Selected Bibliography for Text Mining for Biomedicine

Sophia Ananiadou
National Centre for Text Mining www.nactem.ac.uk
School of Computer Science, University of Manchester

(i) General overviews and position papers on Text Mining for Biomedicine

[2007] Pierre Zweigenbaum, Dina Demner-Fushman, Hong Yu, Kevin B. Cohen: Frontiers of biomedical text mining: current progress. Briefings in Bioinformatics 8(5): 358-375 (2007)

[2006] Ananiadou, S. & J. McNaught (eds) (2006) Text Mining for Biology and Biomedicine, Artech House.

[2006] Ananiadou, S., Kell, D.B. and Tsujii, J. (2006) Text Mining and its Potential Applications in Systems Biology, in Trends in Biotechnology (TIBTECH) 24(12):571-579

[2006] Lars Juhl Jensen, J. Saric and P. Bork (2006) "Literature mining for the biologist: from information retrieval to biological discovery", Nature Reviews Genetics, Vol. 7, Feb. 2006, pp 119-129

[2005] Cohen, A. M., and W. R. Hersh, "A Survey of Current Work in Biomedical Text Mining," Briefings in Bioinformatics, Vol. 6, 2005, pp. 57--71.

[2005] MacMullen, W.J, and S.O. Denn, "Information Problems in Molecular Biology and Bioinformatics," Journal of the American Society for Information Science and Technology, Vol. 56, No. 5, 2005, pp. 447--456.

[2005] Rebholz-Schuhmann, D., H. Kirsch, and F. Couto, "Facts from Text—Is Text Mining Ready to Deliver? " PLoS Biology, Vol. 3, No. 2, 2005, pp. 0188--0191, http://www.plosbiology.org, June 2005.

[2005] Shatkay H. (2005) Hairpins in bookstacks: information retrieval from biomedical text. Brief Bioinform 2005;6(3):222–38.

[2005] Tsujii, Jun-ichi and Sophia Ananiadou (2005) Thesaurus or logical onotology, which do we need for mining text? Language Resources and Evaluation . 39(1). pp. 77-90, Springer SBM, September 2005.

[2004] Cohen KB and Hunter L. (2004) Natural language processing and systems biology. In: Dubitzky W, Azuaje F, (eds). Artificial Intelligence Methods and Tools for Systems Biology. Heidelberg: Springer, 2004;147–74.

[2004] Nédellec, C., "Machine Learning for Information Extraction in Genomics---State of the Art and Perspectives." Text Mining and its Applications, pp. 99--118, S. Sirmakessis (ed.), Berlin: Springer-Verlag, Studies in Fuzziness and Soft Computing 138, 2004.

[2004] Shatkay, H., and R. Feldman, "Mining the Biomedical Literature in the Genomic Era: An Overview," Journal of Computational Biology, Vol. 10, No. 6, 2004, pp. 821--855.

[2002] Blaschke, C., L. Hirschman, and A. Valencia, "Information Extraction in Molecular Biology," Briefings in Bioinformatics, Vol. 3, No. 2, 2002, pp. 1--12.

[2002] Hirschman, L., et al., "Accomplishments and Challenges in Literature Data Mining for Biology," Bioinformatics, Vol. 18, No. 12, 2002, pp. 1553--1561.

[2002] Mack, R., and M. Hehenburger, "Text-based Knowledge Discovery: Search and Mining of Life-sciences Documents," Drug Discovery Today, Vol. 7, No. 11, 2002, pp. S89--S98.

[2002] Yandell, M. D., and W. H. Majoros, "Genomics and Natural Language Processing," Nature Reviews/Genetics, Vol. 3, 2002, pp. 601--610.

(ii) Forums for biomedical text mining on the Web

BLIMP http://blimp.cs.queensu.ca/

"BLIMP covers all publications related to the fast-growing field of biomedical literature and text mining. It is a one-stop resource, letting researchers find out who-does-what in the area and where it is published, bridging across the many discipline-specific venues in which biomedical text-mining papers are published."

