Multi-lingual Entity Discovery and Linking

Avirup Sil
IBM Research AI
Yorktown Heights, NY
avi@us.ibm.com

Heng Ji
Rensselaer Polytechnic Institute
Troy, NY
jih@rpi.edu

Dan Roth
University of Pennsylvania
Philadelphia, PA
danroth@seas.upenn.edu

Silviu Cucerzan
Microsoft Research
Redmond, WA
silviu@microsoft.com

1   Overall

We live in a golden age of information, where we have access to vast amount of data in various forms: text, video and audio. Over the last few years, one of the key task that has been studied in support of natural language understanding and information extraction from text, is the task of Entity Linking (previously studied as Wikification). Entity Linking (henceforth, EL) (Bunescu and Pasca, 2006; Cucerzan, 2007; Ratinov et al., 2011) is the task of mapping mentions of entities in a text document to an entry in a large catalog of entities such as Wikipedia or another knowledge base (KB). It has also been one of the major tasks in the Knowledge-Base Population track at the Text Analysis Conference (TAC) (McNamee and Dang, 2009b; Ji and Grishman, 2011; Ji et al., 2014). Most works in the literature have used Wikipedia as this target catalog of entities because of its wide coverage and its frequent updates made by the community. The previous Entity Linking tutorial in ACL 2014 (Roth et al., 2014) addressed mostly EL research which have focused on English, the most prevalent language on the web and the one with the largest Wikipedia datasets. However, in the last few years research has shifted to address the EL task in other languages, some of which have very large web presence, such as Spanish (Fahrni et al., 2013; Ji et al., 2014), and Chinese (Cao et al., 2014; Shi et al., 2014) but also in others. In particular, there has been interest in cross-lingual EL (Tsai and Roth, 2016; Sil and Florian, 2016): given a mention in a foreign language document, map it to the corresponding page in the English Wikipedia. Beyond the motivation that drives the English EL task – knowledge acquisition and information extraction – in the cross-lingual case and especially when dealing with low resource languages, the hope is to provide improved natural language understanding capabilities for the many languages for which we have few linguistic resources and annotation and no machine translation technology. The LoreHLT2016-2017 evaluation[1] and TAC 2017 pilot evaluation [2] target really low-resource languages like Northern Sotho or Kikuyu which only have about 4000 Wikipedia pages (about 1/1000 the size of the English wikipedia).

The primary goals of this tutorial are to review the framework of cross-lingual EL and motivate it as a broad paradigm for the Information Extraction task. We will start by discussing the traditional EL techniques and metrics and address questions relevant to the adequacy of these to across domains and languages. We will then present more recent approaches such as Neural EL, discuss the basic building blocks of a state-of-the-art neural EL system and analyze some of the current results on English EL. We will then proceed to Cross-lingual EL and discuss methods that work across languages. In particular, we will discuss and compare multiple methods that make use of multi-lingual word embeddings. We will also present EL methods that work for both name tagging and linking in very low resource languages. Finally, we will discuss the uses of cross-lingual EL in a variety of applications like search engines and commercial product selling applications. Also, contrary to the 2014 EL tutorial, we will also focus on Entity Discovery which is an essential component of EL.

The tutorial will be useful for both senior and junior researchers (in academia and industry) 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 road-map of recent approaches, perspectives, and results, as well as point to some of our EL resources that are available to the research community.

---

[1] https://lorehlt.nist.gov/
[2] http://nlp.cs.rpi.edu/kbp/2017/taskspec_pilot.pdf

2 Brief Tutorial Outline

2.1 Motivation and Overview [20 mins]

We will motivate the general EL problem (for English) by teaching the general methods that incorporate distance measures (Ratinov et al., 2011; Sil and Yates, 2013; Cheng and Roth, 2013). We will then briefly discuss multi-lingual IE problems and motivate cross-lingual EL (Ji et al., 2014; Sil and Florian, 2016). Then we will motivate the new trend of modeling distributional representations instead of distance.

