UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Ontology enrichment from heterogeneous sources on the web

de Boer, V.

Publication date
2010

Link to publication

Citation for published version (APA):
de Boer, V. (2010). *Ontology enrichment from heterogeneous sources on the web*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:07 Oct 2026

BIBLIOGRAPHY

Eugene Agichtein and Luis Gravano. Snowball: Extracting relations from large plain-text collections. In *Proceedings of the Fifth ACM International Conference on Digital Libraries*, 2000. (Cited on page 38.)

D.D. Ahn, J. van Rantwijk, and M. de Rijke. A cascaded machine learning approach to interpreting temporal expressions. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 420–427, Rochester, New York, April 2007. Association for Computational Linguistics, Association for Computational Linguistics. (Cited on page 68.)

James F. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983. (Cited on pages 58 and 68.)

Omar Alonso, Michael Gertz, and Ricardo Baeza-Yates. On the value of temporal information in information retrieval. *SIGIR Forum*, 41(2):35–41, 2007. ISSN 0163-5840. (Cited on page 41.)

Anjo Anjewierden. toko and sigmund: text analysis support for ontology development and social research. http://www.toko-sigmund.org, 2006. (Cited on pages 9, 75, 78, 81, and 131.)

Anjo Anjewierden, Bob J. Wielinga, and Robert de Hoog. Task and domain ontologies for knowledge mapping in operational processes. *Metis Deliverable 4.2/2003, University of Amsterdam.*, 2004. (Cited on page 17.)

Kedar Bellare and Andrew McCallum. Generalized expectation criteria for bootstrapping extractors using record-text alignment. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 131–140, Singapore, August 2009. Association for Computational Linguistics. (Cited on page 4.)

Tim Berners-Lee and Robert Cailliau. Worldwideweb: Proposal for a hypertext project. Technical report, CERN, 1990. URL `http://www.w3.org/Proposal`. (Cited on page 1.)

Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web: Scientific american. *Scientific American*, May 2001. (Cited on page 1.)

Branimir Boguraev and Rie Kubota Ando. Timeml-compliant text analysis for temporal reasoning. In *IJCAI'05: Proceedings of the 19th international joint conference on Artificial intelligence*, pages 997–1003, San Francisco, CA, USA, 2005. Morgan Kaufmann Publishers Inc. (Cited on page 68.)

Sergey Brin. Extracting patterns and relations from the world wide web. In *WebDB Workshop at 6th International Conference on Extending Database Technology, EDBT'98*, 1998. (Cited on page 38.)

Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998. (Cited on page 14.)

Paul Buitelaar, Philipp Cimiano, and Bernardo Magnini, editors. *Ontology Learning from Text: Methods, Evaluation and Applications*. Amsterdam: IOS Press, 2005. (Cited on page 2.)

Razvan Bunescu and Raymond Mooney. Subsequence kernels for relation extraction. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 171–178. MIT Press, Cambridge, MA, 2006. (Cited on pages 3 and 38.)

Razvan Bunescu and Raymond Mooney. Learning to extract relations from the web using minimal supervision. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 576–583, Prague, Czech Republic, June 2007. Association for Computational Linguistics. (Cited on page 4.)

Mary E. Califf and Raymond J. Mooney. Bottom-up relational learning of pattern matching rules for information extraction. *J. Mach. Learn. Res.*, 4:177–210, 2003. ISSN 1533-7928. (Cited on pages 9 and 77.)

Nancy Chinchor. *Proceedings of the Seventh Message Understanding Conference*. Science Applications International Corporation (SAIC), San Francisco, CA, 1998. (Cited on page 61.)

Rudi Cilibrasi and Paul M. B. Vitanyi. Automatic meaning discovery using google. *http://xxx.lanl.gov/abs/cs.CL/0412098*, 2004. (Cited on pages 34, 76, 109, and 121.)

Philip Cimiano, Lars Schmidt-Thieme, Aleksander Pivk, and Steffen Staab. Learning taxonomic relations from heterogeneous evidence. *Proceedings of the ECAI 2004 Ontology Learning and Population Workshop*, 2004. (Cited on pages 2, 11, 37, and 126.)

Philipp Cimiano. Ontology learning and population. *Proceedings Dagstuhl Seminar Machine Learning for the Semantic Web*, 2005. (Cited on page 12.)

Fabio Ciravegna, Sam Chapman, Alexiei Dingli, and Yorick Wilks. Learning to harvest information for the semantic web. In *Proceedings of the 1st European Semantic Web Symposium*, 2004. (Cited on pages 36 and 126.)

