

Entity Linking

Michael Strube

michael.strube at h-its.org

February 2, 2015

Inhalt

Als *Entity Linking* wird das Disambiguieren von Eigennamen durch deren Verknüpfung mit einer Wissensbasis bezeichnet. Die Wissensbasis wird dabei häufig aus Wikipedia abgeleitet. Konzeptdisambiguierung geht über *Entity Linking* hinaus, indem nicht nur Eigennamen sondern alle Arten von Nomen disambiguiert werden. Beide Aufgaben haben ihren Ursprung in der Lesartendisambiguierung und im Erkennen von Eigennamen. Das Seminar beginnt mit einem Überblick über diese beiden traditionellen Aufgaben, bevor der Stand der Forschung in den Bereichen *Entity Linking* und Konzeptdisambiguierung diskutiert wird. Dabei wird besonderes Augenmerk auf die *Entity Linking*-Wettbewerbe im Rahmen der *Text Analysis Conference* und verwandte Wettbewerbe gelegt (im Rahmen von WWW, SIGIR, NTCIR, etc.), die auch *cross*-linguale Aufgaben behandeln. Anwendungen wie durch *Entity Linking* unterstütztes Information Retrieval, Konzeptdisambiguierung für Diskursverarbeitungsaufgaben, *Entity Linking* in *Microblogs*, etc. werden ebenfalls angesprochen. – Neben der Darstellung wissenschaftlicher Inhalte soll auch diskutiert werden, welche (positiven oder möglicherweise auch negativen) Auswirkungen die angesprochenen Methoden und Anwendungen auf die Gesellschaft haben.

Termine, Themenvorschläge

23.10.2014

Einführung, Terminologie, ...

30.10.2014

fällt aus: EMNLP 2014

06.11.2014

Word Sense Disambiguation

(Schütze, 1992; Yarowsky, 1995; Lin, 1997; Schütze, 1998; Navigli & Lapata, 2007; Navigli, 2009; Yarowsky, 2010; Navigli & Lapata, 2010)

13.11.2014

Named Entity Recognition and Disambiguation

(Bikel et al., 1997; Sekine, 1998; Riloff & Jones, 1999; Nadeau & Sekine, 2007)

at: CoNLL 2002, CoNLL 2003 (Florian et al., 2003; Klein & Manning, 2003; Tjong Kim Sang, 2002; Tjong Kim Sang & De Meulder, 2003; Zhang & Johnson, 2003)

zur Vorbereitung: (Tjong Kim Sang & De Meulder, 2003) und ((Florian et al., 2003) oder (Klein & Manning, 2003) oder (Zhang & Johnson, 2003))

20.11.2014

Frühe Arbeiten mit Wikipedia I:

Referat Julia Kreutzer: (Bunescu & Paşca, 2006; Mihalcea & Csomai, 2007; Csomai & Mihalcea, 2008)

zur Vorbereitung: (Bunescu & Paşca, 2006) oder (Mihalcea & Csomai, 2007) oder (Csomai & Mihalcea, 2008)

optional: (Cucerzan, 2007)

27.11.2014

Frühe Arbeiten mit Wikipedia II:

Referat: Erwin Glockner – WikipediaMiner: (Milne & Witten, 2008a; 2008b; 2013);

zur Vorbereitung: (Milne & Witten, 2008b), **auch:** testen Sie den Wikipedia Miner unter <http://wikipedia-miner.cms.waikato.ac.nz/>, was funktioniert, was nicht? (na ja, bei mir im Moment, 21.11.14, 17:45 Uhr, der angegebene Link nicht, aber das legt sich ja vielleicht noch ...)

optional: (Kulkarni et al., 2009)

04.12.2014

Yago, DBpedia, Freebase:

Referat: Devon Fritz – Freebase: (Bollacker et al., 2008; Lin & Etzioni, 2012)