2.2 Key Challenges and Multi-lingual Embeddings [20 mins]

We will present some key challenges daunting high-performing *traditional* EL systems and candidate generation and transliteration (Tsai and Roth, 2018) from a knowledge-base. We will also present the models for *traditional* cross-lingual EL (Sil and Florian, 2016; Tsai and Roth, 2016) and discuss some of their challenges: matching context between non-English documents with the English Wikipedia. Recently, neural Entity Discovery and Linking (henceforth, EDL) techniques have combated some of these challenges. These systems use multi-lingual embeddings which are essential building blocks for these neural architectures. Hence, before diving into the architectures we will survey multi-lingual embedding techniques (Mikolov et al., 2013c; Faruqui and Dyer, 2014; Ammar et al., 2016) and which ones work best for neural EL systems and motivate neural EL.

2.3 Neural Methods for EDL [30 mins]

Various shared tasks such as TAC-KBP, ACE and CONLL, along with corpora like OntoNotes and ERE have provided the community substantial amount of annotations for both entity mention extraction (1,500+ documents) and entity linking (5,000+ query entities). Therefore supervised models have become popular again for each step of EDL. Among all of the supervised learning frameworks for mention extraction, the most popular one is a combined Deep Neural Networks architecture consisted of Bi-directional Long Short-Term Memory networks (Bi-LSTM) (Graves et al., 2013) and CRFs (Lample et al., 2016). In TAC-KBP2017 many teams trained this framework from the same training data (KBP2015 and KBP2016 EDL corpora) and the same set of features (word and entity embeddings), but got very different results. The mention extraction F-score gap between the best system and the worst system is about 24%. We will provide a systematic comparison and analysis on reasons that cause this big gap. We will also introduce techniques to make the framework more robust to noise in low-resource settings.

We will then teach neural EL architectures (Globerson et al., 2016; Gupta et al., 2017a; Sil et al., 2018) that can tackle some of the challenges of the traditional systems. Then we will proceed to cross-lingual neural EL and survey the pipelines that most of these EL systems employ: cross-lingual NER and in-document coreference resolution. We will talk about how to model the contexts using various neural techniques like CNNs, LSTMs etc. and how systems compute similarity metrics of varying granularity (Francis-Landau et al., 2016; Sil et al., 2018).

2.4 Coffee break: [30 minutes]

2.5 Language Universal Methods for Cross-lingual EDL [30 mins]

We will then present some recent advances at developing low-cost approaches to perform cross-lingual EL for 282 Wikipedia languages, such as deriving silver-standard annotations by transferring annotations from English to other languages through cross-lingual links and KB properties, refining annotations through self-training and topic selection, deriving language-specific morphology features from anchor links, and mining word translation pairs from cross-lingual links (Pan et al., 2017a). We will also introduce some recent extensions along this line of work, including extending the number of entity types from five to thousands, and its impact on other NLP applications such as Machine Translation.

2.6 Multiple Knowledge Bases [25 mins]

A task that is similar to multi-lingual EL in both definition and approaches is domain-specific linking of entities in documents based on a given set of domains/corresponding knowledge repositories (Gao and Cucerzan, 2017). This task is important for applications such as the analysis and indexing of corporate document repositories, in which many of the entities of interest are not part of the general-knowledge but are rather company-specific and may need to be kept private. Conflating such terminologies and knowledge into one single knowledge model would be both daunting and undesirable. Thus, similarly to handling multiple languages, a system built for multiple-

domain linking, has to model each domain separately. We will discuss a multi-KB entity linking framework that employs one general-knowledge KB and a large set of domain-specific KBs as linking targets that extends the work from (Cucerzan, 2007, 2014a), as well as a supervised model with a large and diverse set of features to detect when a domain-specific KB matches a document targeted for entity analysis (Gao and Cucerzan, 2017).

