

Wikification and Beyond: The Challenges of Entity and Concept Grounding

Dan Roth, Heng Ji, Ming-Wei Chang, Taylor Cassidy

Contextual disambiguation and grounding of concepts and entities in natural language text are essential to moving forward in many natural language understanding related tasks and are fundamental to many applications. The *Wikification* task (Bunescu and Pasca, 2006; Mihalcea and Csomai, 2007; Ratinov et al., 2011) aims at automatically identifying concept mentions appearing in a text document and link it to (or “ground it in”) a concept referent in a knowledge base (KB) (e.g., Wikipedia). For example, consider the sentence, “*The Times report on Blumental (D) has the potential to fundamentally reshape the contest in the Nutmeg State.*”, a Wikifier should identify the key entities and concepts (Times, Blumental, D and the Nutmeg State), and disambiguate them by mapping them to an encyclopedic resource revealing, for example, that “D” here represents the Democratic Party, and that “the Nutmeg State” refers Connecticut.

Wikification may benefit both human end-users and Natural Language Processing (NLP) systems. When a document is *Wikified* a reader can more easily comprehend it, as information about related topics and relevant enriched knowledge from a KB is readily accessible. From a system-to-system perspective, a Wikified document conveys the meanings of its key concepts and entities by grounding them in an encyclopedic resource or a structurally rich ontology. Indeed, there is evidence that Wikification output can improve broad NLP down-stream tasks, including coreference resolution (Ratinov and Roth, 2012), text classification (Gabrilovich and Markovitch, 2007; Chang et al., 2008; Vitale et al., 2012), and applications such as cultural heritage (Fernando and Stevenson, 2012), user interest discovery (Michelson and Macskassy, 2010; Xu et al., 2011) and user recommendation and searching (Weng et al., 2010).

This task has received increased attention over the past several years from the NLP and Data Mining communities, partly fostered by the U.S. NIST Text Analysis Conference Knowledge Base Population (KBP) track (Ji et al., 2010; Ji et al., 2011) and several versions of it has been studied. These include Wikifying all concept mentions in a single text document; Wikifying a cluster of co-referential named entity mentions that appear across multiple documents (Entity Linking), and Wikifying a whole document to a single concept (Genc et al., 2011). Other works relate this task to coreference resolution within and across documents and study it in the context of multiple text genres. A large number of papers on these topics have been published in major conferences in the last couple of years.

The primary goals of this tutorial are to review the framework of Wikification and motivate it as a broad paradigm for cross-source linking for knowledge enrichment. We will present and discuss multiple dimensions of the task definition, present the basic building blocks of a state-of-the-art Wikifier system, share some key lessons learned from the analysis of evaluation results, and discuss recently proposed ideas for advancing work in this area along with some of the key challenges. We will also suggest some research questions brought up by new applications, including interactive Wikification, social media, and censorship. The tutorial will thus be useful for both senior and junior researchers with interests in cross-source information extraction and linking, knowledge acquisition, and the use of acquired knowledge in natural language processing and information extraction. We will try to provide a concise roadmap of recent perspectives and results, as well as point to some of our Wikification resources that are available to the research communities.

Brief Tutorial Outline

After shortly motivating and introducing the general task the first part of the tutorial will be a methodological presentation of a skeletal Wikification system that will allow us to focus on some of the key challenges and computational directions. We will then describe in detail some of the obstacles including the scarcity of supervision signals, issues related to mention detection and candidate selection in different scenarios, and issues that arise when dealing with diverse text genres. Advanced methods that address these obstacles will be surveyed carefully. We will conclude with a discussion of some key remaining challenges and future work.

1. Motivation and Task Definition [30 minutes]

We will describe the notion of Wikification as a generic cross-source linking problem, motivate the task both human reader and system perspectives, and exemplify some applications. We will then lay out multiple dimensions of the task definition and illustrate how different settings are appropriate for different applications and what impact they might have on computational approaches.

2. A Skeletal View of Wikification Systems [45 minutes]

We will present the general architecture of a systematic approach to Wikification, and use it to survey existing approaches from a fairly unified perspective. In doing that we will address the key computational steps – mention identification, candidate identification and decision making and, within it, key issues such as knowledge representation, local and global context analysis, relevant statistical features, the role of machine learning and the utilization of unlabeled data.