BIONLP http://www.ccs.neu.edu/home/futrelle/bionlp/

Bob Futrelle's NLP for Biotext Mining

http://www.text-mining.org/

A comprehensive collection of text mining resources, including links to publications, commercial suppliers, news items, research groups, events, etc

(iii) Named Entity Recognition and Terminology Management

[2008] Gaudan,G., et al. (2008) Identifying GO terms in text. EURASIP JBSB, Hindawi Publishing Group. (accepted)

[2008] Yoshimasa Tsuruoka, John McNaught, and Sophia Ananiadou (2008) Normalizing biomedical terms by minimizing ambiguity and variability.BMC Bioinformatics (in press).

[2007] Yoshida, Kazuhiro and Jun'ichi Tsujii. Reranking for Biomedical Named-Entity Recognition. BioNLP 2007. Prague, Czech Republic.

[2007] Yoshimasa Tsuruoka, John McNaught, Jun'ichi Tsujii, and Sophia Ananiadou (2007) Learning string similarity measures for gene/protein name dictionary look-up using logistic regression. *Bioinformatics,* Vol. 23, No. 20, pp. 2768-2774.

[2006] Ananiadou, S. & Nenadic, G. (2006) Automatic Terminology Management in Biomedicine, in Text Mining for Biology and Biomedicine, Artech house, pp.67-98

[2006] Chang, J.T. and Schütze, H. (2006) Abbreviations in biomedical text. In Ananiadou, S. and McNaught, J. (Eds.). Text Mining for Biology and Biomedicine, London Artech House Inc, pp. 99–119.

[2006] Daisuke Okanohara, Yusuke Miyao, Yoshimasa Tsuruoka, and Jun'ichi Tsujii (2006) Improving the Scalability of Semi-Markov Conditional Random Fields for Named Entity Recognition. COLING/ACL 2006, pp. 465-472.

[2006] Liu H., Hu Z., Zhang J. and Wu C., (2006). BioThesaurus: a web-based thesaurus of protein and gene names. Bioinformatics 22(1):103-105.

[2006] Okazaki, N. and Ananiadou, S. (2006) Building an Abbreviation Dictionary using a Term Recognition Approach, in Bioinformatics, 22(24):3089-3095

[2006] Okazaki, N. and Ananiadou, S. (2006) Clustering acronyms in biomedical text for disambiguation, in Proceedings of LREC.

[2006] Park, J.C. & Kim, J.J. (2006) Named Entity Recognition, in Text Mining for Biology and Biomedicine, Artech house, pp. 121- 142

[2006] Rebholz-Schuhmann,D., et al. (2006a) Annotation and Disambiguation of Semantic Types in Biomedical Text: a Cascaded Approach to Named Entity Recognition. Workshop on "Multi-Dimensional Markup in NLP", EACL 2006, Trente, Italy.

[2005] Hakenberg,J., et al. (2005). Systematic feature evaluation for gene name recognition. BMC Bioinformatics; 6 Suppl 1:S9.

[2005] Hanisch D, Fundel K, Mevissen HT, et al. Prominer: rule based protein and gene entity recognition. BMC Bioinformatics 2005;6(Suppl 1):(S14).

[2005] Hirschman,L., et al. (2005) Overview of BioCreAtIvE task 1B: normalized gene lists. BMC Bioinformatics, 6(Suppl 1):S11.

[2005] Kazama, Jun'ichi and Jun'ichi Tsujii (2005) Maximum Entropy Models with Inequality Constraints: A case study on text categorization. Machine Learning 60(1-3). pp. 169-194, Springer SBM, September 2005.

[2005] Koike, A., Y. Niwa, and T. Takagi, "Automatic Extraction of Gene/Protein Biological Functions from Biomedical Text", Bioinformatics, Vol. 21, No.7, 2005, pp.1227--1236.

[2005] McDonald R, Pereira F. Identifying gene and protein mentions in text using conditional random fields. BMC Bioinform 2005;6:(Suppl)(1:S6).

[2005] Spasic, I., and S. Ananiadou, "A Flexible Measure of Contextual Similarity for Biomedical Terms" Proceedings of Pacific Symposium on Biocomputing (PSB, 2005), Hawaii, USA, 2005.

[2005] Spasic, I., S. Ananiadou, and J. Tsujii, "MaSTerClass: a Case-based Reasoning System for the Classification of Biomedical Terms", Bioinformatics, Vol.21, No.11, 2005, pp.2748--2758.