2.7 New Tasks, Trends and Open Questions [15 mins]

Here, we will address some of the new settings: multi-lingual EL for search engines (Pappu et al., 2017; Tan et al., 2017). We will discuss some open questions such as improving the title candidate generation process for situations where the corresponding titles only exist in the English Wikipedia and also investigate the topological structure of related languages and exploit cross-lingual knowledge transfer to enhance the quality of extraction and linking (Tsai and Roth, 2018). We will also discuss EL for noisy data like social media (Meij et al., 2012; Guo et al., 2013). Finally, we will discuss the possibilities of extending the ideas taught in this EL tutorial to other multi-lingual IE tasks.

2.8 System Demos and Resources [10 mins]

Finally, we will show some demos of multi-lingual EL systems from the industry and academia. We will also provide pointers to resources, including data sets and software.

3 Tutorial Instructors

- **Name:** Avirup Sil
  Affiliation: IBM Research AI
  Email: avi@us.ibm.com
  Website: Avi's Webpage
  Avi is a Research Staff Member and the chair of the NLP community at IBM Research AI. His research interests are in multi-lingual information extraction from large text collection (cross-lingual entity extraction, disambiguation and slot filling), machine learning and knowledge representation. Avi has published several papers on Entity Linking and his systems at IBM have obtained top scores in TAC-KBP annual multi-lingual entity linking evaluations. Avi is an area chair for Information Extraction at NAACL 2018 and also for COLING 2018. He is also organizing the workshop on the "Relevance of Linguistic Structure in Neural NLP" at ACL 2018.

- **Name:** Heng Ji
  Affiliation: Rensselaer Polytechnic Institute
  Email: jih@rpi.edu
  Website: Heng's Webpage
  Heng Ji is the Edward G. Hamilton Development Chair 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 since 2010 and has published many papers on entity discovery and linking. Heng has co-taught the "Wikification and Beyond: The Challenges of Entity and Concept Grounding" tutorial with Dan Roth at ACL 2014.

- **Name:** Dan Roth
  Affiliation: University of Pennsylvania
  Email: danroth@seas.upenn.edu
  Website: Dan's Webpage
  Dan Roth is the Eduardo D. Glandt Distinguished Professor at the Department of Computer and Information Science, University of Pennsylvania. He is a fellow of AAAS, AAAI, ACL, and the ACM and the winner of the IJCAI-2017 John McCarthy Award, for "major conceptual and theoretical advances in the modeling of natural language understanding, machine learning, and reasoning." Roth has published broadly in machine learning, natural language processing, knowledge representation and reasoning, and has developed several machine learning based natural language processing systems that are widely used in the computational linguistics community and in industry. Over the last few years he has worked on Entity Linking and Wikification. He has taught several tutorials at ACL/NAACL/ECL and other forums. Dan has co-taught the "Wikification and Beyond: The Challenges of Entity and Concept Grounding" tutorial with Heng Ji at ACL 2014.

- **Name:** Silviu Cucerzan
  Affiliation: Microsoft Research
  Email: silviu@microsoft.com
  Website: Silviu's Webpage
  Silviu Cucerzan is a Principal Researcher at Microsoft Research and the Bing Knowledge Graph group. His research has focused on topics at the intersection of NLP

and IR with concrete applications to industry, including multilingual spelling correction, question answering, entity recognition and linking, query suggestion, vertical search, and ads selection. Many of the technologies developed by Silviu have been shipped with Microsoft products. The NEMO entity linking system developed by Silviu has scored the top performance during the four consecutive years it participated in the TAC-KBP evaluations organized by NIST and LDC.

References

Waleed Ammar, George Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A Smith. 2016. Massively multilingual word embeddings. *arXiv preprint arXiv:1602.01925* .

Ivo Anastácio, Bruno Martins, Pável Calado, et al. 2011. Supervised learning for linking named entities to knowledge base entries.

Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and space-efficient entity linking for queries. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, pages 179–188.

Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Conference on Artificial Intelligence*. EPFL-CONF-192344.

R. Bunescu and M. Pasca. 2006. Using encyclopedic knowledge for named entity disambiguation. In *EACL*.

Ziqiang Cao, Sujian Li, and Heng Ji. 2014. Joint learning of chinese words, terms and keywords. In *EMNLP*.

David Carmel, Ming-Wei Chang, Evgeniy Gabrilovich, Bo-June Paul Hsu, and Kuansan Wang. 2014. Erd'14: entity recognition and disambiguation challenge. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*.

Jianpeng Cheng, Li Dong, and Mirella Lapata. 2016. Long short-term memory-networks for machine reading. *CoRR* abs/1601.06733. http://arxiv.org/abs/1601.06733.

X. Cheng and D. Roth. 2013. Relational inference for wikification. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Andrew Chisholm and Ben Hachey. 2015. Entity disambiguation with web links. *Transactions of the Association for Computational Linguistics* 3:145–156.

S. Cucerzan. 2014a. Named entities made obvious: The participation in the erd 2014 evaluation. In *SIGIR Forum 2014*.

Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on wikipedia data. In *EMNLP-CoNLL*. pages 708–716.

Silviu Cucerzan. 2014b. The MSR System for Entity Linking at TAC 2014. In *Text Analysis Conference*.

Silviu Cucerzan and Avirup Sil. 2013. The MSR Systems for Entity Linking and Temporal Slot Filling at TAC 2013. In *Text Analysis Conference*.

Hong-Jie Dai, Richard Tzong-Han Tsai, Wen-Lian Hsu, et al. 2011. Entity disambiguation using a markov-logic network. In *IJCNLP*.

Nilesh N. Dalvi, Ravi Kumar, and Bo Pang. 2012. Object matching in tweets with spatial models. In *WSDM*. pages 43–52.

Alexandre Davis, Adriano Veloso, Altigran S da Silva, Wagner Meira Jr, and Alberto HF Laender. 2012. Named entity disambiguation in streaming data. In *ACL*.

Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. 2012. Zencrowd: leveraging probabilistic reasoning and crowdsourcing techniques for large-scale entity linking. In *Proceedings of the 21st international conference on World Wide Web*.

Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2015. Question answering over freebase with multi-column convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. volume 1, pages 260–269.

Greg Durrett and Dan Klein. 2014. A joint model for entity analysis: Coreference, typing, and linking. In *Transactions of the Association for Computational Linguistics*.

Angela Fahrni, Benjamin Heinzerling, Thierry Göckel, and Michael Strube. 2013. Hits monolingual and cross-lingual entity linking system at TAC 2013. In *Text Analysis Conference*.

Manaal Faruqui and Chris Dyer. 2014. Improving vector space word representations using multilingual correlation. ACL.

Matthew Francis-Landau, Greg Durrett, and Dan Klein. 2016. Capturing semantic similarity for entity linking with convolutional neural networks. In *Proc. NAACL 2016*.

Octavian-Eugen Ganea, Marina Ganea, Aurelien Lucchi, Carsten Eickhoff, and Thomas Hofmann. 2016. Probabilistic bag-of-hyperlinks model for entity linking. In *Proceedings of the 25th International*

Conference on World Wide Web. International World Wide Web Conferences Steering Committee, pages 927–938.

N. Gao and S. Cucerzan. 2017. Entity linking to one thousand knowledge bases. In *ECIR 2017*.

Felix A Gers, Jürgen Schmidhuber, and Fred Cummins. 2000. Learning to forget: Continual prediction with lstm. *Neural computation* 12(10):2451–2471.

Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2016. Collective entity resolution with multi-focal attention. *ACL* .

Alan Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. 2013. Hybrid speech recognition with deep bidirectional lstm. In *Automatic Speech Recognition and Understanding, 2013 IEEE Workshop on*.

