

Seminar Sommersemester 2011: Wissensextraktion aus Wikipedia

Michael Strube
Heidelberger Institut für Theoretische Studien
Schloss-Wolfsbrunnenweg 35
69118 Heidelberg
<http://michael.kimstrube.de>

18. Juli 2011

Termine, Themenvorschläge

14.04.2011

Überblick: Wissen in der CL – WordNet, Cyc, Ontologielernen, Wikipedia: Medelyan et al. (2009)

21.04.2011

Alternativen: WordNet, Cyc, Ontologielernen aus Texten: Lenat & Guha (1990), Lenat et al. (1990), Hearst (1992), Hearst (1998), Fellbaum (1998), Berland & Charniak (1999), Caraballo (1999), Riloff & Jones (1999), Agichtein & Gravano (2000), Pantel & Ravichandran (2004), Cimiano et al. (2004), Matuszek et al. (2005), Ahn et al. (2006), Girju et al. (2006), Pantel & Pennacchiotti (2006), Snow et al. (2006)

zur Vorbereitung: Hearst (1992) oder (Berland & Charniak, 1999) oder Snow et al. (2006)

28.04.2011

WikiRelate!, WikiTaxonomy, WikiNet, semantische Verwandtschaft, semantische Ähnlichkeit, BabelNet: Strube & Ponzetto (2006), Ponzetto & Strube (2007b), Ponzetto & Strube (2007a), Ponzetto & Strube (2011), Nastase & Strube (2008), Nastase et al. (2010), Navigli & Ponzetto (2010)

zur Vorbereitung: Strube & Ponzetto (2006) oder Ponzetto & Strube (2007a)

05.05.2011

fällt aus, *CoSyne Pre-Review Meeting* und externer Vortrag

12.05.2011

fällt aus, *Google Faculty Summit*

19.05.2011

Referat: Hans-Martin Ramsel – Yago: Suchanek et al. (2007), Kasneci et al. (2008), Suchanek et al. (2008), Weikum et al. (2009)

Referat: Julian Bastow – DBPedia: Auer et al. (2007), Bizer et al. (2009)

zur Vorbereitung: Suchanek et al. (2007) oder Auer et al. (2007)

26.05.2011

Referat: Martina Trognitz – Kylin, KOG, Intelligence in Wikipedia: Wu & Weld (2010), Weld et al. (2008b)

Referat: Eric Hildebrand – Relation Extraction: (Nguyen et al., 2007; Nguyen & Moschitti, 2011)

zur Vorbereitung: Wu & Weld (2010) oder Weld et al. (2008b) oder Nguyen et al. (2007)

optional: Wu & Weld (2007), Wu & Weld (2008), Weld et al. (2008a)

09.06.2011

Vertretung durch Vivi Nastase

Referat: Xiaoxi Pang – Text in Wikipedia, semantische Verwandtschaft: Gabrilovich & Markovitch (2007), Gabrilovich & Markovitch (2009)

Referat: Ulzhan Kadirbayeva – Hyperlinks, semantische Verwandtschaft: Milne & Witten (2008a)

zur Vorbereitung (bitte die Fragen auch an vivi.nastase@h-its.org schicken): Gabrilovich & Markovitch (2007) oder Milne & Witten (2008a)

16.06.2011

fällt aus, *CoSyne Review Meeting*

30.06.2011

Referat: Saskia Vola – Automatische Übersetzung: Yasud & Sumita (2008), Yuncong & Fung (2010)

Referat: Rebecca Mayr – Ressourcen für Eigennamenerkennung: Bunesco & Paşca (2006), Cucerzan (2007), Wentland et al. (2008)

07.07.2011

Referat: Gesa Stupperich – Konzeptdisambiguierung, Wikification: Milne & Witten (2008b), Turdakov & Lizorkin (2009)

Referat: Thierry Göckel – Konzeptdisambiguierung, Wikification: Kulkarni et al. (2009)

zur Vorbereitung: Milne & Witten (2008b) oder Turdakov & Lizorkin (2009) oder Kulkarni et al. (2009)

optional – Konzeptdisambiguierung, Wikification, Wortartendisambiguierung: Mihalcea & Csomai (2007), Csomai & Mihalcea (2008), Turdakov & Lizorkin (2010), Zhou (2010), Ponzetto & Navigli (2010), Bentivogli et al. (2010)