Referat: Jani Takhsha – DBpedia (Mendes et al., 2011; 2012) und (Lehmann et al., 2014) (als Intro zu DBpedia)

optional: Yago/AIDA: (Hoffart et al., 2011; 2014)

zur Vorbereitung: (Lehmann et al., 2014) (Abschnitte 1-3 und 5, den Rest flüchtig) oder ((Bollacker et al., 2008) und Lektüre von <https://www.freebase.com/>)

11.12.2014

Fortsetzung Referat: Jani Takhsha – DBpedia (Mendes et al., 2011; 2012) und (Lehmann et al., 2014) (als Intro zu DBpedia)

Monolingual Entity Linking:

Referat: Sabrina Mänz – at TAC (<http://www.nist.gov/tac/>): (Ji & Grishman, 2011; Ji et al., 2011)

optional: (Ratinov et al., 2011; Ratinov & Roth, 2011; Chang et al., 2013; Cheng & Roth, 2013)

zur Vorbereitung: (Li et al., 2011) (http://www.nist.gov/tac/publications/2011/additional.papers/KBP2011_annotation_overview_proceedings.pdf) und *Overview of the TAC2011 Knowledge Base Population Track* (http://www.nist.gov/tac/publications/2011/presentations/KBP2011_overview_presentation.pdf)

18.12.2014

Monolingual Entity Linking:

Referat: Attila Azgin – at TAC (<http://www.nist.gov/tac/>): (Cucerzan, 2012; 2013)

Cross-lingual Entity Linking:

Referat: Mengfei Zhou – at TAC (<http://www.nist.gov/tac/>): (Miao et al., 2014) und (Fahrni et al., 2014; Merhav et al., 2014; Wang et al., 2014)

Cross-lingual Link Discovery:

at NTCIR (<http://research.nii.ac.jp/ntcir/ntcir-9/>, <http://research.nii.ac.jp/ntcir/ntcir-10/>): (Fahrni et al., 2011; Kim & Gurevych, 2011; Knoth et al., 2011; Tang et al., 2011b; 2011a; 2014)

zur Vorbereitung: (Cucerzan, 2007) oder (http://www.nist.gov/tac/publications/2013/presentations/KBP2013_EL_overview.presentation.pdf und http://www.nist.gov/tac/publications/2012/presentations/KBP2012_Entity_Linking_tasks_overview.pdf)

08.01.2015

Cross-lingual Link Discovery:

Michael Strube – at NTCIR (<http://research.nii.ac.jp/ntcir/ntcir-9/>, <http://research.nii.ac.jp/ntcir/ntcir-10/>): (Fahrni et al., 2011; Kim & Gurevych, 2011; Knoth et al., 2011; Tang et al., 2011b; 2011a; 2014)

Entity Linking in Microblogs:

Referat: Zoia Bylinovich – (Fang & Chang, 2014); at NEEL (<http://www.scc.lancs.ac.uk/microposts2014/challenge/>): (Chang et al., 2014)

zur Vorbereitung: <http://www.scc.lancs.ac.uk/microposts2014/challenge/> und ((Habib et al., 2014) oder (Yosef et al., 2014))

optional:

(Liu et al., 2011; Abel et al., 2011; Meij et al., 2012; Liu et al., 2012; Guo et al., 2013)

15.01.2015

fällt aus: Vortrag Trier

22.01.2015

Entity Linking in Microblogs:

Referat: Katharina Sowa – (Abel et al., 2011) – verschoben auf den 05.02.13

Michael Strube – (Liu et al., 2013; Guo et al., 2013)

zur Vorbereitung: (Guo et al., 2013)

optional: (Liu et al., 2011; Meij et al., 2012; Liu et al., 2012; Fang & Chang, 2014)

Michael Strube – Evaluierung, B³ und B³+

29.01.2015

Konzeptdisambiguierung:

Referat: Mirjam Eppinger – (Cheng & Roth, 2013)

optional: (Bentivogli et al., 2010; Ratinov et al., 2011; Fahrni & Strube, 2012; 2014)