[Coffee break: 30 minutes]

3. Key Challenges and Recent Advances [35 minutes]

We will address some of the key challenges facing high-performing end-to-end Wikification approaches once the basic algorithmic solutions are in place. In doing that we will touch upon all stages of the pipeline: mention identification, candidate generation, ranking of candidates and the identification of concepts and entities that are outside the knowledge base. We will discuss solutions that advance joint modeling of some of these computational steps, joint inference of Wikification with an application (e.g., coreference) or an additional process (e.g., relation identification) and new training models, and exhibit their impact on various steps in the Wikification pipeline.

4. New Tasks, Trends and Applications [30 minutes]

In this part of the tutorial we will address some of the new challenges that arise from extending the Wikification task to new settings – social media, cross-lingual Wikification, censored data, etc.. We will present some of the solutions that have started to emerge in this area (e.g., to deal with short and noisy text), along with some recent and interesting applications.

5. What's Next? [10 minutes]

We will conclude with a discussion of some of the open issues in this domain. These include the challenge of dealing with multiple knowledge bases of different levels of quality, difficulties that arise when interacting with users at multiple levels of expertise and those that result from using cross-genre data. We will also provide pointers to resources, including data sets, software and on-line demos.

Tutorial Instructors

Name: Dan Roth

Affiliation: University of Illinois at Urbana-Champaign

Email address: danr@illinois.edu

Website: <http://L2R.cs.uiuc.edu/>

Research Interests and Area of Expertise:

Dan Roth is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign and the Beckman Institute of Advanced Science and Technology (UIUC) and a University of Illinois Scholar. He is a fellow of AAAI, ACL and the ACM. Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning and received several paper, teaching and research awards. He has developed several machine learning based natural language processing systems that are widely used in the computational linguistics community and in industry and has presented invited talks and tutorials in several major conferences. Over the last few years he has worked on Entity Linking and Wikification. He has given several tutorials at ACL/NAACL/ECL and other forums.

Name: Heng Ji

Affiliation: Rensselaer Polytechnic Institute

Email address: jih@rpi.edu

Website: <http://nlp.cs.rpi.edu/hengji.html>

Research Interests and Area of Expertise:

Heng Ji is Edward G. Hamilton Development Chair Associate Professor in Computer Science Department of Rensselaer Polytechnic Institute. Her research interests focus on Natural Language Processing, especially on Cross-source Information Extraction and Knowledge Base Population. She coordinated the NIST TAC Knowledge Base Population task in 2010, 2011 and 2014 and has published several papers on entity linking and Wikification.

Name: Ming-Wei Chang

Affiliation: Microsoft Research

Email address: minchang@microsoft.com

Website: <http://research.microsoft.com/en-us/um/people/minchang/>

Research Interests and Area of Expertise:

Ming-Wei Chang is a researcher at Microsoft Research. His research interests are in machine learning and natural language understanding. He currently focuses on using large-scale structured and unstructured data for semantic understanding. Specially, he is interested in developing algorithms for entity linking that are effective for short and noisy text.

Name: Taylor Cassidy

Affiliation: U.S. Army Research Laboratory & IBM Research

Email address: taylor.cassidy.ctr@mail.mil

Website: <http://www.linkedin.com/pub/taylor-cassidy/10/b/160>

Research Interests and Area of Expertise:

Taylor Cassidy is a Postdoctoral Researcher at U.S. Army Research Laboratory & IBM Research. His research interests include Cross-lingual Entity Linking and Wikification for social media.

Expected attendance:

Based on the level of interest in this topic and the broad range of applications we expect around 40-50 participants.

Requirement:

Internet access.