14.07.2011

Linguistisches Wissen:

Referat Nadya Georgieva – Spelling Correction: Nelken & Yamangil (2008), Max & Wisniewski (2010)

Referat: Eugen Ruppert – Linguistische Hecken: Ganter & Strube (2009), Farkas et al. (2010), Ganter (2010)

Referat: Hobli Taffame – Wortgrenzen: Gabay et al. (2008)

zur Vorbereitung: zwei Papiere aus Nelken & Yamangil (2008, pp.32-33) oder Max & Wisniewski (2010) oder Ganter & Strube (2009) oder Gabay et al. (2008)

21.07.2011

Referat: Dustin Heckmann – Qualität, Vertrauen (*trust*), ...: Dalip et al. (2009)

optional: Beschastnikh et al. (2008), Spinellis & Louridas (2008), Stvilia et al. (2008), Gorgeon & Swanson (2009), Halfaker et al. (2009), Wöhner & Peters (2009), Arazy & Nov (2010), Lucassen & Schraagen (2010), Niederer & van Dijk (2010)

Referat: Mareike Hartmann – Topic Indexing: Medelyan et al. (2008)

Referat: Daniel Jovanovic – Wikipedia und Semantic Web: Kröttsch et al. (2005)

optional: Blohm et al. (2008)

zur Vorbereitung: Dalip et al. (2009) oder Medelyan et al. (2008) oder Kröttsch et al. (2005)

Weitere Themen:

Document Clustering: Huang et al. (2009)

Automatische Zusammenfassung: Nelken & Yamangil (2008), Yamangil & Nelken (2008)

Bemerkungen

Leistungsnachweise: Lektüre und aktive Teilnahme (1/3), Referat (1/3), Hausarbeit (1/3). Hausarbeit: 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Bibliographie. Die Hausarbeit kann auch per Email an mich geschickt werden, aber *nicht* als Word-Datei sondern nur als PDF-Datei. – Ich empfehle, wissenschaftliche Texte mit Latex und Bibtex zu verfassen.

Regelmäßige Teilnahme (d.i. nicht mehr als einmal unentschuldigtes Fehlen) ist Voraussetzung für den Scheinerwerb. Zu jeder Sitzung müssen jeweils zwei Fragen (!) zu

einem Papier abgegeben werden, das in der aktuellen Sitzung vorgestellt wird. Abgabe entweder per Email bis spätestens 13 Uhr am Tag der Sitzung oder schriftlich direkt vor der Sitzung. Dies geht in die Bewertung für aktive Teilnahme am Seminar ein.

Literatur: Viele Papiere können direkt aus der *ACL Anthology* kopiert werden (<http://acl.ldc.upenn.edu/>), insbesondere alle Papiere der (*E/NA*)*ACL*-, *Coling*- und *EMNLP*-Konferenzen, alle Workshops, die im Rahmen dieser Konferenzen veranstaltet wurden und die Zeitschrift *Computational Linguistics*. Papiere, die von der *AAAI* publiziert wurden (*AAAI*-Konferenz, *AAAI*-Workshops, *AAAI*-Symposia, etc.) sind in der *AAAI Digital Library* verfügbar (<http://www.aaai.org/Library>). – Die meisten weiteren Zeitschriften sind elektronisch verfügbar über die UB (<http://rzblx1.uni-regensburg.de/ezeit/search.phtml?bibid=UBHE>) – oder stehen dort im Regal.

Sprechstunde: Auf Vereinbarung (Email, Telefon) bei mir im Büro, ggf. auch im Anschluß an das Seminar.

Hausarbeiten:

Maximal 8-10 Seiten (Proseminar), 12-15 Seiten (Hauptseminar) inkl. Abbildungen, inkl. Literaturverzeichnis.