Konzeptdisambiguierung und Diskursverarbeitung:

Referat: Thomas Haider – (Durrett & Klein, 2014)

optional: (Bryl et al., 2010; Tonelli et al., 2013; Fahrni & Strube, 2014)

zur Vorbereitung: (Ratinov et al., 2011) oder (Ratinov & Roth, 2012)

05.02.2015

Entity Linking in Microblogs:

Referat: Katharina Sowa – (Abel et al., 2011)

Zusammenfassung, Diskussion

zur Vorbereitung: Fragen zum ganzen Seminar: Was blieb offen? Weiterführendes? – Außerdem: Was funktioniert? Ist Entity Linking schon als Vorverarbeitung einsetzbar? Sind manche Ansätze schon zu kompliziert? Für welchen Zweck brauchen wir welche Methode, welches Wissen? Empfehlungen?

Weitere Themenvorschläge:

Entity Recognition and Disambiguation Challenge at SIGIR 2014: (<http://web-ngram.research.microsoft.com/ERD2014/>) (Cucerzan, 2014; Lipczak et al., 2014; Procházka et al., 2014)

Evaluierung: (Hachey et al., 2013; Tang et al., 2014)

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). Ich schätze Wikipedia als Gegenstand meiner Forschung sehr, nicht aber als Quelle für wissenschaftliche Arbeiten. Hausarbeiten, die Wikipedia (oder auch andere allgemeine Enzyklopädien) als Beleg zitieren, werde ich zurückweisen. Bitte lesen und zitieren Sie Fachliteratur!

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

Abgabetermin: bis spätestens 8. März 2015; per Email als PDF-Datei (kein Microsoft Word!) oder ausgedruckt per Post – Matrikelnummer und Studiengang nicht vergessen!

References

- Abel, Florian, Qi Gao, Geert-Jan Houben & Ke Tao (2011). Semantic enrichment of twitter posts for user profile construction on the social web. In *Proceedings of the 8th Extended Semantic Web Conference*, Heraklion, Crete, Greece, 29 May – 2 June 2011, pp. 375–389.
- 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.
- Bikel, Daniel M., Scott Miller, Richard Schwartz & Ralph Weischedel (1997). Nymble: A high-performance learning name-finder. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington, D.C., 31 March – 3 April 1997, pp. 194–201.
- Bollacker, Kurt, Colin Evans, Praveen Paritosh, Tim Sturge & Jamie Taylor (2008). Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, Vancouver, B.C., Canada, 10–12 June 2008, pp. 1247–1250.
- Bryl, Volha, Claudio Giuliano, Luciano Serafini & Kateryna Tymoshenko (2010). Supporting natural language processing with background knowledge: Coreference resolution case. In *Proceedings of the 9th International Semantic Web Conference, Revised Selected Papers, Part I*, Shanghai, China, 7-11 November 2010, pp. 80–95.
- 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.
- Chang, Kai-Wei, Rajhans Samdani & Dan Roth (2013). A constrained latent variable model for coreference resolution. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 601–612.
- Chang, Ming-Wei, Bo-June Hsu, Hao Ma, Ricky Loynd & Kuansan Wang (2014). E2E: An end-to-end entity linking system for short and noisy text. In *Proceedings of the 4th Workshop on Making Sense of Microposts*, Seoul, Korea, 7 April 2014.
- Cheng, Xiao & Dan Roth (2013). Relational inference for Wikification. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 1787–1796.
- 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.
- Cucerzan, Silviu (2012). TAC entity linking by performing full-document entity extraction and disambiguation. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.
- Cucerzan, Silviu (2013). The MSR system for entity linking at TAC 2012. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 5–6 November 2012.
- Cucerzan, Silviu (2014). Name entities made obvious. In *Proceedings of the ERD’14 Workshop on the Entity Recognition and Disambiguation Challenge*, Gold Coast, Queensland, Australia, 11 July 2014.
- Durrett, Greg & Dan Klein (2014). A joint model for entity analysis: Coreference, typing, and linking. *Transactions of the Association of Computational Linguistics*, 2:477–490.
- Fahrni, Angela, Benjamin Heinzerling, Thierry Göckel & Michael Strube (2014). HITS’ monolingual and cross-lingual entity linking system at TAC 2013. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 18–19 November 2013.
- Fahrni, Angela, Vivi Nastase & Michael Strube (2011). HITS’ graph-based system at the

