [19] Christian Blaschke, Miguel A. Andrade, Christos A. Ouzounis, and Alfonso Valencia. Automatic extraction of biological information from scientific text: Protein-protein interactions. In *ISMB1999*, pages 60–67, 1999.
- [20] Christian Blaschke, Miguel A Andrade, Christos A Ouzounis, and Alfonso Valencia. Automatic extraction of biological information from scientific text: protein-protein interactions. In *Ismb*, volume 7, pages 60–67, 1999.
- [21] Christian Blaschke and Alfonso Valencia. BioCreAtIvE: Functional extraction. In *ISMB-WS2003A*, 2003.

- [22] Jin bok Lee and Jong C. Park. Text data mining for automatic gene ontology extension. In *ISMB-WS2002A*, 2002.
- [23] Karsten M. Borgwardt, Cheng Soon Ong, Stekan Schönauer, S. V. N. Vishwanathan, Alex J. Smola, and Hans-Peter Kriegel. Protein function prediction via graph kernels. In *ISMB2005*, pages 47–56, 2005.
- [24] Peer Bork, Lars Juhl Jensen, Christian von Mering, Arun K. Ramani, Insuk Lee, and Edward M. Marcotte. Protein interaction networks from yeast to human. *Curr Opin Struct Biol*, 14(3):292–299, June 2004.
- [25] Roger Brent and Jehoshua Bruck. Can computers help to explain biology. *Nature*, 440:416–417, March 23, 2006.
- [26] Ted Briscoe. Language acquisition: The bioprogram hypothesis and the Baldwin Effect, 1997. <http://www.cl.cam.ac.uk/users/ejb/bioprogram.ps.gz>.
- [27] Razvan Bunescu, Ruifang Ge, Rohit G. Kate, Edward M. Marcotte, Raymond J. Mooney, Arun K. Ramani, and Yuk Wah Wong. Comparative experiments on learning information extractors for proteins and their interactions. *Artificial Intelligence in Medicine*, 33:139–155, 2005.
- [28] Razvan C. Bunescu and Raymond J. Mooney. A shortest path dependency kernel for relation extraction. In *HLT/EMNLP2005*, pages 724–731, October 2005.
- [29] Razvan C. Bunescu and Raymond J. Mooney. Subsequence kernels for relation extraction. In *NIPS2005*, 2005.
- [30] Razvan C. Bunescu and Raymond J. Mooney. Extracting relations from text: From word sequences to dependency paths. Forthcoming, 2006.
- [31] Atul J. Butte and Isaac S. Kohane. Creation and implications of a phenome-genome network. *Nature Biotechnology*, 24:55–62, 2006.
- [32] Atul J Butte and Isaac S Kohane. Creation and implications of a phenome-genome network. *Nature biotechnology*, 24(1):55–62, 2006.
- [33] David Campos, Sergio Matos, and Jose Oliveira. Gimli: open source and high-performance biomedical name recognition. *BMC Bioinformatics*, 14(1):54, 2013.
- [34] Gabriel Cardona, Francesc Rossello, and Gabriel Valiente. A perl package and an alignment tool for phylogenetic networks, 2007.
- [35] Jeffrey T. Chang, Hinrich Schütze, and Russ B. Altman. GAPSCORE: Finding gene and protein names one word at a time. *Bioinformatics*, 20(2):216–225, 2004.

- [36] Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*, 2001:34–301, 2001.
- [37] H. Chen and B. M. Sharp. Content-rich biological network constructed by mining pubmed abstracts. *BMC Bioinformatics*, 5(1), October 2004.
- [38] J. Chen, W. Hsu, M. L. Lee, and S. K. Ng. Increasing confidence of protein interactomes using network topological metrics. *Bioinformatics*, 22(16):1998–2004, August 2006.
- [39] J. Y. Chen, C. Shen, and A. Y. Sivachenko. Mining alzheimer disease relevant proteins from integrated protein interactome data. *Pac Symp Biocomput*, pages 367–378, 2006.
- [40] Jake Yue Chen, Changyu Shen, and Andrey Y. Sivachenko. Mining alzheimer disease relevant proteins from integrated protein interactome data. In *Pacific Symposium on Biocomputing*, pages 367–378. World Scientific Publishing Co. Pte. Ltd., 2006.
- [41] Aaron M. Cohen. Unsupervised gene/protein named entity normalization using automatically extracted dictionaries. In *ISMB-WS2005A*, 2005.
- [42] Jacques Cohen. Bioinformatics - an introduction for computer scientists. *ACM Computing Surveys*, 36(2):122–158, 2004.
- [43] Jacques Cohen. Bioinformatics an introduction for computer scientists. *ACM Computing Surveys (CSUR)*, 36(2):122–158, 2004.
- [44] K. Bretonnel Cohen, Lynne Fox, Philip V. Ogren, and Lawrence Hunter. Corpus design for biomedical natural language processing. In *ISMB-WS2005A*, 2005.
- [45] K. Bretonnel Cohen and Lawrence Hunter. Natural language processing and systems biology. In Werner Dubitzky and Francisco Azuaje, editors, *Artificial Intelligence Methods and Tools for Systems Biology*, pages 147–174. Springer Verlag, 2004.
- [46] K Bretonnel Cohen and Lawrence Hunter. Natural language processing and systems biology. In *Artificial intelligence methods and tools for systems biology*, pages 147–173. Springer, 2004.
- [47] William W. Cohen. *A Computer Scientist's Guide to Cell Biology*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.
- [48] William W Cohen. *A computer scientist's guide to cell biology*. Springer Science & Business Media, 2007.

- [49] William W. Cohen, Zhenzhen Kou, and Robert F. Murphy. Extracting information from text and images for location proteomics. In *Proceedings of the 3rd ACM SIGKDD Workshop on Data Mining in Bioinformatics (BIOKDD 2003)*, pages 2–9, 2003.
- [50] William W. Cohen, Richard Wang, and Robert F. Murphy. Understanding captions in biomedical publications. In *KDD2003*, pages 499–504, 2003.
- [51] Vittoria Colizza, Alessandro Flammini, Amos Maritan, and Alessandro Vespignani. Characterization and modeling of protein-protein interaction networks. *Physica A*, 352(1):1–27, July 2005.
- [52] James W. Cooper and Aaron Kershenbaum. Discovery of protein-protein interactions using a combination of linguistic, statistical and graphical information. *BMC Bioinformatics*, 6(1), June 2005.
- [53] James W Cooper and Aaron Kershenbaum. Discovery of protein-protein interactions using a combination of linguistic, statistical and graphical information. *BMC bioinformatics*, 6(1):1, 2005.
- [54] David P. A. Corney, Bernard F. Buxton, William B. Langdon, and David T. Jones. BioRAT: Extracting biological information from full-length papers. *Bioinformatics*, 20(17):3206–3213, 2004.
- [55] Mark Craven. Learning to extract relations from MEDLINE. In *AAAI-WS1999A*, 1999.
- [56] Mark Craven and Johan Kumlien. Constructing biological knowledge bases by extracting information from text sources. In *ISMB1999*, pages 77–86, 1999.
- [57] Noa P. Cruz, Manuel J. Magaña, Jacinto Mata Vázquez, and Victoria Pachón. A machine-learning approach to negation and speculation detection in clinical texts. *J. Am. Soc. Inf. Sci. Technol.*, 63(7):1398–1410, July 2012.
- [58] Nikolai Daraselia, Anton Yuryev, Sergei Egorov, Svetalana Novichkova, Alexander Nikitin, and Ilya Mazo. Extracting human protein interactions from MEDLINE using a full-sentence parser. *Bioinformatics*, 20(5):604–611, 2004.
- [59] Berry de Bruijn and Joel Martin. Getting to the (c)ore of knowledge: Mining biomedical literature. *International Journal of Medical Informatics*, 67(1-3):7–18, December 2002.
- [60] Antonio del Sol, Hiroto Fujihashi, and Paul O’Meara. Topology of small-world networks of protein-protein complex structures. *Bioinformatics*, 21(8), April 15, 2005.