References:

- [1] Ivo Anastacio, Bruno Martins and Pavel Calado. 2011. [Supervised Learning for Linking Named Entities to Knowledge Base Entries](#). Proc. Text Analysis Conference (TAC2011).
- [2] R. Bunescu and M. Pasca. 2006. [Using encyclopedic knowledge for named entity disambiguation](#). In Proceedings of the European Chapter of the ACL (EACL).
- [3] Praveen Bysani, Kranthi Reddy, Vijay Bharath Reddy, Sudheer Kovelamudi, Prasad Pingali and Vasudeva Varma. 2010. [IIIT Hyderabad in Guided Summarization and Knowledge Base Population](#). Proc. TAC 2010 Workshop.
- [4] Taylor Cassidy, Zheng Chen, Javier Artiles, Heng Ji, Hongbo Deng, Lev-Arie Ratinov, Jing Zheng, Jiawei Han and Dan Roth. 2011. [CUNY-UIUC-SRI TAC-KBP2011 Entity Linking System Description](#). Proc. Text Analytics Conference (TAC2011).
- [5] T. Cassidy, H. Ji, L. Ratinov, A. Zubiaga, and H. Huang. 2012. [Analysis and enhancement of Wikification for microblogs with context expansion](#). In Proceedings of COLING 2012.
- [6] D. Ceccarelli, C. Lucchese, S. Orlando, R. Perego, and S. Trani. 2013. [Learning relatedness measures for entity linking](#). Proc. CIKM '13: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management.
- [7] Ming-Wei Chang, Lev Ratinov, Dan Roth, and Vivek Srikumar. 2008. [Importance of semantic representation: dataless classification](#). In *Proceedings of the 23rd national conference on Artificial intelligence -Volume 2*, pages 830–835. AAAI Press.
- [8] Angel X. Chang, Valentin I. Spitzkovsky, Eric Yeh, Eneko Agirre and Christopher D. Manning. 2010. [Stanford-UBC Entity Linking at TAC-KBP](#). Proc. TAC2010.
- [9] Eric Charton, Marie-Jean Meurs, Ludovic Jean-Louis and Michel Gagnon. 2013. SemLinker System for KBP2013: A Disambiguation Algorithm based on Mutual Relations of Semantic Annotations inside a Document. Proc. Text Analysis Conference (TAC2013).

- [10] Z. Chen and H. Ji. 2011. [Collaborative ranking: A case study on entity linking](#). In Proc. EMNLP2011.
- [11] Zheng Chen, Suzanne Tamang, Adam Lee, Xiang Li, Wen-Pin Lin, Matthew Snover, Javier Artiles, Marissa Passantino and Heng Ji. 2010. [CUNY-BLENDER TAC-KBP2010 Entity Linking and Slot Filling System Description](#). *Proc. TAC 2010 Workshop*.
- [12] Yubo Chen, Guangyou Zhou, Liheng Xu, Shizhu He, Kang Liu and Jun Zhao. 2013. The CASIA Entity linking System at TAC 2013. Proc. Text Analysis Conference (TAC2013).
- [13] X. Cheng and D. Roth. 2013. [Relational inference for wikification](#). In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.
- [14] M. Cornolti, P. Ferragina, and M. Ciaramita. 2013. [A Framework for Benchmarking Entity-Annotation Systems](#). Proceedings of the 22nd international conference on World Wide Web, WWW 2013, Rio de Janeiro, Brazil, pp. 249–260.
- [15] Silviu Cucerzan. 2007. [Large-scale named entity disambiguation based on wikipedia data](#). In EMNLP-CoNLL2007.
- [16] Silviu Cucerzan. 2011. [Tac entity linking by performing full-document entity extraction and disambiguation](#). In Proc. TAC 2011 Workshop.
- [17] Jeffrey Dalton and Laura Dietz. 2013. [UMass CIIR at TAC KBP 2013 Entity Linking](#). Proc. Text Analysis Conference (TAC2013).
- [18] Jeffrey Dalton and Laura Dietz. 2013. [A Neighborhood Relevance Model for Entity Linking](#). Proceedings of the 10th International Conference in the RIAO series (OAIR), 2013.
- [19] A. Davis, A. Veloso, A. S. da Silva, W. Meira, Jr., and A. H. F. Laender. 2012. [Named entity disambiguation in streaming data](#). In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL), pages 815–824.
- [20] G. Demartini, D. E. Difallah, and P. Cudre-Mauroux. 2012. [Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking](#). In The International World Wide Web Conference , pages 469–478, New York, NY, USA. ACM.
- [21] Laura Dietz and Jeffrey Dalton. 2012. [Across-Document Neighborhood Expansion: UMass at TAC KBP 2012 Entity Linking](#). Proc. Text Analysis Conference (TAC2012).