- NTCIR-9 cross-lingual link discovery task. In *Proceedings of the 9th NTCIR Workshop Meeting*, Tokyo, Japan, 6-9 December 2011, pp. 473–480.
- Fahrni, Angela & Michael Strube (2012). Jointly disambiguating and clustering concepts and entities with Markov logic. In *Proceedings of the 24th International Conference on Computational Linguistics*, Mumbai, India, 8–15 December 2012, pp. 815–832.
- Fahrni, Angela & Michael Strube (2014). A latent variable model for discourse-aware concept and entity disambiguation. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden, 26–30 April 2014, pp. 491–500.
- Fang, Yuan & Ming-Wei Chang (2014). Entity linking in microblogs with spatial and temporal signals. *Transactions of the Association for Computational Linguistics*, 2:259–272.
- Florian, Radu, Abe Ittycheriah, Hongyan Jing & Tong Zhang (2003). Named entity recognition through classifier combination. In *Proceedings of the 7th Conference on Computational Natural Language Learning*, Edmonton, Alberta, Canada, 31 May – 1 June 2003, pp. 168–171.
- Guo, Yuhang, Bing Qin, Ting Liu & Sheng Li (2013). Microblog entity linking by leveraging extra posts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, Wash., 18–21 October 2013, pp. 863–868.
- Habib, Mena B., Maurice van Keule & Zemin Zhu (2014). Named entity extraction and linking challenge: University of Twente at #microposts2014. In *Proceedings of the 4th Workshop on Making Sense of Microposts*, Seoul, Korea, 7 April 2014.
- Hachey, Ben, Will Radford, Joel Nothman, Matthew Honnibal & James R. Curran (2013). Evaluating entity linking with Wikipedia. *Artificial Intelligence*, 194:130–150.
- Hoffart, Johannes, Yasemin Altun & Gerhard Weikum (2014). Discovering emerging entities with ambiguous names. In *Proceedings of the 23rd World Wide Web Conference*, Seoul, Korea, 7–11 April, 2014, pp. 385–395.
- Hoffart, Johannes, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater & Gerhard Weikum (2011). Robust disambiguation of named entities in text. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, U.K., 27–29 July 2011, pp. 782–792.
- Ji, Heng & Ralph Grishman (2011). Knowledge base population: Successful approaches and challenges. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Portland, Oreg., 19–24 June 2011, pp. 1148–1158.
- Ji, Heng, Ralph Grishman & Hoa Dang (2011). Overview of the TAC 2011 knowledge base population track. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.
- Kim, Juni & Iryna Gurevych (2011). UKP at CrossLink: Anchor text translation for cross-lingual link discovery. In *Proceedings of the 9th NTCIR Workshop Meeting*, Tokyo, Japan, 6-9 December 2011, pp. 487–494.
- Klein, Dan & Christopher D. Manning (2003). Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pp. 423–430.
- Knoth, Petr, Lukas Zilka & Zdenek Zdrahal (2011). KMI, The Open University at NTCIR-9 CrossLink: Cross-lingual link discovery in Wikipedia using Explicit Semantic Analysis. In *Proceedings of the 9th NTCIR Workshop Meeting*, Tokyo, Japan, 6-9 December 2011.
- 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.
- Lehmann, Jens, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo M. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer & Christian Bizer (2014). DBpedia – A large scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*. To appear.
- Li, Xuansong, Joe Ellis, Kira Griffit, Stephanie M. Strassel, Robert Parker & Jonathan Wright