[2005] Tsuruoka, Y., S. Ananiadou, and J. Tsujii, "A Machine Learning Approach to Automatic Acronym Generation", Proc. of Bio-LINK, ISMB, 2005.

[2005] Wren, J.D., et al. (2005) Biomedical term mapping databases. Nucleic Acids Res, 33.

[2004] Adar, E., "Sarad: a Simple and Robust Abbreviation Dictionary," Bioinformatics, Vol. 20, No. 4, 2004, pp. 527--533.

[2004] Ananiadou, S., Friedman, C. & Tsujii, J (eds) (2004) Named Entity Recognition in Biomedicine, Special Issue, Journal of Biomedical Informatics, vol. 37 (6).

[2004] Harkema, H., et al., "A Large-Scale Terminology Resource for Biomedical Text Processing", Proceedings of BioLINK, 2004, pp.53--60.

[2004] Kim, J., et al., "Introduction to the Bio-Entity Recognition Task at JNLPBA," Proc. Int. Workshop on Natural Language Processing in Biomedicine and its Applications, 2004, pp. 70--75.

[2004] Krauthammer, M., and G. Nenadic, "Term Identification in the Biomedical Literature", Journal of Biomedical Informatics, Special Issue on Named Entity Recognition in Biomedicine, Vol.37, No. 6, 2004, pp.512--526.

[2004] Lee, K.J., et al., "Biomedical Named Entity Recognition using Two-Phase Model based on SVMs", Journal of Biomedical Informatics, Special Issue, Named Entity Recognition in Biomedicine, Vol.37, No. 6, 2004, pp. 436--447.

[2004] Mika, S., and B. Rost, "Protein Names Precisely Peeled off Free Text," Bioinformatics, Vol. 20, Suppl. 1, 2004, pp. i241--i247.

[2004] Morgan, A., et al., "Gene Name Identification and Normalization using a Model Organism Database," Journal of Biomedical Informatics, Vol. 37, 2004, pp. 396--410.

[2004] Nenadic, G., S. Ananiadou, and J. McNaught, "Enhancing Automatic Term Recognition through Recognition of Variation", Proceedings of COLING 2004, Geneva, Switzerland, 2004, pp. 604--610.

[2004] Ogren, P., et al.,"The Compositional Structure of Gene Ontology Terms", Proc. PSB, 2004, pp.214--225.

[2004] Tsuruoka,Y. and J.Tsujii, "Improving the Performance of Dictionary-based Approaches in Protein Name Recognition" Journal of Biomedical Informatics, Special Issue on Named Entity Recognition in Biomedicine, Vol.37, No. 6, 2004, pp. 461--470.

[2004] Tuason, O., L. Chen, H. Liu, J.A. Blake, and C. Friedman, "Biological Nomenclature: A Source of Lexical Knowledge and Ambiguity", Proceedings of Pac Symp Biocomputing, Hawaii, 2004, pp. 238--249.

[2004] Zhou, G., et al., "Recognizing Names in Biomedical Texts: A Machine Learning Approach," Bioinformatics, Vol. 20, No. 4, 2004, pp. 1178--1190.

[2003] Cantor M.N., et al., "An Evaluation of Hybrid Methods for Matching Biomedical Terminologies: Mapping the Gene Ontology to the UMLS", Stud. Health Technol. Inform. (95), 2003, pp.62—67

[2003] Hanisch, D., et al., "Playing Biology's Name Game: Identifying Protein Names in Scientific Text," Proc. Pacific Symp. on Biocomputing, 2003, pp. 403--414.

[2003] Morgan, A., et al., "Gene Name Extraction Using FlyBase Resources", Proceedings of ACL Workshop, NLP in Biomedicine, Sapporo, Japan, 2003, pp.1--8.

[2003] Narayanaswamy, M., K.E. Ravikumar, and K. Vijay-Shanker, " A Biological Named Entity Recognizer", PSB, 2003, pp.427--438.

[2003] Nenadic, G., I. Spasic, and S. Ananiadou, "Terminology-Driven Mining of Biomedical Literature", Bioinformatics, Vol. 19, No.8, 2003, pp.938--943.

[2003] Schwartz, A.S. and M.A. Hearst, "A Simple Algorithm for Identifying Abbreviation Definitions in Biomedical Text," Pacific Symposium on Biocomputing, 2003, pp. 451--462.