- [61] Shipra Dingare, Malvina Nissim, Jenny Finkel, Claire Grover, and Christopher Manning. A system for identifying named entities in biomedical text: How results from two evaluations reflect on both the system and the evaluations. In *ISMB-WS2004A*, 2004.
- [62] Pavel B. Dobrokhotov, Cyril Goutte, Anne-Lisa Veuthey, and Eric Gaussier. Combining NLP and probabilistic categorisation for document and term selection for SWISS-PROT medical annotation. In *ISMB2003*, pages 91–94, 2003.
- [63] Ian Donaldson, Joel Martin, Berry de Bruijn, Cheryl Wolting, Vicki Lay, Brigitte Tuekam, Shudong Zhang, Berivan Baskin, Gary D. Bader, Katerina Michalockova, Tony Pawson, and Christopher W. V. Hogue. Pre-BIND and Textomy - mining the biomedical literature for protein-protein interactions using a support vector machine. In *ISMB-WS2003A*, 2003.
- [64] Richard Durbin, Sean R. Eddy, Anders Krogh, and Graeme Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. CUP, 1998.
- [65] Ramón A-A. Erhardt, Reinhard Schneider, and Christian Blaschke. Status of text-mining techniques applied to biomedical text. *Drug Discovery Today*, 11(7-8):315–325, April 2006.
- [66] Güneş Erkan and Dragomir R. Radev. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research (JAIR)*, 22:457–479, 2004.
- [67] Eleazar Eskin, William Noble Grundy, and Yoram Singer. Protein family classification using sparse Markov transducers. In *ISMB2000*, pages 134–145, 2000.
- [68] Eleazar Eskin, William Noble Grundy, and Yoram Singer. Using mixtures of common ancestors for estimating the probabilities of discrete events in biological sequences. In *ISMB2001*, pages 22–29, 2001.
- [69] Joseph Felsenstein. PHYLIP: Phylogeny inference package. Technical report, Department of Genome Sciences, University of Washington, 1995. <http://evolution.genetics.washington.edu/phylip/doc/main.html>.
- [70] David Fenstermacher. Introduction to bioinformatics. *JASIST*, 56(5):440–446, March 2005.
- [71] Jenny Finkel, Shipra Dingare, Christopher D. Manning, Malvina Nissim, Beatrice Alex, and Claire Grover. Exploring the boundaries: Gene and protein identification in biomedical text. *Bioinformatics*, 6, 2005.
- [72] Walter M. Fitch and Emanuel Margoliash. Construction of phylogenetic trees. *Science*, 155(3760):279–284, January 1967.

- [73] Carol Friedman, Pauline Kra, Hong Yu, Michael Krauthammer, and Andrey Rzhetsky. GENIES: A natural-language processing system for the extraction of molecular pathways from journal articles. In *ISMB2001*, pages 74–82, 2001.
- [74] Sahil Garg, Aram Galstyan, Ulf Hermjakob, and Daniel Marcu. Extracting biomolecular interactions using semantic parsing of biomedical text. *arXiv preprint arXiv:1512.01587*, 2015.
- [75] Filip Ginter, Jorma Boberg, Jouni Jarvinen, and Tapio Salakoski. New techniques for disambiguation in natural language and their application to biological text. *JMLR*, 5:605–621, June 2004.
- [76] Filip Ginter, Tapio Pahikkala, Sampo Pyysalo, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. Extracting protein-protein interaction sentences by applying rough set data analysis. In Shusaku Tsumoto, Roman Slowinski, Jan Komorowski, and et al., editors, *Proceedings of the Fourth International Conference on Rough Sets and Current Trends in Computing (RSCTC 2004)*, Lecture Notes in Artificial Intelligence , Vol. 3066, pages 780–785. Springer Verlag, 2004.
- [77] Aristides Gionis and Heikki Mannila. Finding recurrent sources in sequences. In *Seventh Annual Conference on Computational Biology (RECOMB'03)*, Berlin, Germany, 2003.
- [78] L. Giot, J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aa-nensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C.A. Stanyon, R.L. Jr. Finley, K.P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R.A. Shimkets, M.P. McKenna, J. Chant, and J.M. Rothberg. A protein interaction map of drosophila melanogaster. *Science*, 302(5651):1727–1736, Dec 5 2003.
- [79] D. S. Goldberg and F. P. Roth. Assessing experimentally derived interactions in a small world. *Proc Natl Acad Sci U S A*, 100(8):4372–4376, April 2003.
- [80] Graciela Gonzalez, Juan C. Uribe, Luis Tari, Colleen Brophy, and Chitta Baral. Mining gene-disease relationships from biomedical literature: Weighting protein-protein interactions and connectivity measures. *Pacific Symposium on Biocomputing*, 12:28–39, 2007.
- [81] Graciela Gonzalez, Juan C. Uribe, Luis Tari, Colleen Brophy, and Chitta Baral. Mining gene-disease relationships from biomedical literature: Weighting protein-protein interactions and connectivity measures. In *Pacific Symposium on Biocomputing 2007*, pages 28–39, 2007.