Mark Craven, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Seán Slattery. Learning to extract symbolic knowledge from the world wide web. In *AAAI '98/IAAI '98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, pages 509–516, Menlo Park, CA, USA, 1998. American Association for Artificial Intelligence. ISBN 0-262-51098-7. (Cited on page 2.)

Valter Crescenzi and Giansalvatore Mecca. Automatic information extraction from large websites. *J. ACM*, 51(5):731–779, 2004. ISSN 0004-5411. (Cited on page 38.)

Viktor de Boer, Maarten van Someren, and Bob J. Wielinga. A redundancy-based method for relation instantiation from the web. In Gerhard Brewka, Silvia Coradeschi, Anna Perini, and Paolo Traverso, editors, *ECAI 2006, 17th European Conference on Artificial Intelligence, August 29 - September 1, 2006, Riva del Garda, Italy, Including Prestigious Applications of Intelligent Systems (PAIS 2006), Proceedings*, pages 749 – 750. IOS Press, inc., 2006a. isbn=1-58603-642-4. (Cited on page 11.)

Viktor de Boer, Maarten van Someren, and Bob J. Wielinga. Extracting instances of relations from web documents using redundancy. In Y. Sure and J. Domingue, editors, *Proceedings of the Third European Semantic Web Conference*, number 4011 in LNCS, pages 245–258. Springer-Verlag Berlin Heidelberg, 2006b. (Cited on page 11.)

Viktor de Boer, Maarten van Someren, and Bob J. Wielinga. Extracting art style periods from the web. In Paolo Bouquet, Roberto Brunelli, Jean-Pierre Chanod, Claudia Niederée, and Heiko Stoermer, editors, *Proceedings of the ESWC 2006 Workshop on Mastering the Gap: From Information Extraction to Semantic Representation, Budva, Montenegro, June 12, 2006*, number online http://ceur-ws.org/Vol-187/2.pdf in CEUR Workshop Proceedings ISSN 1613-0073, 2006c. (Cited on page 41.)

Viktor de Boer, Maarten van Someren, and Bob J. Wielinga. Relation instantiation for ontology population using the web. In C. Freksa, M. Kohlhase, and K. Schill, editors, *Proceedings of the 29th Annual German Conference on AI, KI 2006, Bremen, Germany, june 14-17, 2006*, number 4314 in LNCS. Springer-Verlag Berlin Heidelberg, 2006d. (Cited on page 11.)

Viktor de Boer, Maarten van Someren, and Bob J. Wielinga. A redundancy-based method for the extraction of relation instances from the web. *International Journal of Human-Computer Studies*, 65(9):816–831, 2007. (Cited on page 11.)

Doug Downey, Oren Etzioni, and Stephen Soderland. A probabilistic model of redundancy in information extraction. In *IJCAI*, pages 1034–1041, 2005. (Cited on page 38.)

Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.*, 165(1): 91–134, 2005. ISSN 0004-3702. (Cited on page 38.)

Christiane Fellbaum. *WordNet: An Electronical Lexical Database*. The MIT Press, Cambridge, MA, 1998. (Cited on pages 6, 50, 83, and 126.)

Lisa Ferro. Time expression recognition and normalization evaluation. website, 2004. http://timex2.mitre.org/tern.html. (Cited on page 62.)

Lisa Ferro, Laurie Gerber, Inderjeet Mani, Beth Sundheim, and George Wilson. Tides 2005 standard for the annotation of temporal expressions, 2005. http://fofoca.mitre.org/annotation_guidelines/2005_timex2_standard_v1.1.pdf. (Cited on pages 44, 61, and 130.)

Gijs Geleijnse and Jan Korst. Automatic ontology population by googling. In *Proceedings of the Seventeenth Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 2005)*, pages 120 – 126, Brussels, Belgium, 2005. (Cited on page 38.)

Gijs Geleijnse, Jan Korst, and Viktor de Boer. Instance classification using co-occurrences on the web. In *Proceedings of the ISWC 2006 workshop on Web Content Mining (WebConMine)*, Athens, GA, November 2006. (Cited on pages 34, 38, 76, 81, 102, and 130.)

Risto Gligorov, Warner T. Kate, Zharko Aleksovski, and Frank van Harmelen. Using google distance to weight approximate ontology matches. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 767–776, New York, NY, USA, 2007. ACM Press. ISBN 9781595936547. (Cited on page 3.)

Trond Grenager, Dan Klein, and Christopher D. Manning. Unsupervised learning of field segmentation models for information extraction. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 371–378, Morristown, NJ, USA, 2005. Association for Computational Linguistics. (Cited on page 3.)

Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6): 907–928, 1995. ISSN 1071-5819. (Cited on page 2.)

Marti. A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 529–545, 1992. (Cited on pages 12, 38, and 126.)

Jerry R. Hobbs and Feng Pan. An ontology of time for the semantic web. *CM Transactions on Asian Language Information Processing*, 3(1), March 2004. (Cited on pages 61 and 69.)

Neil Ireson, Fabio Ciravegna, Mary Elaine Califf, Dayne Freitag, Nicholas Kushmerick, and Alberto Lavelli. Evaluating machine learning for information extraction. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 345–352, New York, NY, USA, 2005. ACM. ISBN 1-59593-180-5. (Cited on page 4.)

Sophia Katrenko. *A closer look at learning relations from text*. PhD thesis, Universiteit van Amsterdam, 2009. (Cited on page 135.)

Sophia Katrenko and Pieter W. Adriaans. Learning relations from biomedical corpora using dependency trees. *Knowledge Discovery and Emergent Complexity in Bioinformatics, Lecture Notes in Bioinformatics. LNBI*, 4366:61–80, 2007. (Cited on page 134.)

Jan Korst, Gijs Geleijnse, Nick de Jong, and Michael Verschoor. Ontology-based extraction of information from the World Wide Web. In *Intelligent Algorithms in Ambient and Biomedical Computing*, Philips Research Book Series, pages 149 – 167. Springer, 2006. (Cited on page 126.)

Nicholas Kushmerick, Daniel S. Weld, and Robert Doorenbos. Wrapper induction for information extraction, 1997. (Cited on pages 4 and 12.)

Erich L. Lehmann and H. J. M. D'Abrera. *Nonparametrics: Statistical Methods Based on Ranks*. NJ: Prentice-Hall, 1998. (Cited on page 34.)

Wentian Li. Random texts exhibit zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6):1842–, 1992. (Cited on pages 86, 104, and 133.)

Winston Lin, Roman Yangarber, and Ralph Grishman. Bootstrapped learning of semantic classes from positive and negative examples. In *Proceedings of the ICML-2003 Workshop on Continuum from Labeled to Unlabeled Data*, Washington DC, 2003. (Cited on page 15.)

Dolores Llido, Rafael Berlanga Llavori, and Maria José Aramburu Cabo. Extracting temporal references to assign document event-time periods. In *Proceedings of the 12th International Conference on Database and Expert Systems Applications*, pages 62–71, September 2001. (Cited on page 68.)

Alexander Maedche and Steffen Staab. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 13:993, 2001. (Cited on pages 2 and 11.)

Véronique Malaisé, Antoine Isaac, Luit Gazendam, and Hennie Brugman. Anchoring dutch cultural heritage thesauri to wordnet: Two case studies. In *Proceedings of the Workshop on Language Technology for Cultural Heritage Data (LaTeCH 2007).*, pages 57–64, Prague, Czech Republic, June 2007. Association for Computational Linguistics. (Cited on page 3.)

Inderjeet Mani. Recent developments in temporal information extraction. In Nicolas Nicolov, Kalina Bontcheva, Galia Angelova, and Ruslan Mitkov, editors, *RANLP*, volume 260 of *Current Issues in Linguistic Theory (CILT)*, pages 45–60. John Benjamins, Amsterdam/Philadelphia, 2003. ISBN 1-58811-618-2. (Cited on page 42.)

Inderjeet Mani and George Wilson. Robust temporal processing of news. In *ACL '00: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 69–76, Morristown, NJ, USA, 2000. Association for Computational Linguistics. (Cited on page 68.)

Andrew McCallum. Information extraction: Distilling structured data from unstructured text. *Queue*, 3(9):48–57, 2005. ISSN 1542-7730. (Cited on page 3.)

Luke McDowell and Michael Cafarella. Ontology-driven information extraction with ontosyphon. In *5th International Semantic Web Conference (ISWC2006)*, 2006. (Cited on page 126.)

Matthew Michelson and Craig A. Knoblock. Unsupervised information extraction from unstructured, ungrammatical data sources on the world wide web. *International Journal on Document Analysis*, 10(3-4):211–226, December 2007. (Cited on page 4.)

Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997. (Cited on page 113.)

Raymond J. Mooney and Razvan Bunescu. Mining knowledge from text using information extraction. *SIGKDD Explor. Newsl.*, 7(1):3–10, June 2005. ISSN 1931-0145. doi: 10.1145/1089815.1089817. (Cited on pages 3 and 4.)

Matteo Negri and Lucca. Marseglia. Recognition and normalization of time expressions: Itc-irst at tern 2004, technical report. Technical report, ITC-irst, Trento, 2004. (Cited on page 68.)