[2003] Takeuchi, K. and N. Collier, "Bio-medical Entity Extraction using Support Vector Machines," Proceedings of ACL Workshop NLP in Biomedicine, Sapporo, Japan, 2003, pp. 57--64.

[2003] Torii, M., S. Kamboj, and K. Vijay-Shanker, "An Investigation of Various Information Sources for Classifying Biological Names," Proceedings of ACL Workshop NLP in Biomedicine, Sapporo, Japan, 2003, pp.113--120.

[2003] Tsuruoka, Y., and J. Tsujii, "Probabilistic Term Variant Generator for Biomedical Terms," Proceedings of 26th Annual ACM SIGIR Conference, 2003, pp.167--173.

[2003] Yamamoto, K., et al. "Protein Name Tagging for Biomedical Annotation in Text", Proc. of ACL Workshop NLP in Biomedicine, Sapporo, Japan, 2003, pp.65--72.

[2003] Yu, H., and E. Agichtein, "Extracting Synonymous Gene and Protein Terms from Biological Literature," Bioinformatics, 19, Suppl 1, 2003, pp.I340--349.

[2003] Yu, Z., Y. Tsuruoka, and J. Tsujii, "Automatic Resolution of Ambiguous Abbreviations in Biomedical Texts using Support Vector Machines and One Sense per Discourse Hypothesis," SIGIR 2003 Workshop on Text Analysis and Search for Bioinformatics, 2003, pp. 57--62.

[2002] Bodenreider, O., J.A. Mitchell, and A.T. McCray, "Evaluation of the UMLS as a Terminology and Knowledge Resource for Biomedical Informatics", Proc. AMIA, 2002, pp.61--65.

[2002] Cohen, K.B., G.K. Acquaah-Mensah, A.E. Dolbey, and L. Hunter, "Contrast and Variability in Gene Names," Proceedings of ACL Workshop on NLP in the Biomedical Domain, Philadelphia, USA, 2002, pp.14--20.

[2002] Franzen, K., et al., "Protein Names and How to Find Them," International Journal of Medical Informatics, Vol. 67, No. 1--3, 2002, pp. 49--61.

[2002] Hirschman, L., A. Morgan, and A.S. Yeh, "Rutabaga by Any Other Name: Extracting Biological Names". Journal of Biomedical Informatics, Vol.35, No. 4, 2002, pp.247--259.

[2002] Liu, H., S.B. Johnson, and C. Friedman, "Automatic Resolution of Ambiguous Terms based on Machine Learning and Conceptual Relations in the UMLS," Journal of the American Medical Informatics Association, Vol. 9, No. 4, 2002, pp. 621--636.

[2002] Nenadic, G., H. Mima, I. Spasic, I., S. Ananiadou, and J. Tsujii, "Terminology-based Literature Mining and Knowledge Acquisition in Biomedicine", International Journal of Medical Informatics, Vol. 67, No.(1-3), 2002, pp.33--48.

[2002] Pakhomov, S., "Semi-Supervised Maximum Entropy Based Approach to Acronym and Abbreviation Normalization in Medical Texts", Proceedings of 40th ACL Conference, 2002, pp.160--167.

[2002] Pustejovsky, J., et al., "Medstract: Creating Large-Scale Information Servers for Biomedical Libraries," ACL Workshop on Natural Language Processing in the Biomedical Domain, 2002, pp. 85—92.

[2002] Raychaudhuri, S., J.T. Chang, P.D. Sutphin, and R.B. Altman, "Associating Genes with Gene Ontology Codes Using a Maximum Entropy Analysis of Biomedical Literature," Genome Res, Vol.12, No.1, 2002, pp.203--214.

[2002] Tanabe, L., and W.J. Wilbur, "Tagging Gene and Protein Names in Biomedical Text", Bioinformatics, 18(8), 2002, pp.1124--1132.

[2002] Thelen, M., and E. Riloff, "A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts", Proceedings of EMNLP, 2002.

[2001] Aronson, A.R, "Effective Mapping of Biomedical Text to the UMLS Metathesaurus: the MetaMap Program," Proceedings of AMIA, 2001, pp.17--21.

[2001] Hatzivassiloglou, V., P.A. Duboue, and A. Rzhetsky, "Disambiguating Proteins, Genes, and RNA in Text: A Machine Language Approach" Bioinformatics, 17, Suppl 1, 2001, pp.97--106.