Stephen Guo, Ming-Wei Chang, and Emre Kıcıman. 2013. To link or not to link? a study on end-to-end tweet entity linking. In *NAACL*.

Zhaochen Guo and Denilson Barbosa. 2014. Robust entity linking via random walks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, pages 499–508.

Nitish Gupta, Sameer Singh, and Dan Roth. 2017a. Entity linking via joint encoding of types, descriptions, and context. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. pages 2671–2680.

Nitish Gupta, Sameer Singh, and Dan Roth. 2017b. Entity linking via joint encoding of types, descriptions, and context. In *EMNLP*. http://cogcomp.org/papers/GuptaSiRo17.pdf.

Ben Hachey, Will Radford, and James R. Curran. 2011. Graph-based named entity linking with wikipedia. In *The 12th International Conference on Web Information System Engineering (WISE)*.

X. Han, L. Sun, and J. Zhao. 2011a. Collective entity linking in web text: a graph-based method. In *SIGIR*.

Xianpei Han, Le Sun, and Jun Zhao. 2011b. Collective entity linking in web text: a graph-based method. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, pages 765–774.

Zhengyan He, Shujie Liu, Mu Li, Ming Zhou, Longkai Zhang, and Houfeng Wang. 2013. Learning entity representation for entity disambiguation. In *ACL (2)*. pages 30–34.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. 2013. Yago2: A spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence* 194:28–61.

Neil Houlsby and Massimiliano Ciaramita. 2014. A scalable gibbs sampler for probabilistic entity linking. In *Advances in Information Retrieval*, Springer, pages 335–346.

Zhiting Hu, Poyao Huang, Yuntian Deng, Yingkai Gao, and Eric P Xing. 2015. Entity hierarchy embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*. volume 1, pages 1292–1300.

Hongzhao Huang, Larry Heck, and Heng Ji. 2015. Leveraging deep neural networks and knowledge graphs for entity disambiguation. *arXiv preprint arXiv:1504.07678* .

Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*. volume 1, pages 1681–1691.

Heng Ji, HT Dang, J Nothman, and B Hachey. 2014. Overview of tac-kbp2014 entity discovery and linking tasks. In *Proc. Text Analysis Conference (TAC2014)*.

Heng Ji and Ralph Grishman. 2011. Knowledge base population: Successful approaches and challenges. In *Proceedings of ACL2011*.

Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2011. Overview of the tac2011 knowledge base population track. *TAC* .

Heng Ji, Joel Nothman, Ben Hachey, and Radu Florian. 2015. Overview of tac-kbp2015 tri-lingual entity discovery and linking. In *TAC*.

Alex Judea, Benjamin Heinzerling, Angela Fahrni, and Michael Strube. 2014. Hits monolingual and cross-lingual entity linking system at TAC 2014. In *Text Analysis Conference*.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188* .

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882* .

Guillaume Lample, Miguel Ballesteros, Kazuya Kawakami, Sandeep Subramanian, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceeddings of the 2016 Conference of*

the North American Chapter of the Association for Computational Linguistics Human Language Technologies.

Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. Plato: A selective context model for entity resolution. *Transactions of the Association for Computational Linguistics* 3:503–515.

Hailun Lin, Zeya Zhao, Yantao Jia, Yuanzhuo Wang, Jinhua Xiong, Xiaojing Li, Yuxiao Chang, Manling Li, Hongbo Xu, and Xueqi Cheng. 2014. Openkn at tac kbp 2014. In *Text Analysis Conference*.

T. Lin, P. Pantel, M. Gamon, A. Kannan, and A. Fuxman. 2012a. Active objects: actions for entity-centric search. In *Proceedings of the 21st international conference on World Wide Web*. pages 589–598.

Thomas Lin, Mausam, and Oren Etzioni. 2012b. Entity Linking at Web Scale. In *AKBC-WEKEX*.

Thomas Lin, Mausam, and Oren Etzioni. 2012c. No Noun Phrase Left Behind: Detecting and Typing Unlinkable Entities. In *EMNLP*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*. pages 2181–2187.