Inhalt: Fokus auf das vorgestellte Papier; NICHT *Related Work*-Kapitel referieren, wenn die entsprechenden Papiere nicht gelesen wurden; Evaluierung berichten; WICHTIG: mit eigener Meinung oder Bewertung abschließen.

Stil: Wissenschaftlichkeit drückt sich nicht durch lange, komplizierte Sätze und exzessiven Gebrauch von Fremdwörtern aus – deshalb bitte kurze Sätze, einfache Sprache; Hausarbeiten vor der Abgabe Korrektur lesen oder Korrektur lesen lassen (s. auch *Dos and donts: Hinweise zur Abfassung wissenschaftlicher Arbeiten* von Prof. Frank – http://www.cl.uni-heidelberg.de/~frank/materials/dos_and_donts.pdf).

Seminararbeit (d.i. eine praktische Arbeit) ist auch möglich. Sollte durch 5-6 Seiten Bericht begleitet werden.

Abgabetermin: bis spätestens 15. August 2011; per Email als PDF-Datei (kein Microsoft Word!) oder ausgedruckt per Post – Matrikelnummer und Studiengang nicht vergessen!

Literatur

- Agichtein, Eugene & Luis Gravano (2000). Extracting relations from large plain text collections. In *Proceedings of the Fifth ACM Conference on Digital Libraries*, pp. 85–94.
- Ahn, Luis von, Mihir Kedia & Manuel Blum (2006). Verbosity: A gaem for collecting common-sense facts. In *Proceedings of the Conference on Human Factors in Computing Systems*, Montréal, Québec, Canada, 22–27 April 2006, pp. 75–78.
- Arazy, Ofer & Oded Nov (2010). Determinants of Wikipedia quality: the roles of global and local contribution inequality. In *Proceedings of the ACM 2010 Conference on Computer Supported Cooperative Work (CSCW-10)*, Savannah, Ga., 6–10 February 2010, pp. 233–236.
- Auer, Sören, Christian Bizer, Jens Lehmann, Georgi Kobilarov, Richard Cyganiak & Zachary Ives (2007). DBpedia: A nucleus for a Web of open data. In *Proceedings of the 6th International Semantic Web Conference and 2nd Asian Semantic Web Conference*, Busan, Korea, November 11–15, 2007, pp. 722–735.
- Bentivogli, Luisa, Pamela Forner, Claudio Giuliano, Alessandro Marchetti, Emanuele Pianta & Kateryna Tymoshenko (2010). Extending English ACE 2005 corpus annotation with ground-truth links to Wikipedia. In *Proceedings of the 2nd Workshop on The People’s Web: Collaboratively Constructed Semantic Resources*, Beijing, China, 28 August 2010, pp. 19–27.
- Berland, Matthew & Eugene Charniak (1999). Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Md., 20–26 June 1999, pp. 57–64.
- Beschastnikh, Ivan, Travis Kriplean & David W. McDonald (2008). Wikipedian self-governance in action: Motivating the policy lens. In *Proceedings of the 2nd International Conference on Weblogs and Social Media*, Seattle, Wash., Penn., March 30 – 2 April 2008.
- Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak & Sebastian Hellmann (2009). DBpedia – A crystallization point for the web of data. *Journal of Web Semantics*, 7:154–165.
- Blohm, Sebastian, Markus Krötzsch & Philipp Cimiano (2008). The fast and the numerous – Combining machine and community intelligence for semantic annotation. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 1–6.
- Bunescu, Razvan & Marius Paşca (2006). Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, Italy, 3–7 April 2006, pp. 9–16.
- Caraballo, Sharon A. (1999). Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, College Park, Md., 20–26 June 1999, pp. 120–126.
- Cimiano, Philipp, Siegfried Handschuh & Steffen Staab (2004). Towards the self-annotating web. In *Proceedings of the 13th World Wide Web Conference*, New York, N.Y., 17–22 May 2004, pp. 462–471.
- Csomai, Andras & Rada Mihalcea (2008). Linking documents to encyclopedic knowledge. *IEEE Intelligent Systems*, 23(5):34–41.
- Cucerzan, Silviu (2007). Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Language Learning*, Prague, Czech Republic, 28–30 June 2007, pp. 708–716.
- Dalip, Daniel Hasan, Marcos André Gonçalves, Marco Cristo & Pável Calada (2009). Automatic quality assessment of content created collaboratively by web communities: A case study of Wikipedia. In *Proceedings of the 9th Joint Conference on Digital Libraries*, Austin, Texas, 14–19 June 2009, pp. 295–304.
- Farkas, Richárd, Veronika Vincze, György Móra, János Csirik & György Szarvas (2010). The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the Shared Task of 14th Conference on Computational Natural Language Learning*, Uppsala, Sweden, 15–16 July 2010, pp. 1–12.