[2001] Krauthammer, M., A. Rzhetsky, P. Morozov, and C. Friedman, "Using BLAST for identifying gene and protein names in journal articles" Gene, Vol. 259, No. (1–2), 2001, pp.245--252.

[2001] Liu, H., Y.A. Lussier, and C. Friedman, "A Study of Abbreviations in the UMLS," Proceedings of the American Medical Informatics Association Annual Symposium, 2001, pp. 393--397.

[2001] Mima, H., S. Ananiadou, and G. Nenadic, "The ATRACT Workbench: Automatic Term Recognition and Clustering for Terms". In Text, Speech and Dialogue Matousek, V., Mautner, P., Moucek, R. and Tauser, K. (eds.), Lecture Notes in Artificial Intelligence 2166, Springer Verlag, Heidelberg, pp.126--133, 2001.

[2001] Pustejovsky, J., et al., "Automatic Extraction of Acronym--Meaning Pairs from Medline Databases," Medinfo, Vol. 10, 2001, pp. 371--375.

[2000] Yoshida, M., K. Fukuda, and T. Takagi, "Pnad-css: a Workbench for Constructing a Protein Name Abbreviation Dictionary," Bioinformatics, Vol. 16, 2000, pp. 169--75.

[1999] Nobata, C., N. Collier, and J. Tsujii, "Automatic Term Identification and Classification in Biological Texts," Proceedings of Natural Language Pacific Rim Symposium, 1999, pp.369--374.

[1998] Fukuda, K., et al., "Towards Information Extraction: Identifying Protein Names from Biological Papers", PSB, Hawaii, USA, 1998, pp.707--718.

[1998] Proux, D., et al., "Detecting Gene Symbols and Names in Biomedical Texts: A First Step toward Pertinent Information," Proc. 9th Workshop on Genome Informatics, 1998, pp. 72--80.

(iv) Relation mining, event extraction

[2008] Kim, Jin-Dong, Tomoko Ohta and Jun'ichi Tsujii. Corpus annotation for mining biomedical events from lterature. BMC Bioinformatics. 9(1). pp. 10, BioMed Central, 2008.

[2008] Oda, Kanae, Jin-Dong Kim, Tomoko Ohta, Daisuke Okanohara, Takuya Matsuzaki, Yuka Tateisi and Jun'ichi Tsujii. (2008) New challenges for text mining: Mapping between text and manually curated pathways. In Christopher JO Baker and Su Jian (Eds.), BMC Bioinformatics.2008. To appear.

[2007] Ahlers CB, Fiszman M, Demner-Fushman D, et al. Extracting semantic predications from MEDLINE citations for pharmacogenomics. In: Pac Symp Biocomput 12. Maui, Hawaii, 2007:209–20.

[2007] Krallinger,M., Leitner,F., and Valencia,A. (2007) Assessment of the Second BioCreative PPI task: Automatic Extraction of Protein-Protein Interactions. Proc Second BioCreative Challenge Evaluation Workshop.

[2007] Rebholz-Schuhmann D, Kirsch H, Arregui M, et al. EBIMed-text crunching to gather facts for proteins from Medline. Bioinformatics 2007;23:e237–44

[2007] Rebholz-Schuhmann,D., et al. (2007b) Text processing through Web services: Calling Whatizit. Bioinformatics 2007 Nov 21

[2007] Rinaldi F, Schneider G, Kaljurand K, et al. Mining of relations between proteins over biomedical scientific literature using a deep-linguistic approach. Artif Intell Med 2007;39(2):127–36.

[2007] Sætre, Rune, Kazuhiro Yoshida, Akane Yakushiji, Yusuke Miyao, Yuichiro Matsubayashi and Tomoko Ohta. (2007) AKANE System: Protein-Protein Interaction Pairs in BioCreAtIvE2 Challenge, PPI-IPS subtask. Proceedings of the Second BioCreative Challenge Evaluation Workshop.

[2006] Chun, H., et al., "Extraction of Gene-Disease Relations from MedLine using Domain Dictionaries and Machine Learning, " Proc. Pacific Symp. on Biocomputing (PSB), 2006, pp. 4--15.