- [22] M. Dredze, P. McNamee, D. Rao, A. Gerber, and T. Finin. 2010. [Entity disambiguation for knowledge base population](#). Proceedings of the 23rd International Conference on Computational Linguistics.
- [23] Angela Fahrni and Michael Strube. 2011. [HITS' Cross-lingual Entity Linking System at TAC2011: One Model for All Languages](#). Proc. TAC2011.
- [24] Angela Fahrni, Thierry Gockel and Michael Strube. 2012. [HITS' Monolingual and Cross-lingual Entity Linking System at TAC 2012: A Joint Approach](#). Proc. Text Analysis Conference (TAC2012).
- [25] Angela Fahrni, Benjamin Heinzerling, Thierry Gockel and Michael Strube. 2013. HITS' Monolingual and Cross-lingual Entity Linking System at TAC 2013. Proc. Text Analysis Conference (TAC2013).
- [26] N. Fernandez, J. A. Fisteus, L. Sanchez, and E. Martin. 2010. [Webtlab: A cooccurrence-based approach to kbp 2010 entity-linking task](#). In Proc. TAC 2010 Workshop.
- [27] Samuel Fernando and Mark Stevenson. [Adapting Wikification to Cultural Heritage](#) In proceedings of the EACL 2012 workshop on "Language Technology for Cultural Heritage, Social Sciences, and Humanities", April 24 2012, Avignon, France.
- [28] P. Ferragina and U. Scaiella. 2010. [Tagme: on-the-fly annotation of short text fragments \(by wikipedia entities\)](#). In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM 2010).
- [29] Evgeniy Gabrilovich and Shaul Markovitch. 2007. [Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge](#). Proc. AAI2007.
- [30] Sanyuan Gao, Yichao Cai, Si Li, Zongyu Zhang, Jingyi Guan, Yan Li, Hao Zhang, Weiran Xu and Jun Guo. 2010. [PRIS at TAC2010 KBP Track](#). Proc. TAC 2010 Workshop.
- [31] Y. Genc, Y. Sakamoto, and J. V. Nickerson. 2011. [Discovering context: classifying tweets through a semantic transform based on wikipedia](#). In *Proceedings of the 6th international conference on Foundations of augmented cognition: directing the future of adaptive systems*, FAC'11, pages 484–492.
- [32] Anna Lisa Gentile, Ziqi Zhang, Lei Xia, José Iria. 2009. [Graph-based semantic relatedness for named entity disambiguation](#). Proc. International Conference on Software, Services & Semantic Technologies 2009.

- [33] E. Gonzalez, H. Rodriguez, J. Turmo, P.R. Comas, A. Naderi, A. Ageno, E. Sapena, M. Vila and M.A. Marti. 2012. [The TALP participation at TAC-KBP 2012](#). Proc. Text Analysis Conference (TAC2012).
- [34] Swapna Gottipati and Jing Jiang. 2010. [SMU-SIS at TAC 2010 – KBP Track Entity Linking](#). *Proc. TAC 2010 Workshop*.
- [35] S. Gottipati and J. Jiang. 2011. [Linking entities to a knowledge base with query expansion](#). In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.
- [36] Y. Guo, W. Che, T. Liu, and S. Li. 2011. [A graph-based method for entity linking](#). In Proc. IJCNLP2011.
- [37] Zhaochen Guo, Ying Xu, Filipe Mesquita, Denilson Barbosa, Grzegorz Kondrak. 2012. [Ualberta at TAC-KBP 2012: English and Cross-Lingual Entity Linking](#). Proc. Text Analysis Conference (TAC2012).
- [38] Yuhang Guo, Bing Qin, Yuqin Li, Ting Liu, Sheng Li. 2013. [Improving Candidate Generation for Entity Linking](#). Natural Language Processing and Information Systems.
- [39] Sayali Kulkarni, Amit Singh, Ganesh Ramakrishnan, Soumen Chakrabarti. 2009. [Collective annotation of Wikipedia entities in web text](#). Proc. SIGKDD2009.
- [40] Ben Hachey, Will Radford, James R. Curran. 2011. [Graph-based named entity linking with wikipedia](#). Proc. WISE 2011.
- [41] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. Curran. 2013. [Evaluating entity linking with wikipedia](#). Artificial. Intelligence.
- [42] Sherzod Hakimov, Salih Atalay Oto, Erdogan Dogdu. 2012. [Named entity recognition and disambiguation using linked data and graph-based centrality scoring](#). Proc. the 4th International Workshop on Semantic Web Information Management SWIM12 (2012)
- [43] X. Han and L. Sun. 2011. [A generative entity-mention model for linking entities with knowledge base](#). In Proc. ACL2011.
- [44] X. Han and J. Zhao. 2009. [Named entity disambiguation by leveraging wikipedia semantic knowledge](#). In Proceedings of the 18th ACM conference on Information and knowledge management, CIKM 2009.