Ying Lin, Chin-Yew Lin, and Heng Ji. 2017. List-only entity linking. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. volume 2, pages 536–541.

Dan Liu, Wei Lin, Shiliang Zhang, Si Wei, and Hui Jiang. 2016. Neural networks models for entity discovery and linking. *arXiv preprint arXiv:1611.03558* .

Xiaoqiang Luo, Radu Florian, and Todd Ward. 2009. Improving coreference resolution by using conversational metadata. In *NAACL*. Association for Computational Linguistics, pages 201–204.

Xiaoqiang Luo, Abe Ittycheriah, Hongyan Jing, Nanda Kambhatla, and Salim Roukos. 2004. A mention-synchronous coreference resolution algorithm based on the bell tree. In *ACL*. page 135.

James Mayfield. 2013. Overview of the kbp 2013 entity linking track.

P. McNamee and H. T Dang. 2009a. Overview of the TAC 2009 Knowledge Base Population Track. In *Text Analysis Conference*.

Paul McNamee and Hoa Trang Dang. 2009b. Knowledge base population: Successful approaches and challenges. In *Proceedings of TAC2009*.

Paul McNamee, Mark Dredze, Adam Gerber, Nikesh Garera, Tim Finin, James Mayfield, Christine Piatko, Delip Rao, David Yarowsky, and Markus Dreyer. 2009. HLTCOE Approaches to Knowledge Base Population at TAC 2009. In *Text Analysis Conference*.

Paul McNamee, James Mayfield, Douglas W Oard, Tan Xu, Ke Wu, Veselin Stoyanov, and David Doermann. 2011. Cross-language entity linking in maryland during a hurricane.

Edgar Meij, Wouter Weerkamp, and Maarten de Rijke. 2012. Adding semantics to microblog posts. In *WSDM*.

Pablo N. Mendes, Max Jakob, and Christian Bizer. 2011. Evaluating DBpedia Spotlight for the TAC-KBP Entity Linking Task. In *TAC*.

Yuval Merhav, Joel Barry, James Clarke, David Murgatroyd, and One Alewife Center. 2013. Basis technology at tac 2013 entity linking .

Qingliang Miao, Ruiyu Fang, Yao Meng, and Shu Zhang. 2013. Frdc's cross-lingual entity linking system at tac 2013.

Rada Mihalcea and Andras Csomai. 2007. Wikify!: Linking documents to encyclopedic knowledge. In *CIKM*. pages 233–242.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013b. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168* .

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013c. Distributed representations of words and phrases and their compositionality. In *NIPS*. pages 3111–3119.

Sean Monahan, Dean Carpenter, Maxim Gorelkin, Kevin Crosby, and Mary Brunson. 2014. Populating a knowledge base with entities and events. In *Text Analysis Conference*.

Thien Huu Nguyen, Yifan He, Maria Pershina, Xiang Li, and Ralph Grishman. 2014. New york university 2014 knowledge base population systems. In *Text Analysis Conference*.

Mark Palatucci, Dean Pomerleau, Geoffrey E Hinton, and Tom M Mitchell. 2009. Zero-shot learning with semantic output codes. In *Advances in neural information processing systems*. pages 1410–1418.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017a. Cross-lingual name tagging and linking for 282 languages. In *Proc. the 55th Annual Meeting of the Association for Computational Linguistics (ACL2017)*.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017b. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. volume 1, pages 1946–1958.

Aasish Pappu, Roi Blanco, Yashar Mehdad, Amanda Stent, and Kapil Thadani. 2017. Lightweight multilingual entity extraction and linking. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, pages 365–374.

Maria Pershina, Yifan He, and Ralph Grishman. 2015. Personalized page rank for named entity disambiguation. In *HLT-NAACL*. pages 238–243.

Xuan-Hieu Phan, Le-Minh Nguyen, and Susumu Horiguchi. 2008. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In *Proceedings of the 17th international conference on World Wide Web*.