[2006] Chun, Hong-woo, Yoshimasa Tsuruoka, Jin-Dong Kim, Rie Shiba, Naoki Nagata, Teruyoshi Hishiki, Jun'ichi Tsujii (2006) Automatic Recognition of Topic-Classified Relations between Prostate Cancer and Genes using MEDLINE Abstracts. BMC-Bioinformatics. 7(Suppl 3). pp. S4, November 2006.

[2006] Kim JJ, Zhang Z, Park JC, et al. BioContrasts: extracting and exploiting protein-protein contrastive relations from biomedical literature. Bioinformatics 2006;22:597-605.

[2006] Kirsch,H., et al. (2006) Distributed modules for text annotation and IE applied to the biomedical domain. Int. J. Med. Inform. 75(6):496-500.

[2006] McNaught, J. & Black, W. (2006) Information Extraction, in Text Mining for Biology and Biomedicine, Ananiadou & McNaught (eds), Artech House, pp.143-178

[2006] Miyao, Yusuke, Tomoko Ohta, Katsuya Masuda, Yoshimasa Tsuruoka, Kazuhiro Yoshida, Takashi Ninomiya and Jun'ichi Tsujii (2006) Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. Proceedings of COLING-ACL 2006. Sydney, Australia, pp. 1017--1024, July 2006.

[2006] Saric,J., et al. (2006) Extraction of regulatory gene/protein networks from Medline. Bioinformatics 22(6):645-650.

[2005] Hoffmann,R. and Valencia,A. (2005) Implementing the iHOP concept for navigation of biomedical literature. Bioinformatics 21 (Suppl 2):ii252-8.

[2005] Hu, Z., et al., "Literature Mining and Database Annotation of Protein Phosphorylation Using a Rule-based System," Bioinformatics, Vol. 21, 2005, 2759--2765.

[2005] Jelier,R., et al. (2005) Co-occurrence based meta-analysis of scientific texts retrieving biological relationships between genes. Bioinformatics, 21(9):2049-58.

[2005] Leroy, G., and H. Chen, "Genescene: An Ontology-enhanced Integration of Linguistic and Co-occurrence Based Relations in Biomedical Texts," Journal of the American Society for Information Science and Technology, Vol. 56, No. 5, 2005, pp. 457--468.

[2005] Yakushiji, A., et al., "Biomedical Information Extraction with Predicate-Argument Structure Patterns," Proc. Int. Symp. on Semantic Mining in Biomedicine, 2005, pp. 60--69.

[2005] Šarić, J., L. J. Jensen, and I. Rojas, "Large-scale Extraction of Gene Regulation for Model Organisms in an Ontological Context," In Silico Biology, Vol. 5, No. 0004, 2004, http://www.bioinfo.de.isb/2004/05/0004/, June 2005

[2004] Alphonse, E., et al., "Event-based Information Extraction for the Biomedical Domain:The Caderige Project," Proc. Int. Workshop on Natural language Processing in Biomedicine and its Applications (JNLPBA), 2004, pp. 43--49.

[2004] Corney, D. P., et al., "BioRAT: Extracting Biological Information from Full-length Papers," Bioinformatics, Vol. 20, No. 17, 2004, pp. 3206--3213.

[2004] Daraselia, N., et al., "Extracting Human Protein Interactions from MEDLINE using a Full-sentence Parser," Bioinformatics, Vol. 20, No. 5, 2004, pp. 604--611.

[2004] Huang,M., et al. (2004) Discovering patterns to extract protein–protein interactions from full texts. Bioinformatics 20(18):3604-3612

[2004] Kim, J., and J. C. Park, "BioIE: Retargetable Information Extraction and Ontological Annotation of Biological Interactions from the Literature," Journal of Bioinformatics and Computational Biology, Vol. 3, No. 3, 2004, pp. 551--568.

[2004] McDonald, D.M., et al., "Extracting Gene Pathway Relations Using a Hybrid Grammar: The Arizona Relation Parser," Bioinformatics, Vol. 20, No. 18, 2004, pp. 3370--3378.

[2004] Pyysalo, S., et al., "Analysis of Link Grammar on Biomedical Dependency Corpus Targeted at Protein-Protein Interactions," Proc. Int. Joint Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA), 2004, pp. 15--21.