- [45] X. Han, L. Sun, and J. Zhao. 2011. [Collective entity linking in web text: A graph-based method](#). In Proc. SIGIR2011.
- [46] Xianpei Han, Le Sun. 2012. [An entity-topic model for entity linking](#). Proc. EMNLP-CoNLL 2012.
- [47] J. He, M. de Rijke, M. Sevenster, R. van Ommering, and Y. Qian. 2011. [Generating links to background knowledge: A case study using narrative radiology reports](#). In Proceedings of the 20th ACM international conference on Information and knowledge management. ACM.
- [48] Z. He, S. Liu, Y. Song, M. Li, M. Zhou, and H. Wang. 2013. [Efficient collective entity linking with stacking](#). In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing.
- [49] Zhengyan He, Shujie Liu, Mu Li, Ming Zhou, Houfeng Wang, Longkai Zhang. 2013. [Learning Entity Representation for Entity Disambiguation](#). Proc. ACL2013.
- [50] J. Hoffart, M. A. Yosef, I. Bordino, H. Fürstenau, M. Pinkal, M. Spaniol, B. Taneva, S. Thater, and G. Weikum. 2011. [Robust Disambiguation of Named Entities in Text](#). Proc. Conference on Empirical Methods in Natural Language Processing. Proc. EMNLP 2011, Edinburgh, Scotland, pp. 782–792.
- [51] H. Huang, Z. Wen, D. Yu, H. Ji, Y. Sun, J. Han, and H. Li. 2013. [Resolving entity morphs in censored data](#). In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).
- [52] H. Ji, R. Grishman, H.T. Dang, K. Griffitt, and J. Ellis. 2010. [Overview of the tac 2010 knowledge base population track](#). In Text Analysis Conference (TAC) 2010.
- [53] H. Ji, R. Grishman, and H.T. Dang. 2011. [Overview of the tac 2011 knowledge base population track](#). In Text Analysis Conference (TAC) 2011.
- [54] Saurabh S. Kataria, Krishnan S. Kumar, Rajeev R. Rastogi, Prithviraj Sen, Srinivasan H. Sengamedu. 2011. [Entity disambiguation with hierarchical topic models](#). Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 2011.
- [55] Zornitsa Kozareva and Sujith Ravi. 2011. [Unsupervised Name Ambiguity Resolution Using A Generative Model](#). Proc. EMNLP2011 Workshop on Unsupervised Learning in NLP.
- [56] Z. Kozareva, K. Voevodski, and S. Teng. 2011. [Class label enhancement via related instances](#). In Proc. EMNLP2011.