- Fellbaum, Christiane (Ed.) (1998). *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.
- Gabay, David, Ziv Ben-Eliahu & Michael Elhadad (2008). Using Wikipedia links to construct word segmentation corpora. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 61–63.
- Gabrilovich, Evgeniy & Shaul Markovitch (2007). Computing semantic relatedness using Wikipedia-based explicit semantic analysis. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, Hyderabad, India, 6–12 January 2007, pp. 1606–1611.
- Gabrilovich, Evgeniy & Shaul Markovitch (2009). Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research*, 34:443–498.
- Ganter, Viola (2010). Detecting vagueness in two languages and two domains. In *Proceedings of the Konferenz zur Verarbeitung Natürlicher Sprache (KONVENS)*, Saarbrücken, Germany, 6–8 September, 2010, pp. 141–145.
- Ganter, Viola & Michael Strube (2009). Finding hedges by chasing weasels: Hedge detection using Wikipedia tags and shallow linguistic features. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Singapore, 2–7 August 2009, pp. 173–176.
- Girju, Roxana, Adriana Badulescu & Dan Moldovan (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.
- Gorgeon, Arnaud & E. Burton Swanson (2009). Organizing the vision for Web 2.0: A study of the evolution of the concept in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, Orlando, Fl., 25–27 October 2009. 4 pages.
- Halfaker, Aaron, Aniket Kittur, Robert Kraut & John Riedl (2009). A jury of peers: Quality, experience and ownership in Wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, Orlando, Fl., 25–27 October 2009. 10 pages.
- Hearst, Marti (1998). Automated discovery of WordNet relations. In Christiane Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*, pp. 131–151. Cambridge, Mass.: MIT Press.
- Hearst, Marti A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 15th International Conference on Computational Linguistics*, Nantes, France, 23–28 August 1992, pp. 539–545.
- Huang, Anna, David Milne, Eibe Frank & Ian H. Witten (2009). Clustering documents with active learning using Wikipedia. In *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Bangkok, Thailand, pp. 628–636.
- Kasneci, Gjergji, Maya Ramnath, Fabian M. Suchanek & Gerhard Weikum (2008). The YAGO-NAGA approach to knowledge discovery. *SIGMOD Record*, 37(4):41–47.
- Krötzsch, Markus, Denny Vrandečić & Max Völkel (2005). Wikipedia and the semantic web — the missing links. In *Proceedings of The First International Wikimedia Conference (Wikimania 2005)*, Frankfurt am Main, Germany.
- Kulkarni, Sayali, Amit Singh, Ganesh Ramakrishnan & Soumen Chakrabarti (2009). Collective annotation of Wikipedia entities in web text. In *Proceedings of the 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Paris, France, 28 June – 1 July 2009, pp. 457–466.
- Lenat, Douglas B. & R. V. Guha (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, Mass.: Addison-Wesley.
- Lenat, Douglas B., R.V. Guha, Karen Pittman, Dexter Pratt & Mary Shepherd (1990). Cyc: Towards programs with common sense. *Communications of the ACM*, 33(8):30–49.
- Lucassen, Teun & Jan Maarten Schraagen (2010). Trust in Wikipedia: How users trust information from an unknown source. In *Proceedings of the 4th WWW Workshop on Information Credibility on the Web in conjunction with the 19th World Wide Web Conference*, Raleigh, N.C., 27 April 2010.
- Matuszek, Cynthia, Michael Witbrock, Robert C. Kahlert, John Cabral, Dave Schneider, Purvesh Shah & Doug Lenat (2005). Searching for common sense: Populating Cyc from the web. In *Proceedings of the 20th National Conference on Artificial Intelligence*, Pittsburgh, Penn., 9–13 July 2005, pp. 1430–1435.
- Max, Aureliën & Guillaume Wisniewski (2010). Mining naturally-occurring corrections and