- [82] Claire Grover, Mirella Lapata, and Alex Lascarides. A comparison of parsing technologies for the biomedical domain. *NLE*, 11(1):27–65, March 2005.
- [83] Cigdem Gunduz, Bulent Yener, and S. Humayun Gultekin. The cell-graphs of cancer. In *ISMB2004*, pages 145–151, 2004.
- [84] Dan Gusfield. *Algorithms On Strings, Trees, and Sequences: Computer Science and Computational Biology*. CUP, 1997.
- [85] Luke Hakes, John W Pinney, David L. Robertson, and Simon C Lovell. Protein-protein interaction networks and biology-what’s the connection? *Nature Biotechnology*, 26:69–72, 2008.
- [86] Eran Halperin, Gad Kimmel, and Ron Shamir. Tag SNP selection in genotype data for maximizing SNP prediction accuracy. In *ISMB2005*, pages 195–203, 2005.
- [87] Y. Hao and et al. Discovering patterns to extract protein-protein interactions from the literature: Part II. *Bioinformatics*, 21:3294–3300, 2005.
- [88] Henk Harkema, John N. Dowling, Tyler Thornblade, and Wendy W. Chapman. Context: An algorithm for determining negation, experienter, and temporal status from clinical reports. *J. of Biomedical Informatics*, 42(5):839–851, October 2009.
- [89] Henk Harkema, Ian Roberts, Rob Gaizauskus, and Mark Hepple. A Web service for biomedical term look-up. In *ISMB-WS2004A*, 2004.
- [90] Vasileios Hatzivassiloglou, Pablo A. Duboué, and Andrey Rzhetsky. Disambiguating proteins, genes, and RNA in text: A machine learning approach. In *ISMB2001*, pages 97–106, 2001.
- [91] Marti A A. Hearst, Anna Divoli, Harendra Guturu, Alex Ksikes, Preslav Nakov, Michael A A. Wooldridge, and Jerry Ye. Biotext search engine: beyond abstract search. *Bioinformatics*, June 2007.
- [92] Bradley M. Hemminger, Trish Losi, and Anne Bauers. Survey of bioinformatics programs in the united states. *JASIST*, 56(5):529–537, March 2005.
- [93] Henning Hermjakob, Luisa Montecchi-Palazzi, Chris Lewington, Sugath Mudali, Samuel Kerrien, Sandra Orchard, Martin Vingron, Bernd Roechert, Peter Roepstorff, Alfonso Valencia, Hanah Margalit, John Armstrong, Amos Bairoch, Gianni Cesareni, David Sherman, and Rolf Apweiler. IntAct: An open source molecular interaction database. *Nucleic Acids Research*, 32:452–455, 2004.
- [94] William Hersh. TREC genomics pre-track workshop report. *SIGIR Forum*, 36(2):81–82, 2002.

- [95] L. Hirschman, A. Yeh, C. Blaschke, and A. Valencia. Overview of biocreative: critical assessment of information extraction for biology. *BMC Bioinformatics*, 6 Suppl 1, 2005.
- [96] Lynette Hirschman, Jong C. Park, Junichi Tsujii, Limsoon Wong, and Cathy H. Wu. Accomplishments and challenges in literature data mining for biology. *Bioinformatics*, 18(12):1553–1561, 2002.
- [97] R. Hoffmann and A. Valencia. Protein interaction: Same network, different hubs. *Trends Genet.*, 19(12):681–683, December 2003.
- [98] R. Hoffmann and A. Valencia. A gene network for navigating the literature. *Nature Genet.*, 36:664, 2004.
- [99] R. Hoffmann and A. Valencia. Implementing the ihop concept for navigation of biomedical literature. *Bioinformatics*, 21 Suppl 2:ii252–ii258, September 2005.
- [100] Robert Hoffmann and Alfonso Valencia. Life cycles of successful genes. *TRENDS in Genetics*, 19(2):79–81, February 2003.
- [101] Robert Hoffmann and Alfonso Valencia. Life cycles of successful genes. *Trends Genet*, pages 79–81, 2007.
- [102] Ramin Homayouni, Kevin Heinrich, Lai Wei, and Michael W. Berry. Gene clustering by latent semantic indexing of MEDLINE abstracts. *Bioinformatics*, 21(1):104–115, 2005.
- [103] Ramin Homayouni, Kevin Heinrich, Lai Wei, and Michael W Berry. Gene clustering by latent semantic indexing of medline abstracts. *Bioinformatics*, 21(1):104–115, 2005.
- [104] D. Hristovski, B. Peterlin, J. A. Mitchell, and S. M. Humphrey. Using literature-based discovery to identify disease candidate genes. *Int J Med Inform*, 74(2-4):289–298, March 2005.
- [105] Haiyan Hu, Xifeng Yan, Yu Huang, Jiawei Han, and Xianghong Jasmine Zhou. Mining coherent dense subgraphs across massive biological networks for functional discovery. In *ISMB2005*, pages 213–221, 2005.
- [106] X. Hu. Mining and analyzing acal-free protein-protein interaction network. *Int. J. Bioinformatics Research and Applications*, 1(1):81–101, 2005.
- [107] Xiaohua Hu, Illhoi Yoo, Il-Yeol Song, Min Song, Jianchao Han, and Mark Lechner. Extracting and mining protein-protein interaction network from biomedical literature. In *Proceedings of the 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, pages 244–251, 2004.

- [108] Minlie Huang, Xiaoyan Zhu, Yu Hao, Donald G. Payan, Kunbin Qu, and Ming Li. Discovering patterns to extract protein-protein interactions from full texts. *Bioinformatics*, 20(18):3604–3612, 2004.
- [109] K. Humphreys, G. Demetriou, , and R. Gaizauskas. Two applications of information extraction to biological science journal articles: Enzyme interactions and protein structures. In *Proceedings of Pacific Symposium on Biocomputing (PSB-00)*, pages 502–513, Honolulu, Hawaii, USA, 2000.
- [110] Kevin Humphreys, George Demetriou, and Robert Galizauskas. Information extraction from biological science journal articles. In *IJCAI-WS1999B*, 1999.
- [111] Hsieh-Chuan Hung, Richard Tzong-Han Tsai, and Wen-Lian Hsu. Identifying protein interaction abstracts with contextual bag of words. In *AAAI 2007*, pages 1870–1871, 2007.
- [112] Eugene Ie, Jason Weston, William Stafford Noble, and Christina Leslie. Multi-class protein fold detection using adaptive codes. In *ICML2005*, 2005.
- [113] I. Iliopoulos, A. J. Enright, and C. A. Ouzounis. Textquest: Document clustering of Medline abstracts for concept discovery in molecular biology. In *Proceedings of Pacific Symposium on Biocomputing (PSB-01)*, pages 384–395, Honolulu, Hawaii, USA, 2001.
- [114] Takashi Ito, Kosuke Tashiro, Shigeru Muta, Ritsuko Ozawa, Tomoko Chiba, Mayumi Nishizawa, Kiyoshi Yamamoto, Satoru Kuhara, and Yoshiyuki Sakaki. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between yeast proteins. *PNAS*, 97(3):1143–1147, February 1, 2000.
- [115] Tommi Jaakkola, Mark Diekhans, and David Haussler. Using the Fisher kernel method to detect remote protein homologies. In *ISMB1999*, pages 149–158, 1999.
- [116] Hyunchul Jang, Jaesoo Lim, Joon-Ho Lim, Soo-Jun Park, Kyu-Chul Lee, and Seon-Hee Park. Finding the evidence for protein-protein interactions from pubmed abstracts. *Bioinformatics*, 22(14):e220–e226, July 2006.
- [117] Hyunchul Jang, Jaesoo Lim, Joon-Ho Lim, Soo-Jun Park, Kyu-Chul Lee, and Seon-Hee Park. Finding the evidence for protein-protein interactions from pubmed abstracts. *Bioinformatics*, 22(14):e220–e226, 2006.
- [118] Frizo Janssens, Wolfgang Glänzel, and Bart De Moor. Dynamic hybrid clustering of bioinformatics by incorporating text mining and citation analysis. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 360–369, New York, NY, USA, 2007. ACM Press.