[2004] Rinaldi, F., et al., "Mining Relations in the GENIA Corpus," Proc. Second European Workshop on Data Mining and Text Mining for Bioinformatics, 2004, pp. 61--68.

[2004] Rzhetsky, A., et al., "GeneWays: A System for Extracting, Analyzing, Visualizing and Integrating Molecular Pathway Data," Journal of Biomedical Informatics, Vol. 37, No. 1, 2004, pp. 43--53.

[2004] Wattarujeekrit T, Shah PK, Collier N. PASBio: predicate argument structures for event extraction in molecular biology. BMC Bioinform 2004;5:155.

[2003] Ding, J., et al., "Extracting Biochemical Interactions from MEDLINE Using a Link Grammar Parser, " Proc. 15th IEEE Int. Conf. on Tools with Artificial Intelligence (ICTAI 2003), Sacramento, CA, Nov. 3--5, 2003,   pp. 467--473.

[2003] Gaizauskas, R., et al., "Protein Structures and Information Extraction from Biological Texts: The PASTA System," Bioinformatics, Vol. 19, No. 1, 2003, pp. 135--143.

[2003] Leroy, G., H. Chen, and J.D. Martinez, "A Shallow Parser based on Closed-Class Words to Capture Relations in Biomedical Text," Journal of Biomedical Informatics, Vol. 36, No. 3, 2003, pp.145--158.

[2003] Novichkova, S., S. Egorov, and N. Daroselia, "MedScan, a Natural Language Processing Engine for MEDLINE Abstracts," Bioinformatics, Vol. 19, 2003, pp. 1699--1706.

[2003] Temkin,J.M. and Gilder,M.R. (2003) Extraction of protein interaction information from unstructured text using a context-free grammar. Bioinformatics. 19(16):2046-53.

[2002] De Bruijn, B., and J. Martin, "Getting to the (C)ore of Knowledge: Mining Biomedical Literature," International Journal of Medical Informatics, Vol. 67, 2002, pp. 7--18.

[2002] Leroy, G., and H. Chen, "Filling Preposition-based Templates to Capture Information from Medical Abstracts," Proc. Pacific Symp. on Biocomputing 7 (PSB), 2002, pp. 362--373.

[2001] Friedman, C., et al., "GENIES: A Natural-language Processing System for the Extraction of Molecular Pathways from Journal Articles," Bioinformatics, Vol. 17, 2001, pp. S74--S82.

[2001] Ono, T., et al., "Extraction of Information on Protein-Protein Interactions from the Biological Literature," *Bioinformatics*, Vol. 17, No. 2, 2001, pp. 155--161.

[2001] Park, J.C. et al. (2001) Bidirectional incremental parsing for automatic pathway identification with combinatory categorical grammar. Pac Symp Biocomput :396-407.

[2001] Park, J.C., "Using Combinatory Categorial Grammar to Extract Biomedical Information," IEEE Intelligent Systems, Vol. 16, No. 1, 2001, pp.62--67.

[2001] Pustejovsky,J. et al. (2001) Robust relational parsing over biomedical literature: extracting inhibit relations. Pac Symp Biocomput :362-73.

[2001] Yakushiji, A., et al., "Event Extraction from Biomedical Papers Using a Full Parser,"   Proc. Pacific Symp. on Biocomputing (PSB 2001), Kauai, Hawaii, Jan. 3--7, 2001,   pp. 408--419.

(v) Annotation of Biomedical Corpora

[2008] Thompson, Paul, Giuila Venturi, John McNaught, Simonetta Montemagni and Sophia Ananiadou (2008). Categorising Modality in Biomedical Texts. LREC 2008 workshop "Building and Evaluating resources for biomedical text mining" Marrakech, Morocco.

[2008] Thompson, Paul, Philip Cotter, John McNaught, Sophia Ananiadou, Simonetta Montemagni, Andrea Trabucco and Giulia Venturi (2008). Building a bio-event annotated corpus for the acquisition of semantic frames from biomedical corpora. Sixth International Conference on Language Resources and Evaluation (LREC 2008), Marrakech, Morocco

[2006] Chou WC, Tsai RTH, Su YS, et al. A semi-automatic method for annotating a biomedical proposition bank. Proceedings of workshop on frontiers in linguistically annotated corpora 2006. Association for Computational Linguistics.Sydney, Australia, 2006:5–12.