- (2011). Linguistic resources for 2011 knowledge base population evaluation. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.
- Lin, Dekang (1997). Using syntactic dependency as local context to resolve word sense ambiguity. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and of the 8th Conference of the European Chapter of the Association for Computational Linguistics*, Madrid, Spain, 7–12 July 1997, pp. 64–71.
- Lin, Thomas & Oren Etzioni (2012). Entity linking at web scale. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction (AKBC-WEKEX)*, Montréal, Québec, Canada, 7–8 June 2012, pp. 84–88.
- Lipczak, Marek, Arash Koushkestani & Evangelos Milios (2014). Tulip: Lightweight entity recognition and disambiguation using Wikipedia-based topic centroids. In *Proceedings of the ERD'14 Workshop on the Entity Recognition and Disambiguation Challenge*, Gold Coast, Queensland, Australia, 11 July 2014.
- Liu, Xiaohua, Yitong Li, Haocheng Wu, Ming Zhou, Furu Wei & Yi Lu (2013). Entity linking for tweets. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Sofia, Bulgaria, 4–9 August 2013, pp. 1304–1311.
- Liu, Xiaohua, Shaodian Zhang, Furu Wei & Ming Zhou (2011). Recognizing named entities in Tweets. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Portland, Oreg., 19–24 June 2011, pp. 359–367.
- Liu, Xiaohua, Ming Zhou, Xiangyang Zhou, Zhongyang Fu & Furu Wei (2012). Joint inference of named entity recognition and normalization for Tweets. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Jeju Island, Korea, 8–14 July 2012, pp. 526–535.
- Meij, Edgar, Wouter Weerkamp & Maarten de Rijke (2012). Adding semantics to microblog posts. In *Proceedings of the 5th International Conference on Web Search and Web Data Mining*, Seattle, Wash., 8–12 February 2012, pp. 563–572.
- Mendes, Pablo N., Joachim Daiber, Max Jakob & Christian Bizer (2012). Evaluating DBpedia spotlight for the TAC-KBP entity linking task. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.
- Mendes, Pablo N., Max Jakob, Andrés García-Silva & Christian Bizer (2011). DBpedia spotlight: Shedding light on the web of documents. In *I-Semantics '11: Proceedings of the 7th International Conference on Semantic Systems*, Graz, Austria, 7–9 September 2011, pp. 1–8.
- Merhav, Yuval, Joel Barry, James Clarke & David Murgatroyd (2014). Basis Technology at TAC 2013 entity linking. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 18–19 November 2013.
- Miao, Qingliang, Ruiyu Fang, Yao Meng & Shu Zhang (2014). FRDC's cross-lingual entity linking system at TAC 2013. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 18–19 November 2013.
- 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.
- Milne, David & Ian H. Witten (2013). An open-source toolkit for mining Wikipedia. *Artificial Intelligence*, 194:222–239.
- Nadeau, David & Satoshi Sekine (2007). A survey of named entity recognition and classification. *Linguisticae Investigationes*, 30(1):3–26.
- Navigli, Roberto (2009). Word sense disambiguation: A survey. *ACM Computing Surveys*,

41(2):1–69.