- [119] Magesh Jayapandian, Adriane Chapman, V. Glenn Tarcea, Cong Yu, Aaron Elkiss, Angela Ianni, Bin Liu, Arnab Nandi, Carlos Santos, Philip Andrews, Brian Athey, David States, and H. V. Jagadish. Michigan Molecular Interactions (MiMI): putting the jigsaw puzzle together. *Nucleic Acids Research*, 35, November 2006.
- [120] Lars Juhl Jensen, Jasmin Saric, and Peer Bork. Literature mining for the biologist: From information retrieval to biological discovery. *Nature Reviews Genetics*, 7:119–129, February 2006.
- [121] H. Jeong, S.P. Mason, A.L. Barabási, and Z.N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, May 3, 2001.
- [122] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, October 2000.
- [123] Cliff A. Joslyn, Susan M. Mniszewski, Andy Fulmer, and Gary Heaton. The gene ontology categorizer. In *ISMB2004*, pages 169–177, 2004.
- [124] Raja Jothi, Maricel G. Kann, and Teresa M. Przytycka. Predicting protein-protein interaction by searching evolutionary tree automorphism space. In *ISMB2005*, pages 241–250, 2005.
- [125] V. Kandiah and D. L. Shepelyansky. Google matrix analysis of DNA sequences. *ArXiv e-prints*, January 2013.
- [126] Noam Kaplan and Michal Linial. False annotations of proteins: Automatic detection via keyword-based clustering. In *ISMB2004*, 2004.
- [127] Peter D. Karp. An ontology for biological function based on molecular interactions. *Bioinformatics*, 16(3):269–285, 2000.
- [128] S. Katrenko and et al. Learning biological interactions from Medline abstracts. In *ICML2005*, 2005.
- [129] Sophia Katrenko and Pieter Adriaans. Learning relations from biomedical corpora using dependency tree levels. In *Proceedings of the Fifteenth Dutch-Belgian Conference on Machine Learning*, May 12, 2006.
- [130] Sophia Katrenko and Pieter Adriaans. Learning relations from biomedical corpora using dependency trees. In *Knowledge Discovery and Emergent Complexity in Bioinformatics*, pages 61–80. Springer, 2007.
- [131] Mehmet Kayaalp, Alan R. Aronson, Susanne M. Humphrey, Nicholas C. Ide, Lorraine K. Tanabe, Lawrence H. Smith, Dina Demner-Fushman, Russell R. Loane, James G. Mork, and Olivier Bodenreider. Methods for accurate retrieval of MEDLINE citations in functional genomics. In *TREC 2003*, pages 441–450, 2003.

- [132] Halil Kilicoglu and Dina Demner-Fushman. Bio-scores: A smorgasbord architecture for coreference resolution in biomedical text. *PloS one*, 11(3):e0148538, 2016.
- [133] Jee-Hyub Kim, Alex Mitchell, Teresa K. Attwood, and Melanie Hilario. Learning to extract relations for protein annotation. *Bioinformatics*, 23:i256–i263, 2007.
- [134] Jin-Dong Kim, Tomoko Ohta, Yuka Tateisi, and Jun’ichi Tsujii. GENIA corpus - a semantically annotated corpus for bio-textmining. In *ISMB2003*, pages 180–182, 2003.
- [135] Jin-Dong Kim, Yuka Tateisi, Jun’ichi Tsujii, and Tomoko Ohta. Corpus-based approach to biological entity recognition. In *ISMB-WS2002A*, 2002.
- [136] Jung-Jae Kim, Zhuo Zhang, Jong C. Park, and See-Kiong Ng. Bio-contrasts: extracting and exploiting proteinprotein contrastive relations from biomedical literature. *Bioinformatics*, 22(5):597–605, March 2006. <http://www.ingentaconnect.com/content/oup/cabios/2006/00000022/00000005/art00597>.
- [137] A. Koike, Y. Kobayashi, and T. Takagi. Kinase pathway database: An integrated protein-kinase and NLP-based protein-interaction resource. *Genome Research*, 13:1231–1243, 2003.
- [138] Corinna Kolárik, Martin Hofmann-Apitius, Marc Zimmermann, and Juliane Fluck. Identification of new drug classification terms in textual resources. *Bioinformatics*, 23:i264–i272, 2007.
- [139] Jan A. Kors, Martijn J. Schuemie, Bob J. A. Schijvenaars, Marc Weeber, and Barend Mons. Combination of genetic databases improving identification of genes and proteins in text. In *ISMB-WS2005A*, 2005.
- [140] Zhenzhen Kou, William W. Cohen, and Robert F. Murphy. High-recall protein entity recognition using a dictionary. In *ISMB2005*, pages 266–273, 2005.
- [141] Mehmet Koyutürk, Ananth Grama, and Wojciech Szpankowski. An efficient algorithm for detecting frequent subgraphs in biological networks. In *ISMB2004*, pages 200–207, 2004.
- [142] M. Krallinger and A. Valencia. Text-mining and information-retrieval services for molecular biology. *Genome Biol.*, 6:224, 2005.
- [143] Michael Krauthammer, Pauline Kra, Ivan Iossifov, Shawn M. Gomez, George Hripcsak, Vasileios Hatzivassiloglou, Carol Friedman, and Andrey Rzhetsky. Of truth and pathways: Chasing bits of information through myriads of articles. In *ISMB2002*, pages 249–257, 2002.
- [144] Mark-A. Krogel and Tobias Scheffer. Multi-relational learning, text mining, and semi-supervised learning for functional genomics. *ML*, 57(1-2):61–81, October 2004.