- paraphrases from Wikipedia’s revision history. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010.
- Medelyan, Olena, David Milne, Catherine Legg & Ian H. Witten (2009). Mining meaning from Wikipedia. *International Journal of Human-Computer Interaction*, 67(9):716–754.
- Medelyan, Olena, Ian H. Witten & David Milne (2008). Topic indexing with Wikipedia. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 19–24.
- Mihalcea, Rada & Andras Csomai (2007). Linking documents to encyclopedic knowledge. In *Proceedings of the ACM 16th Conference on Information and Knowledge Management (CIKM 2007)*, Lisbon, Portugal, 6–9 November 2007, pp. 233–242.
- Milne, David & Ian H. Witten (2008a). An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 25–30.
- Milne, David & Ian H. Witten (2008b). Learning to link with Wikipedia. In *Proceedings of the ACM 17th Conference on Information and Knowledge Management (CIKM 2008)*, Napa Valley, Cal., USA, 26–30 October 2008, pp. 1046–1055.
- Nastase, Vivi & Michael Strube (2008). Decoding Wikipedia category names for knowledge acquisition. In *Proceedings of the 23rd Conference on the Advancement of Artificial Intelligence*, Chicago, Ill., 13–17 July 2008, pp. 1219–1224.
- Nastase, Vivi, Michael Strube, Benjamin Börschinger, Căcilia Zirn & Anas Elghafari (2010). WikiNet: A very large scale multi-lingual concept network. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010.
- Navigli, Roberto & Simone Paolo Ponzetto (2010). BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 11–16 July 2010, pp. 216–225.
- Nelken, Rani & Elif Yamangil (2008). Mining Wikipedia’s article revision history for training computational linguistics algorithms. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 31–36.
- Nguyen, Dat P.T., Yutaka Matsuo & Mitsuru Ishizuka (2007). Relation extraction from Wikipedia using subtree mining. In *Proceedings of the 22nd Conference on the Advancement of Artificial Intelligence*, Vancouver, B.C., Canada, 22–26 July 2007, pp. 1414–1420.
- Nguyen, Truc-Vien T. & Alessandro Moschitti (2011). End-to-end relation extraction using distant supervision from external semantic repositories. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, Oreg., USA, 19–24 June 2011.
- Niederer, Sabine & José van Dijk (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. *New Media & Society*, 12(8):1368–1387.
- Pantel, Patrick & Marco Pennacchiotti (2006). Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 17–21 July 2006, pp. 113–120.
- Pantel, Patrick & Deepak Ravichandran (2004). Automatically labeling semantic classes. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, Boston, Mass., 2–7 May 2004, pp. 321–328.
- Ponzetto, Simone Paolo & Roberto Navigli (2010). Knowledge-rich word sense disambiguation rivaling supervised systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 11–16 July 2010, pp. 1522–1531.
- Ponzetto, Simone Paolo & Michael Strube (2007a). Deriving a large scale taxonomy from Wikipedia. In *Proceedings of the 22nd Conference on the Advancement of Artificial Intelligence*, Vancouver, B.C., Canada, 22–26 July 2007, pp. 1440–1445.
- Ponzetto, Simone Paolo & Michael Strube (2007b). Knowledge derived from Wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research*, 30:181–212.
- Ponzetto, Simone Paolo & Michael Strube (2011). Taxonomy induction based on a collabora-