[2006] Kim, J.D. & Tsujii, J. (2006) Corpora and their Annotation, in Text Mining for Biology and Biomedicine, Ananiadou, S. & McNaught, J. (eds), Artech House, pp. 179-211.

[2005] Cohen, K., et al., "Corpus Design for Biomedical Natural Language Processing," Proc. ACL Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics, 2005, pp. 38--45.

[2005] Smith, L.H., et al., "MedTag: a Collection of Biomedical Annotations," Proc. ACL Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics, 2005, pp. 32--37.

[2005] Tateisi, Y., et al., "Syntax Annotation for the GENIA corpus," Proc. IJCNLP Companion volume, 2005, pp. 222--227.

[2004] Pakhomov, S., A. Coden, and C. Chute, "Creating a Test Corpus of Clinical Notes Manually Tagged for Part-of-Speech Information," Proc. COLING Joint Workshop on Natural Language Processing in Biomedicine and its Applications, 2004, pp. 62--65.

[2004] Tateisi, Y., and J. Tsujii, "Part-of-Speech Annotation of Biology Research Abstracts," Proc. Int. Conf. on Language Resource and Evaluation (LREC), 2004, pp. 1267--1270.

[2003] Erjavec, T., et al., "Encoding Biomedical Resources in TEI: The Case of the GENIA Corpus," Proc. ACL Workshop on Natural Language Processing in Biomedicine, 2003, pp. 97--104.

[2003] Kim, J., et al., "GENIA Corpus---a Semantically Annotated Corpus for Bio-Textmining," Bioinformatics, Vol. 19, Suppl. 1, 2003, pp. i180--i182.

(vi) Tagging and Parsing

[2008] Miyao, Yusuke and Jun'ichi Tsujii. Feature Forest Models for Probabilistic HPSG Parsing. Computational Linguistics. 34(1). pp. 35–80, MIT Press, 2008.

[2007] Ninomiya T, Tsuruoka Y, Miyao Y, et al. Fast and scalable HPSG parsing. TAL 2005 2007;46(2).

[2007] Yoshida, Kazuhiro, Yoshimasa Tsuruoka, Yusuke Miyao and Jun'ichi Tsujii (2007) Ambiguous Part-of-Speech Tagging for Improving Accuracy and Domain Portability of Syntactic Parsers. Twentieth International Joint Conference on Artificial Intelligence (IJCAI)

[2005] Hara, T., Y. Miyao and J. Tsujii, "Adapting a probabilistic disambiguation model of an HPSG parser to a new domain," Proc. Int. Joint Conf. on Natural Language Processing, 2005.

[2005] Miyao, Y., and J. Tsujii, "Probabilistic disambiguation models for wide-coverage HPSG parsing," Proc. ACL, 2005, pp. 83--90.

[2005] Tsuruoka, Y., and J. Tsujii, "Bidirectional Inference with the Easiest-First Strategy for Tagging Sequence Data," Proc. HLT/EMNLP, 2005. pp. 467--474.

[2005] Tsuruoka, Y., et al., "Developing a Robust Part-of-Speech Tagger for Biomedical Text," Proc. 10th Panhellenic Conference on Informatics, 2005, pp. 382-392.

[2004] Smith, L., et al., "MedPost: a part-of-speech tagger for bioMedical text, Bioinformatics, Vol. 20, No. 14, 2004, pp. 2320-2321

[2003] Giménez, J., and L. Màrquez, "Fast and accurate part-of-speech tagging: The SVM approach revisited," Proc. RANLP, 2003, pp. 153-163.

[2003] Toutanova K., et al., "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," Proc. HLT-NAACL, 2003.

[2000] Brants, T., "TnT - A Statistical Part-of-Speech Tagger," Proc. Applied Natural Language Processing Conf., 2000.

[2000] Charniak, E., "A Maximum-Entropy-Inspired Parser," Proc. NAACL, 2000.

[1999] Collins, M., "Head-Driven Statistical Models for Natural Language Parsing," Ph.D. thesis, University of Pennsylvania, 1999.

[1997] Reynar, J., and A. Ratnaparkhi, "A Maximum Entropy Approach to Identifying Sentence Boundaries," Proc. ANLP, 1997, pp. 16-19.