- [145] Rui Kuang, Jason Weston, William Stafford Noble, and Christina Leslie. Motif-based protein ranking by network propagation. *Bioinformatics*, 21(19):3711–3718, 2005.
- [146] Seth Kulick, Mark Liberman, Martha Palmer, and Andrew Schein. Shallow semantic annotation of biomedical corpora for information extraction. In *ISMB-WS2003A*, 2003.
- [147] D. Kuonen. Challenges in bioinformatics for statistical data miners. *Bulletin of the Swiss Statistical Society*, 46:10–17, October 2003.
- [148] Matthew Lease and Eugene Charniak. Parsing biomedical literature. In R. Dale, K.-F. Wong, J. Su, and O. Y. Kwong, editors, *Proceedings of the Second International Joint Conference on Natural Language Processing*, pages 58–69, Jeju Island, Korea, October 11-13, 2005. Springer-Verlag.
- [149] Florian Leitner, Martin Krallinger, Carlos Rodriguez-Penagos, Jorg Hakenberg, Conrad Plake, Cheng-Ju Kuo, Chun-Nan Hsu, Richard Tsai, Hsi-Chuan Hung, William Lau, Calvin Johnson, Rune Saetre, Kazuhiro Yoshida, Yan Chen, Sun Kim, Soo-Yong Shin, Byoung-Tak Zhang, William Baumgartner, Lawrence Hunter, Barry Haddow, Michael Matthews, Xinglong Wang, Patrick Ruch, Frederic Ehrler, Arzucan Ozgur, Gunes Erkan, Dragomir Radev, Michael Krauthammer, ThaiBinh Luong, Robert Hoffmann, Chris Sander, and Alfonso Valencia. Introducing meta-services for biomedical information extraction. *Genome Biology*, 9(2), 2008.
- [150] GONDY Leroy and Hsinchun Chen. Genescene: An ontology-enhanced integration of linguistic and co-occurrence based relations in biomedical texts. *JASIST*, 56(5):457–468, March 2005.
- [151] C. Leslie and R. Kuang. Fast string kernels using inexact matching for protein sequences. *JMLR*, 5:1435–1455, November 2004.
- [152] Jiexun Li, Zhu Zhang, Xin Li, and Hsinchun Chen. Kernel-based learning for biomedical relation extraction. *J. Am. Soc. Inf. Sci. Technol.*, 59(5):756–769, 2008.
- [153] M. Light, X. Y. Qiu, and P. Srinivasan. The language of bioscience: Facts, speculations, and statements in between. *HLT Biolink*, pages 17–24, 2004.
- [154] J. Lin and W. J. Wilbur. Pubmed related articles: a probabilistic topic-based model for content similarity. *BMC Bioinformatics*, 8(1):423, October 2007.
- [155] Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, Chengxiang Zhai, and Bruce Schatz. Generating gene summaries from biomedical literature: A study of semi-structured summarization. *Inf. Process. Manage.*, 43(6):1777–1791, 2007.

- [156] Xu Ling, Jing Jiang, Xin He, Qiaozhu Mei, ChengXiang Zhai, and Bruce R. Schatz. Automatically generating gene summaries from biomedical literature. In *Pacific Symposium on Biocomputing*, pages 40–51, 2006.
- [157] Phillip W. Lord, Robert D. Stevens, Andy Brass, and Carole A. Goble. Semantic similarity measures as tools for exploring the gene ontology. In *Pacific Symposium on Biocomputing*, pages 601–612, 2003.
- [158] Zhiyong Lu, K. Bretonnel Cohen, and Lawrence Hunter. Finding GeneRIFs via gene ontology annotations. In *Proceedings of the Pacific Symposium on Biocomputing*, volume 11, pages 52–63, Maui, Hawaii, January 3-7, 2006. World Scientific Publishing Co. Pte. Ltd.
- [159] Inderjeet Mani, Zhangzhi Hu, Cathy Wu, Seok Bae Jang, Ken Samuel, Matthew Krause, and Jon Phillips. Protein name tagging guidelines: Lessons learned. In *ISMB-WS2004A*, 2004.
- [160] Fenglou Mao, Zhengchang Su, Victor Olman, Phuongan Dam, Zhijie Liu, and Ying Xu. Mapping of orthologous genes in the context of biological pathways: An application of integer programming. *PNAS*, 103(1):129–134, January 3, 2006.
- [161] E. M. Marcotte, I. Xenarios, and D. Eisenberg. Mining literature for protein-protein interactions. *Bioinformatics*, 17:359–363, 2001.
- [162] D. R. Masys, J. B. Welsh, J. Lynn Fink, M. Gribskov, I. Klacansky, and J. Corbeil. Use of keyword hierarchies to interpret gene expression patterns. *Bioinformatics*, 17(4):319–326, April 2001.
- [163] Jason McDermott, Roger Bumgarner, and Ram Samudrala. Functional annotation from predicted protein interaction networks. *Bioinformatics*, 21(15):3217–3226, August 2005.
- [164] Daniel M. McDonald, Hsinchun Chen, Hua Su, and Byron B. Marshall. Extracting gene pathway relations using a hybrid grammar: The Arizona relation parser. *Bioinformatics*, 20(18):3370–3378, 2004.
- [165] Ryan McDonald and Fernando Pereira. Identifying gene and protein mentions in text using conditional random fields. *BMC Bioinformatics*, 6(Suppl 1):S6, 2005.
- [166] Ryan McDonald and Fernando Pereira. Identifying gene and protein mentions in text using conditional random fields. *BMC bioinformatics*, 6(1):1, 2005.
- [167] Bridget McInnes, Ying Liu, Ted Pedersen, Genevieve Melton, and Serguei Pakhomov. Umls:: similarity: Measuring the relatedness and similarity of biomedical concepts. Association for Computational Linguistics, 2013.

- [168] Iain Melvin, Eugene Ie, Jason Weston, William Stafford Noble, and Christina Leslie. Multi-class protein classification using adaptive codes. *J. Mach. Learn. Res.*, 8:1557–1581, 2007.
- [169] Ramgopal R. Mettu, Ryan H. Lilien, and Bruce Randall Donald. High-throughput inference of protein-protein interfaces from unassigned NMR data. In *ISMB2005*, pages 292–301, 2005.
- [170] S. Micciché. Gene-based and semantic structure of the Gene Ontology as a complex network. *ArXiv e-prints*, November 2012.
- [171] Manuel Middendorf, Anshul Kundaje, Chris Wiggins, Yoav Freund, and Christina Leslie. Predicting genetic regulatory response using classification. In *ISMB2004*, pages 232–240, 2004.
- [172] Sven Mika and Burkhard Rost. Protein names precisely peeled off free text. In *ISMB2004*, pages 241–247, 2004.
- [173] David Milward, Marcus Bjärelund, William Hayes, Michelle Maxwell, Lisa Öberg, Nick Tilford, James Thomas, Roger Hale, Sylvia Knight, and Julie Barnes. Ontology-based interactive information extraction from scientific abstracts. In *ISMB-WS2004A*, 2004.
- [174] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, 6, 2016.
- [175] Tomohiro Mitsumori, Masaki Murata, Yasushi Fukuda, Kouichi Doi, and Hirohumi Doi. Extracting protein-protein interaction information from biomedical text with SVM. *IEICE Transactions on Information and Systems*, E89-D(8):2464–2466, August 2006.
- [176] Roser Morante, Anthony Liekens, and Walter Daelemans. Learning the scope of negation in biomedical texts. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 715–724, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.
- [177] Preslav Nakov, Ariel Schwartz, Brian Wolf, and Marti Hearst. Scaling up BioNLP: Application of a text annotation architecture to noun compound bracketing. In *ISMB-WS2005A*, 2005.
- [178] Preslav I. Nakov, Schwartz S. Ariel, and Hearst A. Marti. Citances: Citation sentences for semantic analysis of bioscience text. In *Workshop on Search and Discovery in Bioinformatics*, 2004.
- [179] M. Narayanaswamy, K. E. Ravikumar, and K. Vijay-Shanker. Beyond the clause: Extraction of Phosphorylation information from MedLine abstracts. In *ISMB2005*, pages 319–327, 2005.