- [57] S. Kulkarni, A. Singh, G. Ramakrishnan, and S. Chakrabarti. 2009. [Collective annotation of wikipedia entities in web text](#). In KDD.
- [58] John Lehmann, Sean Monahan, Luke Nezda, Arnold Jung and Ying Shi. 2010. [LCC Approaches to Knowledge Base Population at TAC 2010](#). *Proc. TAC 2010 Workshop*.
- [59] Y. Li, C. Wang, F. Han, J. Han, D. Roth, and X. Yan. 2013. [Mining evidences for named entity disambiguation](#). In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13.
- [60] Thomas Lin, Mausam, Oren Etzioni. 2012. [No noun phrase left behind: detecting and typing unlinkable entities](#). *Proc. EMNLP-CoNLL 2012*.
- [61] X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu. 2013. [Entity linking for tweets](#). In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).
- [62] Paul McNamee, Hoa Trang Dang, Heather Simpson, Patrick Schone and Stephanie M. Strassel. 2010. [An Evaluation of Technologies for Knowledge Base Population](#). *Proc. LREC2010*.
- [63] P. McNamee and H.T. Dang. 2009. Overview of the tac 2009 knowledge base population track. *Proc. Text Analysis Conference (TAC) 2009*.
- [64] P. McNamee, J. Mayfield, D. Lawrie, D. W. Oard, and D. Doermann. 2011. [Cross-language entity linking](#). In *Proc. IJCNLP2011*.
- [65] Olena Medelyan, Catherine Legg, David Milne and Ian H. Witten. 2009. [Mining Meaning from Wikipedia](#). *International Journal of Human-Computer Studies archive*. Volume 67 , Issue 9.
- [66] E. Meij, W. Weerkamp, and M. de Rijke. 2012. [Adding semantics to microblog posts](#). In Proceedings of the fifth ACM international conference on Web search and data mining, WSDM '12.
- [67] P. N. Mendes, M. Jakob, A. García-Silva, and C. Bizer. 2011. [DBpedia spotlight: shedding light on the web of documents](#). Proceedings of the 7th International Conference on Semantic Systems.
- [68] Qingliang Miao, Ruiyu Fang, Yao Meng and Shu Zhang. 2013. FRDC's Cross-lingual Entity Linking System at TAC 2013. *Proc. TAC2013*.
- [69] M. Michelson and S. Macskassy. 2010. [Discovering users' topics of interest on twitter: A first look](#). In Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data, AND '10.

- [70] R. Mihalcea and A. Csomai. 2007. [Wikify!: linking documents to encyclopedic knowledge](#). In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, CIKM'07.
- [71] D. Milne and I.H. Witten. 2008. [Learning to link with wikipedia](#). In Proceeding of the 17th ACM conference on Information and knowledge management, pages 509–518. ACM.
- [72] Sean Monahan, John Lehmann, Timothy Nyberg, Jesse Plymale, and Arnold Jung. 2011. [Cross-lingual cross-document coreference with entity linking](#). In Proceedings of the Text Analysis Conference.
- [73] David Nemeskey, Gabor Recski, Attila Zseder and Andras Kornai. 2010. [BUDAPESTACAD at TAC 2010](#). *Proc. TAC 2010 Workshop*.
- [74] Hien T. Nguyen, Huy H. Minha,^b Tru H. Caob,^c Trong T. Nguyen^b. [JVN-TDT Entity Linking Systems at TAC-KBP2012](#). Proc. Text Analysis Conference (TAC2012).
- [75] Cesar de Pablo-Sanchez, Juan Perea and Paloma Martinez. 2010. [Combining Similarities with Regression based Classifiers for Entity Linking at TAC 2010](#). *Proc. TAC 2010 Workshop*.
- [76] M. Pennacchiotti and P. Pantel. 2009. [Entity extraction via ensemble semantics](#). In Proc. EMNLP2009.
- [77] Anja Pilz, Gerhard Paaß. 2011. [From names to entities using thematic context distance](#). Proc. CIKM 2011.
- [78] Glen Pink, Will Radford, Will Cannings, Andrew Naoum, Joel Nothman, Daniel Tse and James R. Curran. 2013. [SYDNEY CMCRC at TAC 2013](#). Proc. Text Analysis Conference (TAC2013).
- [79] W. Radford, B. Hachey, J. Nothman, M. Honnibal, and J. R. Curran. 2010. [Cmcrc at tac10: Documentlevel entity linking with graph-based re-ranking](#). In Proc. TAC 2010 Workshop.
- [80] L. Ratinov and D. Roth. 2012. [Learning-based multi-sieve co-reference resolution with knowledge](#). In Proc. EMNLP2012.
- [81] L. Ratinov, D. Roth, D. Downey, and M. Anderson. 2011. [Local and global algorithms for disambiguation to wikipedia](#). In Proc. of the Annual Meeting of the Association of Computational Linguistics (ACL).
- [82] Wei Shen, Jianyong Wang, Ping Luo, Min Wang. 2012. [LINDEN: linking named entities with knowledge base via semantic knowledge](#). Proc. WWW2012.

