

6. Bibliography

Allen, D. et al. 2012, **The Unicode Standard**, Version 6.2 ISBN 978-1-936213-07-8, September 2012.

Bellare, K. et al. 2007. **Learning extractors from unlabeled text using relevant databases**, In Sixth International Workshop on Information Integration on the Web, 2007.

Berners-Lee, T. et al. 2011, O. **The Semantic Web**. Scientific American 284 (5): 34-43, May, 2001.

Berners-Lee, T. et al. 2005, **Uniform Resource Identifier (URI): Generic Syntax**, IETF RFP 3986 (standard track), Internet Engineering Task Force, January 2005, Available at: www.ietf.org/rfc/rfc3986.txt

Berners-Lee, T. 2006. **Linked Data - Design Issues**. Retrieved at March 10th, 2014. Available at <http://www.w3.org/DesignIssues/LinkedData.html>.

Bizer, C. et al. 2009. **Linked Data - The Story So Far**. International Journal on Semantic Web and Information Systems 5 (3): 1-22, 2009.

Bray, T. et al. 2008. **Extensible Markup Language (XML) 1.0**, Fifth Edition, November 2008. Available at: <http://www.w3.org/TR/REC-xml/>

Brants, T. 2000. **A Statistical Part-of-Speech Tagger**, In 6th Applied Natural Language Processing Conference, 2000.

Breitman, K. et al. 2006. **Semantic Web: Concepts, Technologies and Applications**. Springer, 2006.

Brickley, D. et al. 2014. **RDF Schema 1.1**, February 2014, Available at: <http://www.w3.org/TR/rdf-schema/>

Cafarella, M. et al. 2011. **Structured Data on the Web**, Communications of the ACM, Vol. 54, No. 2. (February 2011), pp 72-79.

Chowdhury, G. 2003. **Natural Language Processing**, Annual Review of Information Science and Technology, 37. pp 51-89. ISSN 0066-4200

- Collins, M. 2002. **Discriminative training methods for hidden markov models: theory and experiments with perceptron algorithms**. In EMNLP, 2002.
- Craven, M. et al. 1999. **Constructing biological knowledge bases by extracting information from text sources**, In Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology, pages 77–86.
- Curran, J. et al. 2003. **Language independent NER using a maximum entropy tagger**. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003 - Volume 4, pages 164-167, Morristown, NJ, USA, 2003. Association for Computational Linguistics.
- DaConta, M. et al. 2003. **The Semantic Web: A guide to the future of XML, Web Services, and Knowledge Management**, John Wiley & Sons, Jul 21, 2003
- Finkel, J. et al. 2005. **Incorporating non-local information into information extraction systems by gibbs sampling**. In ACL, ACL '05, pages 363-370, 2005 .
- Gerber, D. et al. 2011. **Bootstrapping the Linked Data Web**, 1st Workshop on Web Scale Knowledge Extraction @ ISWC 2011' .
- Gillick, D. 2009. **Sentence Boundary Detection and the Problem with the U.S.**, In Proc. Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL: HLT), Companion Volume: Short Papers (2009), pp. 241-244
- Gruber, T. 1993. **A translation approach to portable ontology specifications**, Knowledge Acquisition, Vol. 5, No. 2, June 1993, pp. 199-220, doi:10.1006/knac.1993.1008
- Heath, T., 2008. **The Linking Open Data Project — Bootstrapping the Web of Data**. Available at: <http://tomheath.com/slides/2008-02-amsterdam-linking-open-data-project.pdf>, 2008.
- Heath, T., 2009. **Linked Data? Web of Data? Semantic Web? WTF?**. Retrieved May 4, 2013, from <http://tomheath.com/blog/2009/03/linked-data-web-of-data-semantic-web-wtf/>, 2009.
- Heath, T. et al., 2011. **Linked Data: Evolving the Web into a Global Data Space**, Synthesis Lectures on the Semantic Web, Morgan & Claypool Publishers 2011
- Hoffmann, R. et al. 2011. **Knowledge-based weak supervision for information extraction of overlapping relations**. In HLT '11 Proceedings of the 49th

Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, pp 541-550

Kiss, T. et al., 2006. **Unsupervised Multilingual Sentence Boundary Detection**, Computational Linguistics, 34(4):485-525

Kotsiantis, S. 2007. **Supervised machine learning: A review of classification techniques**, Informatica, Vol. 31 (2007), pp. 249-268.