- [180] Raheel Nawaz, Paul Thompson, and Sophia Ananiadou. Negated bio-events: analysis and identification. *BMC Bioinformatics*, 14(1):14, 2013.
- [181] William Stafford Noble, Rui Kuang, Christina Leslie, and Jason Weston. Identifying remote protein homologs by network propagation. *FEBS Journal*, 272(20):5119–5128, 2005.
- [182] Svetlana Novichkova, Sergei Egorov, and Nikolai Daraselia. MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics*, 19(13):1699–1706, 2003.
- [183] Martin A. Nowak and Karl Sigmund. Evolutionary dynamics of biological games. *Science*, 303:793–799, February 6, 2004.
- [184] R. Nusse. The Wnt gene homepage, 2004. <http://www.stanford.edu/~rnusse/pathways/cell2.html>.
- [185] Yoshihiro Ohta, Yasunori Yamamoto, Tomoko Okazaki, Ikuo Uchiyama, and Toshihisa Takagi. Automatic construction of knowledge base from biological papers. In *ISMB1997*, pages 218–225, 1997.
- [186] Z. N. Oltvai and A. L. Barbarasi. Network biology: Understanding the cell’s functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.
- [187] T. Ono and et al. Automated extraction of information on protein-protein interactions from the biological literature. *Bioinformatics*, 17:155–161, 2001.
- [188] Jahna Otterbacher. *Novelty and Change in Text*. PhD thesis, University of Michigan, December 2004. PhD Dissertation Proposal.
- [189] Gaurav Pandey, Michael Steinbach, Rohit Gupta, Tushar Garg, and Vipin Kumar. Association analysis-based transformations for protein interaction networks: a function prediction case study. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 540–549, New York, NY, USA, 2007. ACM Press.
- [190] Peter J. Park, Lu Tian, and Isaac S. Kohane. Linking gene expression data with patient survival times using partial least squares. In *ISMB2002*, pages 120–127, 2002.
- [191] C. Perez-Iratxeta, P. Bork, and M. A. Andrade. Association of genes to genetically inherited diseases using data mining. *Nat Genet*, 31(3):316–319, July 2002.
- [192] C. Perez-Iratxeta, Matthias Wjst, P. Bork, and M. A. Andrade. G2d: A tool for mining genes associated with disease. *BMC Genet*, 6(1), August 22, 2005.

- [193] Conrad Plake, Torsten Schiemann, Marcus Pankalla, Jorg Hakenberg, and Ulf Leser. ALIBABA: PubMed as a graph. *Bioinformatics*, page btl408, 2006.
- [194] Denys Proux, François Rechenmann, and Laurent Julliard. A pragmatic information extraction strategy for gathering data on genetic interactions. In *ISMB2000*, pages 279–285, 2000.
- [195] Arun Ramani, Edward Marcotte, Razvan Bunescu, and Raymond Mooney. Using biomedical literature mining to consolidate the set of known human protein-protein interactions. In *ISMB-WS2005A*, 2005.
- [196] Arun Ramani, Edward Marcotte, Razvan Bunescu, and Raymond Mooney. Using biomedical literature mining to consolidate the set of known human protein-protein interactions. In *Proceedings of the ACL-ISMB Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics*, pages 46–53. Association for Computational Linguistics, 2005.
- [197] Barbara A. Rapp and David L. Wheeler. Bioinformatics resources from the National Center for Biotechnology Information: An integrated foundation for discovery. *JASIST*, 56(5):538–550, March 2005.
- [198] Soumya Raychaudhuri, Hinrich Schütze, and Russ B. Altman. Text analysis of scientific literature can automatically determine if a group of genes share common biological function. In *ISMB-WS2002A*, 2002.
- [199] Soumya Raychaudhuri, Hinrich Schütze, and Russ B. Altman. Using text analysis to identify functionally coherent gene groups. *Genome Research*, 12(10):1582–1590, October 2002.
- [200] Soumya Raychaudhuri, Hinrich Schütze, and Russ B Altman. Using text analysis to identify functionally coherent gene groups. *Genome Research*, 12(10):1582–1590, 2002.
- [201] Soumya Raychaudhuri, Hinrich Schütze, and Russ B. Altman. Inclusion of textual documentation in the analysis of multidimensional data sets: Application to gene expression data. *ML*, 52(1-2):119–145, July-August 2003.
- [202] Lawrence H. Reeve, Hyoil Han, and Ari D. Brooks. The use of domain-specific concepts in biomedical text summarization. *Inf. Process. Manage.*, 43(6):1765–1776, 2007.
- [203] David J. Reiss and Benno Schwikowski. Predicting protein-peptide interactions via a network-based motif sampler. In *ISMB2004*, pages 274–282, 2004.

- [204] Lorenza Romano, Milen Kouylekov, Idan Szpektor, Ido Dagan, and Alberto Lavelli. Investigating a generic paraphrase-based approach for relation extraction. In *11th Conference of the European Chapter of the Association for Computational Linguistics (EACL '06)*, pages 409–416. European Chapter of the Association for Computational Linguistics, April 2006.
- [205] Barbara Rosario and Marti A. Hearst. Classifying semantic relations in bioscience texts. In *ACL2004*, pages 430–437, 2004.
- [206] Andrey Rzhetsky, Ivan Iossifov, Tomohiro Koike, Michael Krauthammer, Pauline Kra, Mitzi Morris, Hong Yu, Pablo Ariel Duboué, Wubin Weng, W. John Wilbur, Vasileios Hatzivassiloglou, and Carol Friedman. GeneWays: A system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J Biomed Inform.*, 37:43–53, 2004.
- [207] Andrey Rzhetsky, Ivan Iossifov, Ji Meng Loh, and Kevin P. White. Microparadigms: Chains of collective reasoning in publications about molecular interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 103(13):4940–4945, 2006.
- [208] Olivia Sanchez-Graillet and Massimo Poesio. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics*, 23:i424–i432, 2007.
- [209] Carlos Santos, Daniela Eggle, and David J. States. Wnt pathway curation using automated natural language processing: Combining statistical methods with partial and full parse for knowledge extraction. *Bioinformatics*, 21(8):1653–1658, 2005.
- [210] Jasmin Saric, Lars Juhl J. Jensen, Rossitza Ouzounova, Isabel Rojas, and Peer Bork. Extraction of regulatory gene/protein networks from medline. *Bioinformatics*, July 2005. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=
- [211] Alexander Schliep, Alexander Schonhuth, and Christine Steinhoff. Using hidden Markov models to analyze gene expression time course data. In *ISMB2003*, pages 255–263, 2003.
- [212] M. J. Schuemie, M. Weeber, B. J. A. Schijvenaars, E. M. van Mulligen, C. C. van der Eijk, R. Jelier, B. Mons, and J. A. Kors. Distribution of information in biomedical abstracts and full-text publications. *Bioinformatics*, 20(16):2597–2604, 2004.
- [213] B. Schwikowski, P. Uetz, and S. Fields. A network of protein-protein interactions in yeast. *Nat Biotechnol*, 18(12):1257–1261, December 2000.
- [214] David B. Searls. The language of genes. *Nature*, 420:211–217, November 2002.
- [215] David B Searls. The language of genes. *Nature*, 420(6912):211–217, 2002.