- tively built knowledge repository. *Artificial Intelligence*, 175(9/10):1737–1756.
- Riloff, Ellen & Rosie Jones (1999). Learning dictionaries for information extraction by multi-level bootstrapping. In *Proceedings of the 16th National Conference on Artificial Intelligence*, Orlando, Flo., 18–22 July 1999, pp. 474–479.
- Snow, Rion, Daniel Jurafsky & Andrew Y. Ng (2006). Semantic taxonomy induction from heterogeneous evidence. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia, 17–21 July 2006, pp. 801–808.
- Spinellis, Diomidis & Panagiotis Louridas (2008). The collaborative organization of knowledge. *Communications of the ACM*, 51(8):68–73.
- Strube, Michael & Simone Paolo Ponzetto (2006). WikiRelate! Computing semantic relatedness using Wikipedia. In *Proceedings of the 21st National Conference on Artificial Intelligence*, Boston, Mass., 16–20 July 2006, pp. 1419–1424.
- Stvilia, Beseki, Michael B. Twidale, Linda C. Smith & Les Gasser (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001.
- Suchanek, Fabian, Gjergji Kasneci & Gerhard Weikum (2008). YAGO: A large ontology from Wikipedia and WordNet. *Elsevier Journal of Web Semantics*, 6(3):203–217.
- Suchanek, Fabian M., Gjergji Kasneci & Gerhard Weikum (2007). YAGO: A core of semantic knowledge. unifying WordNet and Wikipedia. In *Proceedings of the 16th World Wide Web Conference*, Banff, Canada, 8–12 May, 2007, pp. 697–706.
- Turdakov, Denis Y. & S. Dmitry Lizorkin (2009). HMM expanded to multiple interleaved chains as a model for word sense disambiguation. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation*, Hong Kong, China, 3–5 December 2009.
- Turdakov, Denis Y. & S. Dmitry Lizorkin (2010). Automatic word sense disambiguation based on document networks. *Programming and Computer Software*, 36(1):11–18.
- Weikum, Gerhard, Gjergji Kasneci, Maya Ramanath & Fabian Suchanek (2009). Database and information-retrieval methods for knowledge discovery. *Communications of the ACM*, 52(4):56–64.
- Weld, Daniel S., Raphael Hoffmann & Fei Wu (2008a). Using Wikipedia to bootstrap open information extraction. *SIGMOD Record*, 37(4):62–68.
- Weld, Daniel S., Fei Wu, Eytan Adar, Saleema Amershi, James Fogarty, Raphael Hoffmann, Kayur Patel & Michael Skinner (2008b). Intelligence in Wikipedia. In *Proceedings of the 23rd Conference on the Advancement of Artificial Intelligence*, Chicago, Ill., 13–17 July 2008, pp. 1609–1615.
- Wentland, Wolodja, Johannes Knopp, Carina Silberer & Matthias Hartung (2008). Building a multilingual lexical resource for named entity disambiguation, translation and transliteration. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, 26 May – 1 June 2008.
- Wöhner, Thomas & Ralf Peters (2009). Assessing the quality of Wikipedia articles with life-cycle based metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, Orlando, Fl., 25–27 October 2009. 10 pages.
- Wu, Fei & Daniel Weld (2007). Automatically semantifying Wikipedia. In *Proceedings of the ACM 16th Conference on Information and Knowledge Management (CIKM 2007)*, Lisbon, Portugal, 6–9 November 2007, pp. 41–50.
- Wu, Fei & Daniel Weld (2008). Automatically refining the Wikipedia infobox ontology. In *Proceedings of the 17th World Wide Web Conference*, Beijing, China, 21–25 April, 2008.
- Wu, Fei & Daniel Weld (2010). Open information extraction using Wikipedia. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, 11–16 July 2010, pp. 118–127.
- Yamangil, Elif & Rani Nelken (2008). Mining Wikipedia revision histories for improving sentence compression. In *Companion Volume to the Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, 15–20 June 2008, pp. 137–140.

- Yasud, Keiji & Eiichiro Sumita (2008). Method for building sentence-aligned corpus from Wikipedia. In *Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy at AAAI-08*, Chicago, Ill., 13 July 2008, pp. 64–66.
- Yuncong, Chen & Pascale Fung (2010). Unsupervised synthesis of multilingual Wikipedia articles. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, 23–27 August 2010, pp. 197–205.
- Zhou, Yiping (2010). Resolving surface forms to Wikipedia topics. In *Proceedings of the 23rd International Conference on Computational Linguistics*, Beijing, China, 23–27 August 2010, pp. 1335–1343.