- [83] W. Shen, J. Wang, P. Luo, and M. Wang. 2013. [Linking named entities in tweets with knowledge base via user interest modeling](#). In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '13.
- [84] Liangcai Shu, Bo Long, Weiyi Meng. 2009. [A Latent Topic Model for Complete Entity Resolution](#). Proc. ICDE 2009.
- [85] A. Sil, E. Cronin, P. Nie, Y. Yang, A.-M. Popescu, and A. Yates. 2012. [Linking named entities to any database](#). In Proceedings of the Conference on Empirical Methods for Natural Language Processing (EMNLP), pages 116–127.
- [86] Avirup Sil, Alexander Yates. 2013. [Re-ranking for Joint Named-Entity Recognition and Linking](#). Proc. CIKM2013.
- [87] Harish Srinivasan, John Chen and Rohini Srihari. 2009. [Cross-Document Person Name Disambiguation Using Entity Profiles](#). Proc. Text Analysis Conference (TAC2009).
- [88] Zareen Syed, Tim Finin, and Anupam Joshi. 2008. [Wikipedia as an Ontology for Describing Documents](#). Proc. the Second International Conference on Weblogs and Social Media.
- [89] D. Vitale, P. Ferragina, and U. Scaiella. 2012. [Classification of short texts by deploying topical annotations](#). In ECIR, pages 376–387.
- [90] C. Wang, K. Chakrabarti, T. Cheng, and S. Chaudhuri. 2012. [Targeted disambiguation of ad-hoc, homogenous sets of named entities](#). In The International World Wide Web Conference, pages 719–728, New York, NY, USA. ACM.
- [91] Zhichun Wang, Juanzi Li, Jie Tang. 2013. [Boosting Cross-lingual Knowledge Linking via Concept Annotation](#). Proc. IJCAI 2013.
- [92] J. Weng, E. Lim, J. Jiang, and Q He. 2010. [Twitterrank: Finding topic-sensitive influential twitterers](#). In Proceedings of the Third ACM International Conference on Web Search and Data Mining, WSDM '10.
- [93] Jian Xu, Qin Lu, Jie Liu and Ruifeng Xu. 2012. [NLPComp in TAC 2012 Entity Linking and Slot-Filling](#). Proc. Text Analysis Conference (TAC2012).
- [94] Z. Xu, R. Lu, L. Xiang, and Q. Yang. 2011. [Discovering user interest on twitter with a modified author topic model](#). In Web Intelligence.

- [95] Dian Yu, Haibo Li, Taylor Cassidy, Qi Li, Zheng Chen and Heng Ji. 2013. RPI-BLENDER TAC-KBP2013 Knowledge Base Population System Description. Proc. Text Analysis Conference (TAC2013).
- [96] W. Zhang, J. Su, C. Tan, and W. Wang. 2010. [Entity linking leveraging automatically generated annotation](#). In Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010).
- [97] W. Zhang, J. Su, and C. L. Tan. 2011. [A wikipedia-lda model for entity linking with batch size changing](#). In Proc. IJCNLP2011.
- [98] Wei Zhang, Yan Chuan Sim, Jian Su, Chew Lim Tan. 2011. [Entity linking with effective acronym expansion, instance selection and topic modeling](#). Proc. IJCAI 2011.
- [99] Zhicheng Zheng, Fangtao Li, Minlie Huang, Xiaoyan Zhu. 2010. [Learning to link entities with knowledge base](#). Proc. HLT2010.
- [100] Zhicheng Zheng, Xiance Si, Fangtao Li, Edward Y. Chang, Xiaoyan Zhu. 2012. [Entity Disambiguation with Freebase](#). Proc. WI-IAT 2012.
- [101] Yiping Zhou, Lan Nie, Omid Rouhani-Kalleh, Flavian Vasile, and Scott Gaffney. 2010. [Resolving surface forms to wikipedia topics](#). In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1335–1343, Beijing, China, August. Coling 2010 Organizing Committee.
- [102] Yiping Zhou, Lan Nie, Omid Rouhani-kalleh Flavian, Vasile Scott. 2010. [Resolving Surface Forms to Wikipedia Topics](#). Computational Linguistics.