Lafferty, J. et al. 2001. **Conditional random fields: Probabilistic models for segmenting and labeling sequence data**. In In The Eighteenth International Conference on Machine Learning, 2001.

Loper, E. et al., 2002. **NLTK: The Natural Language Toolkit**, In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, 19(2):313-330

Manning, C. 2001. **Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?**, Proceedings, Part I. Lecture Notes in Computer Science 6608, pp.171-189 Springer.

McCallum, A. 2005. **Information Extraction: Distilling Structured Data from Unstructured Text**, ACM Queue, Vol. 3, 9 (November 2005), 48-57.

McGuinness, D. et al., 2004. **OWL web ontology language overview**, W3C Recommendation, Vol. 10, Issue 2004-03, pp 10.

Mikheev, A. 2002. **Periods, Capitalized Words, etc.**, Computational Linguistics, 28(3):289-318

Mintz, M. et al., 2009. **Distant supervision for relation extraction without labeled data**, In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2. ACL '09. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 1003-1011, 2009.

Mitchel, M. et al., 1994. **The Penn Treebank: annotating predicate argument structure**, HLT '94 Proceedings of the workshop on Human Language Technology, pp 114-119.

Nguyen, T. et al., 2007. **Keyphrase Extraction**. In Scientific Publications, pp. 317-326. 2007.

Quinlan, J. 1993. **C4.5: Programs for machine learning**. Morgan Kaufmann, San Francisco

- Ratnaparkhi, A. 1996. **A maximum entropy model for part-of-speech tagging.** In EMNLP, Philadelphia, PA, 1996.
- Reynar, J. et al., 1997. **A maximum entropy approach to identifying sentence boundaries.** In Proceedings of the Fifth Conference on Applied Natural Language Processing, pages 16-19.
- Riedel, S. et al. 2010. **Modeling relations and their mentions without labeled text,** In Proceedings of the Sixteenth European Conference on Machine Learning (ECML-2010), 148–163.
- Sahnwaldt, C., 2012. **The DBpedia ontology,** Available at <http://wiki.dbpedia.org/ontology>, 2012.
- Shen, L. et al., 2007. **Guided learning for bidirectional sequence classification,** Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL 2007), pp 760-767.
- Soderland, S. et al., 1999. **Learning Information Extraction Rules for Semi-Structured and Free Text,** In Machine Learning 34, pages 233–272.
- Sogaard, A. et al., 2011. **Semi-supervised condensed nearest neighbor for part-of-speech tagging,** The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT). Portland, Oregon.
- Toutanova, K. et al., 2000. **Enriching the knowledge sources used in a maximum entropy part-of-speech tagger.** In EMNLP, 2000.
- Toutanova, K. et al., 2003. **Feature-rich part-of-speech tagging with a cyclic dependency network.** In Proceedings of the AAAI'94 Workshop on CaseBased Reasoning, pages 252-259, 2003.
- Winston, P. 1977. **Artificial Intelligence: Second Edition,** Addison-Wesley Publishing Company, ISBN: 0-201-08259-4, 1977.
- Wu, F. et al. 2007. **Autonomously semantifying wikipedia.** In Proceedings of the International Conference on Information and Knowledge Management (CIKM-2007), pages 41–50.
- Wu, F. et al. 2008. **Automatically refining the Wikipedia infobox ontology,** In Proceedings of the 17th International Conference on World Wide Web, pp. 635-644. ACM, 2008.

Zhou, G. et al., 2002. **Named entity recognition using an HMM-based chunk tagger**. In ACL '02, pages 473-480, 2002.