- [216] Eran Segal, Benjamin Taskar, Audrey Gasch, Nir Friedman, and Daphne Koller. Rich probabilistic models for gene expression. In *ISMB2001*, pages 243–252, 2001.
- [217] Eran Segal, Haidong Wang, and Daphe Koller. Discovering molecular pathways from protein interaction and gene expression data. In *ISMB2003*, pages 264–272, 2003.
- [218] Aditya Kumar Sehgal and Padmini Srinivasan. Manjal - a text mining system for MEDLINE. In *SIGIR2005*, pages 680–680, 2005.
- [219] Kazuhiro Seki and Javed Mostafa. A probabilistic model for identifying protein names and their name boundaries. In *In Proceedings of the 2nd IEEE-CS Bioinformatics Conference (CSB 2003)*, pages 251–258, August 2003.
- [220] Kazuhiro Seki and Javed Mostafa. An application of text categorization methods to gene ontology annotation. In *SIGIR2005*, 2005.
- [221] Kazuhiro Seki and Javed Mostafa. A hybrid approach to protein name identification in biomedical texts. *IPM*, 41(4):723–743, July 2005.
- [222] Anil K. Seth and Gerald M. Edelman. Theoretical neuroanatomy: Analyzing the structure, dynamics, and function of neuronal networks. In Eli Ben-Naim, Hans Frauenfelder, and Zoltan Toroczkai, editors, *Complex Networks*, pages 483–511. Springer Berlin/Heidelberg, 2004. Lecture Notes in Physics - Volume 650.
- [223] B. Settles. ABNER: an open source tool for automatically tagging genes, proteins, and other entity names in text. *Bioinformatics*, 21(14):3191–3192, 2005.
- [224] P. K. Shah, C. Perez-Iratxeta, P. Bork, and M. A. Andrade. Information extraction from full text scientific articles: where are the keywords? *BMC Bioinformatics*, 4(1), May 29, 2003.
- [225] Hagit Shatkay and Ronen Feldman. Mining the biomedical literature in the genomic era: An overview. *Journal of Computational Biology*, 10(6):821–855, December 2003.
- [226] Hagit Shatkay and Ronen Feldman. Mining the biomedical literature in the genomic era: an overview. *Journal of computational biology*, 10(6):821–855, 2003.
- [227] Hagit Shatkay, A. Rzhetsky, and W. J. Wilbur. Searching for high-utility text in the biomedical literature. In *ISMB-WS2005A*, 2005.
- [228] L. H. Smith, L. Tanabe, T. Rindfleisch, and W. John Wilbur. MedTag: A collection of biomedical annotations. In *ISMB-WS2005A*, 2005.

- [229] Tony C. Smith and John G. Cleary. Automatically linking MEDLINE abstracts to the gene ontology. In *ISMB-WS2003A*, 2003.
- [230] Ricard V. Solé, Romualdo Pastor-Satorras, Eric Smith, and Thomas B. Kepler. A model of large-scale proteome evolution. *Advances in Complex Systems*, 5(1):43–54, 2002.
- [231] V. Spirin and L. A. Mirny. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci U S A*, 100(21):12123–12128, October 2003.
- [232] Padmini Srinivasan. Text mining: Generating hypotheses from MEDLINE. *JASIST*, 55(5):396–413, March 2004.
- [233] Padmini Srinivasan and Bisharah Libbus. Mining MEDLINE for implicit links between dietary substances and diseases. In *ISMB2004*, pages 290–296, 2004.
- [234] Padmini Srinivasan and Bisharah Libbus. Mining medline for implicit links between dietary substances and diseases. *Bioinformatics*, 20(suppl 1):i290–i296, 2004.
- [235] M. Stephens, M. Palakal, S. Mukhopadhyay, R. Raje, and J. Mostafa. Detecting gene relations from Medline abstracts. In *Proceedings of Pacific Symposium on Biocomputing (PSB-01)*, pages 483–495, Honolulu, Hawaii, USA, 2001.
- [236] Venkata L. Subramaniam, Sougata Mukherjea, Pankaj Kankar, Biplav Srivastava, Vishal S. Batra, Pasumarti V. Kamesam, and Ravi Kothari. Information extraction from biomedical literature: methodology, evaluation and an application. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 410–417. ACM Press, 2003.
- [237] K. Sugiyama and et al. Extracting information on protein-protein interactions from biological literature based on machine learning approaches. *Genome Informatics*, 14:699–700, 2003.
- [238] B. P. Suomela and M. A. Andrade. Ranking the whole MEDLINE database according to a large training set using text indexing. *BMC Bioinformatics*, 6(75), March 24, 2005.
- [239] Don R. Swanson and Neil R. Smalheiser. Link analysis of MEDLINE titles as an aid to scientific discovery. In *AAAI-FS1998A*, pages 94–97, 1998.
- [240] Rishi Talreja, Andrew Schein, Scott Winters, and Lyle Ungar. GeneTaggerCRF: An entity tagger for recognizing gene names in text. Technical report, PennBioIE - The Biomedical Information Extraction Project at the University of Pennsylvania, May 2004.

- [241] L. Tanabe, N. Xie, L. H. Thom, W. Matten, and W. J. Wilbur. Genetag: a tagged corpus for gene/protein named entity recognition. *BMC Bioinformatics*, 6 Suppl 1, August 25, 2005.
- [242] Lorraine Tanabe and W. John Wilbur. Tagging gene and protein names in biomedical text. *Bioinformatics*, 18(8):1124–1132, 2002.
- [243] Ben Taskar. A framework for learning with combinatorial output structure and applications to protein structure prediction. In *NIPS-WS2004B*, 2004.
- [244] Yuka Tateisi, Tomoko Ohta, Jin-Dong Kim, Huaqing Hong, Su Jian, and Jun-Ichi Tsujii. The GENIA corpus: MEDLINE abstracts annotated with linguistic information. In *ISMB-WS2003A*, 2003.
- [245] J. M. Temkin and M. R. Gilder. Extraction of protein interaction information from unstructured text using a context-free grammar. *Bioinformatics*, 19:2046–2053, 2003.
- [246] Yuanyuan Tian, Richard C. McEachin, Carlos Santos, David J. States, and Jignesh M. Patel. SAGA: A subgraph matching tool for biological graphs. *Bioinformatics*, 23(2):232–239, January 15, 2007.
- [247] Yuanyuan Tian, Richard C. Mceachin, Carlos Santos, David J. States, and Jignesh M. Patel. Saga: a subgraph matching tool for biological graphs. *Bioinformatics*, 23(2):232–239, 2007.
- [248] Vetle I. Torvik and Neil R. Smalheiser. A quantitative model for linking two disparate sets of articles in medline. *Bioinformatics*, pages btm161+, April 2007.
- [249] Vetle I. Torvik, Marc Weeber, Don R. Swanson, and Neil R. Smalheiser. A probabilistic similarity metric for MEDLINE records: A model for author name disambiguation. *JASIST*, 56(2):140–158, January 15, 2005.
- [250] Ono Toshihide, Hishigaki Haretsugu, Tanigami Akira, and Takagi Toshihisa. Automatic extraction of information on protein-protein interaction from scientific literature. *Genome Informatics*, 10:296–297, 1999.
- [251] Koji Tsuda and William Noble. Learning kernels from biological networks by maximizing entropy. In *ISMB2004*, pages 326–333, 2004.
- [252] Aditya Vailaya, Peter Bluvas, Robert Kincaid, Allan Kuchinsky, Michael Creech, , and Annette Adler. An architecture for biological information extraction and representation. *Bioinformatics*, December 2004.
- [253] Alexei Vázquez, Alessandro Flammini, Amos Maritan, and Alessandro Vespignani. Modeling of protein interaction networks. *ComplexUs*, 1:38–44, November 2003.
- [254] Werner P Veldsman and Alan Christoffels. Sniphunter: A snp-based search engine. *Data*, 1(3):17, 2016.