- Navigli, Roberto & Mirella Lapata (2007). Graph connectivity measures for unsupervised word sense disambiguation. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, Hyderabad, India, 6–12 January 2007, pp. 1683–1688.
- Navigli, Roberto & Mirella Lapata (2010). An experimental study of graph connectivity for unsupervised word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):678–692.
- Procházka, Jan, Alan Eckhardt, Juraj Hreško & Otakar Smrž (2014). Disambiguation using entities occurrences for ERD challenge. In *Proceedings of the ERD'14 Workshop on the Entity Recognition and Disambiguation Challenge*, Gold Coast, Queensland, Australia, 11 July 2014.
- Ratinov, Lev & Dan Roth (2011). GLOW TAC-KBP2011 entity linking system. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 14–15 November 2011.
- Ratinov, Lev & Dan Roth (2012). Learning-based multi-sieve co-reference resolution with knowledge. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*, Jeju Island, Korea, 12–14 July 2012, pp. 1234–1244.
- Ratinov, Lev, Dan Roth, Doug Downey & Mike Anderson (2011). Local and global algorithms for disambiguation to Wikipedia. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Portland, Oreg., 19–24 June 2011, pp. 1375–1384.
- 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.
- Schütze, Hinrich (1992). Dimensions of meaning. In *Proceedings of the ACM/IEEE Conference on Supercomputing*, pp. 787–796.
- Schütze, Hinrich (1998). Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- Sekine, Satoshi (1998). NYU: Description of the Japanese NE system used for MET-2. In *Proceedings of the Seventh Message Understanding Conference*, Fairfax, Virginia, 29 April - 1 May, 1998.
- Tang, Ling-Xiang, , Daniel Cavanagh, Andrew Trotman, Shlomo Geva, Yue Xu & Laurianne Sitbon (2011a). Automated cross-lingual link discovery in Wikipedia. In *Proceedings of the 9th NTCIR Workshop Meeting*, Tokyo, Japan, 6-9 December 2011, pp. 512–519.
- Tang, Ling-Xiang, Shlomo Geva, Andrew Trotman, Yue Xu & Kelly Itakura (2011b). Overview of the NTCIR-9 crosslink task: Cross-lingual link discovery. In *Proceedings of the 9th NTCIR Workshop Meeting*, Tokyo, Japan, 6-9 December 2011, pp. 437–463.
- Tang, Ling-Xiang, Shlomo Geva, Andrew Trotman, Yue Xu & Kelly Y. Itakura (2014). An evaluation framework for cross-lingual link discovery. *Information Processing & Management*, 50(1):1–23.
- Tjong Kim Sang, Erik F. (2002). Introduction to the CoNLL-2002 Shared Task: Language-independent Named Entity Recognition. In *Proceedings of the 6th Conference on Computational Natural Language Learning*, Taipei, Taiwan, 31 August – 1 September 2002, p. ??
- Tjong Kim Sang, Erik F. & Fien De Meulder (2003). Introduction to the CoNLL-2003 Shared Task: Language-independent Named Entity Recognition. In *Proceedings of the 7th Conference on Computational Natural Language Learning*, Edmonton, Alberta, Canada, 31 May – 1 June 2003, pp. 142–147.
- Tonelli, Sara, Claudio Giuliano & Kateryna Tymoshenko (2013). Wikipedia-based WSD for multilingual frame annotation. *Artificial Intelligence*, 194:203–221.
- Wang, Yan, Yankai Lin, Zhiyuan Liu & Maosong Sun (2014). THUNLP at TAC KBP 2013 in entity linking. In *Proceedings of the Text Analysis Conference*, National Institute of Standards and Technology, Gaithersburg, Maryland, USA, 18–19 November 2013.
- Yarowsky, David (1995). Unsupervised word sense disambiguation rivalling supervised meth-

- ods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, Cambridge, Mass., 26–30 June 1995, pp. 189–196.
- Yarowsky, David (2010). Word sense disambiguation. In Nitin Indurkha & Fred J. Damerau (Eds.), *Handbook of Natural Language Processing*, pp. 315–338. Boca Raton, Flo.: CRC Press.
- Yosef, Mohamed Amir, Johannes Hoffart, Yusra Ibrahim, Artem Boldyrev & Gerhard Weikum (2014). Adapting AIDA for Tweets. In *Proceedings of the 4th Workshop on Making Sense of Microposts*, Seoul, Korea, 7 April 2014.
- Zhang, Tong & David Johnson (2003). A robust risk minimization based named entity recognition system. In *Proceedings of the 7th Conference on Computational Natural Language Learning*, Edmonton, Alberta, Canada, 31 May – 1 June 2003, pp. 204–207.