Hans Jürgen Ohlbach. Relations between fuzzy time intervals. In *Proceedings 11th. International Symposium on Temporal Representation and Reasoning*, pages 44–51, 2004. (Cited on page 69.)

James Pustejovsky, Jose Castano, Robert Ingria, Roser Sauri, Robert Gauzauskas, Andrea. Setzer, and Graham Katz. Timeml: Robust specification of event and temporal expression in text. *IWCS-5, Fifth International Workshop on Computational Semantics.*, 2003. (Cited on page 62.)

Ellen Riloff and Rosie Jones. Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI '99/IAAI '99: Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligence conference innovative applications of artificial intelligence*, pages 474–479, Menlo Park, CA, USA, 1999. American Association for Artificial Intelligence. ISBN 0-262-51106-1. (Cited on page 12.)

Steven Schockaert and Martine De Cock. Temporal reasoning about fuzzy intervals. *Artif. Intell.*, 172(8-9):1158–1193, 2008. ISSN 0004-3702. (Cited on page 69.)

Steven Schockaert, Martine De Cock, and Etienne E. Kerre. Acquiring vague temporal information from the web. In *Web Intelligence/IAT Workshops*, pages 265–268. IEEE, 2008. (Cited on page 68.)

Guus Schreiber, Alia Amin, Mark van Assem, Viktor de Boer, Lynda Hardman, Michiel Hildebrand, Laura Hollink, Zhisheng Huang, Janneke van Kersen, Marco de Niet, Borys Omelayenko, Jacco van Ossenbruggen, Ronny Siebes, Jos Taekema, Jan Wielemaker, and Bob Wielinga. Multimedian e-culture demonstrator. In *Proceedings of the Fifth International Semantic Web Conference (ISWC2006), Athens, Georgia*, volume 4273 of *Lecture Notes in Computer Science*, pages 951–958, Athens, Georgia, November 2006. Springer-Verlag. Winner Semantic Web Challenge 2006. (Cited on pages 6 and 17.)

Guus Schreiber, Alia Amin, Lora Aroyo, Marc van Assem, Viktor de Boer, Lynda Hardman, Michiel Hildebrand, Borys Omelayenko, Jacco van Ossenbruggen, and Anna Tordai. Semantic annotation and search of cultural-heritage collections: The multimedian e-culture demonstrator. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):243–249, November 2008. ISSN 15708268. (Cited on pages 49 and 107.)

The Getty Foundation. Aat: Art and architecture thesaurus. http://www.getty.edu/research/conducting_research/vocabularies/aat, 2000a. (Cited on pages 17, 49, and 118.)

The Getty Foundation. Tgn: Getty thesaurus of geographic names. http://www.getty.edu/research/conducting_research/vocabularies/tgn, 2000b. (Cited on page 84.)

The Getty Foundation. Ulan: Union list of artist names. http://www.getty.edu/research/conducting_research/vocabularies/ulan, 2000c. (Cited on pages 17, 54, 92, 118, and 121.)

Mike Uschold and Michael Gruninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11:93–136, 1996. (Cited on page 2.)

Willem van Hage, Hap Kolb, and Guus Schreiber. A method for learning part-whole relations. In *The Semantic Web - ISWC 2006*, pages 723–735, 2006. (Cited on page 3.)

Cornelis J. van Rijsbergen. *Information Retrieval*. Butterworths, London, 1979. (Cited on page 18.)

Nicolas Weber and Paul Buitelaar. Web-based ontology learning with isolde. In *Proceedings of ISWC2006 Workshop on Web Content Mining with Human Language Technologies*, 2006. (Cited on page 30.)

Ian H. Witten and Eibe Frank. *"Data Mining: Practical machine learning tools and techniques", 2nd Edition*. Morgan Kaufmann, San Francisco, 2005. (Cited on pages 113 and 116.)

Dmitry Zelenko, Chinatsu Aone, Anthony Richardella, Jaz K, Thomas Hofmann, Tomaso Poggio, and John Shawe-taylor. Kernel methods for relation extraction. *Journal of Machine Learning Research*, 3:2003, 2003. (Cited on pages 3 and 38.)

Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison, 2005. http://www.cs.wisc.edu/~jerryzhu/pub/ssl_survey.pdf. (Cited on page 15.)

George K. Zipf. *Selective Studies and the Principle of Relative Frequency in Language*. Harvard University Press, 1932. (Cited on page 112.)

George K. Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley (Reading MA), 1949. (Cited on pages 86, 104, and 133.)