- [255] Stella Veretnik, J Lynn Fink, and Philip E Bourne. Computational biology resources lack persistence and usability. *PLoS Comput Biol*, 4(7):e1000136, 2008.
- [256] Karin Verspoor. Towards a semantic lexicon for biological language processing. In *ISMB-WS2004A*, 2004.
- [257] K.M. Verspoor. Drawing on millions of biomedical journal publications to do predictive biology. In *Big Data and Smart Computing (BigComp), 2015 International Conference on*, pages 251–253, Feb 2015.
- [258] Jean-Philippe Vert. A tree kernel to analyse phylogenetic profiles. In *ISMB2002*, pages 276–284, 2002.
- [259] Jean-Philippe Vert. Analysis and inference of gene networks from genomic data. In *ISMB-WS2004B*, 2004.
- [260] Jean-Philippe Vert. Kernels for protein sequences. In *NIPS-WS2005F*, 2005.
- [261] Haidong Wang, Eran Segal, Asa Ben-Hur, Daphne Koller, and Douglas L. Brutlag. Identifying protein-protein interaction sites on a genome-wide scale. In *NIPS2004*, pages 1465–1472, 2004.
- [262] Hudong Wang, Shannon Bradshaw, and Marc Light. Automatic highlighting of bioscience literature. In *ISMB-WS2005A*, 2005.
- [263] Ben Wellner. Weakly supervised learning methods for improving the quality of gene name normalization data. In *ISMB-WS2005A*, 2005.
- [264] Ben Wellner, José Castaño, and James Pustejovsky. Adaptive string similarity metrics for biomedical reference resolution. In *ISMB-WS2005A*, 2005.
- [265] Jason Weston, Andre Elisseeff, Dengyong Zhou, Christina Leslie, and William S. Noble. Protein ranking: From local to global structure in the protein similarity network. *PNAS*, 101(17):6559–6563, 2004.
- [266] Daniela Wieser, Ernst Kretschmann, and Rolf Apweiler. Filtering erroneous protein annotation. In *ISMB2004*, pages 342–347, 2004.
- [267] W. J. Wilbur and Y. Yang. An analysis of statistical term strength and its use in the indexing and retrieval of molecular biology texts. *Computers in Biology and Medicine*, 26(3):209–222, 1996.
- [268] W. John Wilbur, Andrey Rzhetsky, and Hagit Shatkay. New directions in biomedical text annotation: Definitions, guidelines and corpus construction. *BMC Bioinformatics*, 7(356), July 25, 2006.

- [269] Adam Wilcox, George Hripcsak, and Carol Friedman. Using knowledge sources to improve classification of medical text reports. In *KDD-WS2000A*, 2000.
- [270] Hugh E. Williams. Introduction to genomic information retrieval. *ACM-TOIS*, 23(1):1–2, January 2005.
- [271] Jérôme Wojcik and Vincent Schächter. Protein-protein interaction map inference using interacting domain profile pairs. In *ISMB2001*, pages 296–305, 2001.
- [272] S. Wuchty, Z. N. Oltvai, and A. L. Barabasi. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat Genet*, 35(2):176–179, October 2003.
- [273] Yoshihiro Yamanishi, Jean-Philippe Vert, and Minoru Kanehisa. Protein network inference from multiple genomic data: A supervised approach. In *ISMB2004*, pages 363–370, 2004.
- [274] Xin Yan, Xue Li, and Dawei Song. Document re-ranking by generality in bio-medical information retrieval. In *WISE2005*, pages 376–389, 2005.
- [275] M. D. Yandell and W. H. Majoros. Genomics and natural language processing. *Nature Rev. Genet.*, 3:601–610, August 2002.
- [276] Alex Yeh, Lynette Hirschman, and Alex Morgan. BioCreAtIvE: Entity extraction. In *ISMB-WS2003A*, 2003.
- [277] Alexander S. Yeh, Lynette Hirschman, and Alexander A. Morgan. Evaluation of text data mining for database curation: Lessons learned from the KDD Challenge Cup. In *ISMB2003*, pages 331–339, 2003.
- [278] Eunji Yi, Gary G. Lee, and Soo-Jun Park. HMM-based protein name recognition with edit distance using automatically annotated corpus. In *ISMB-WS2003A*, 2003.
- [279] Golan Yona, Nathan Linial, Naftali Tishby, and Michal Linial. A map of the protein space: An automatic hierarchical classification of all protein sequences. In *ISMB1998*, pages 212–221, 1998.
- [280] Hong Yu and Eugene Agichtein. Extracting synonymous gene and protein terms from biological literature. In *ISMB2003*, pages 340–349, 2003.
- [281] Hong Yu, Won Kim, Vasileios Hatzivassiloglou, and John Wilbur. A large scale, corpus-based approach for automatically disambiguating biomedical abbreviations. *ACM-TOIS*, 24(3):380–404, July 2006.
- [282] Mohammed J. Zaki, Shinichi Morishita, and Isidore Rigoutsos. Report on BIODDD04: Workshop on data mining in Bioinformatics. *SIGKDD-EN*, 6(2):153–154, 2004.

- [283] Andreas Zanzoni, Luisa Montecchi-Palazzi, Michele Quondam, Gabriele Ausiello, Manuela Helmer-Citterich, and Gianni Cesareni. MINT: A Molecular INTERaction database. *FEBS Letters*, 513(1):135–140, February 20, 2002.
- [284] GuoDong Zhou, Jie Zhang, Jian Su, Dan Shen, and ChewLim Tan. Recognizing names in biomedical texts: A machine learning approach. *Bioinformatics*, 20(7):1178–1190, 2004.