L. Ratinov, D. Roth, D. Downey, and M. Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *ACL*. http://cogcomp.cs.illinois.edu/papers/RRDA11.pdf.

Dan Roth, Heng Ji, Ming-Wei Chang, and Taylor Cassidy. 2014. Wikification and beyond: The challenges of entity and concept grounding. In *ACL Tutorials*.

Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*. International World Wide Web Conferences Steering Committee, pages 373–374.

Xing Shi, Kevin Knight, and Heng Ji. 2014. How to speak a language without knowing it. In *ACL*.

Avirup Sil. 2013. Exploring Re-ranking Approaches for Joint Named-Entity Recognition and Linking. In *ACM PIKM*.

Avirup Sil, Ernest Cronin, Penghai Nie, Yinfei Yang, Ana-Maria Popescu, and Alexander Yates. 2012. Linking Named Entities to Any Database. In *EMNLP-CoNLL*.

Avirup Sil, Georgiana Dinu, and Radu Florian. 2015. The ibm systems for trilingual entity discovery and linking at tac 2015. In *Proceedings of the Eighth Text Analysis Conference (TAC2015)*.

Avirup Sil and Radu Florian. 2016. One for all: Towards language independent named entity linking. *ACL* .

Avirup Sil, Gourab Kundu, Radu Florian, and Wael Hamza. 2018. Neural cross-lingual entity linking. In *Proc. Association for the Advancement of Artificial Intelligence (AAAI)*.

Avirup Sil and Alexander Yates. 2013. Re-ranking for Joint Named-Entity Recognition and Linking. In *CIKM*.

Sameer Singh, Amarnag Subramanya, Fernando Pereira, and Andrew McCallum. 2012. Wikilinks: A large-scale cross-document coreference corpus labeled via links to wikipedia. *University of Massachusetts, Amherst, Tech. Rep. UM-CS-2012-015* .

Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. 2013a. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*. pages 926–934.

Richard Socher, Milind Ganjoo, Christopher D Manning, and Andrew Ng. 2013b. Zero-shot learning through cross-modal transfer. In *Advances in neural information processing systems*. pages 935–943.

Richard Socher, Brody Huval, Christopher D Manning, and Andrew Y Ng. 2012. Semantic compositionality through recursive matrix-vector spaces. In *EMNLP*. pages 1201–1211.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013c. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*. volume 1631, page 1642.

Valentin I Spitkovsky and Angel X Chang. 2012. A cross-lingual dictionary for english wikipedia concepts. In *LREC*. pages 3168–3175.

Yaming Sun, Lei Lin, Duyu Tang, Nan Yang, Zhenzhou Ji, and Xiaolong Wang. 2015. Modeling mention, context and entity with neural networks for entity disambiguation. In *IJCAI*. pages 1333–1339.

Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. Lstm neural networks for language modeling. In *INTERSPEECH*. pages 194–197.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. pages 3104–3112.

Chuanqi Tan, Furu Wei, Pengjie Ren, Weifeng Lv, and Ming Zhou. 2017. Entity linking for queries by searching wikipedia sentences. *arXiv preprint arXiv:1704.02788* .

Chen-Tse Tsai and Dan Roth. 2016. Cross-lingual wikification using multilingual embeddings. In *Proceedings of NAACL-HLT*. pages 589–598.

Chen-Tse Tsai and Dan Roth. 2018. Learning better name translation for cross-lingual wikification. In *Proceedings of The Conference on Artificial Intelligence (AAAI)*. http://cogcomp.org/papers/2018_aaai_semanticilp.pdf.

Chi Wang, Kaushik Chakrabarti, Tao Cheng, and Surajit Chaudhuri. 2012. Targeted disambiguation of ad-hoc, homogeneous sets of named entities. In *Proceedings of the 21st international conference on World Wide Web*.

Han Wang, Jin Guang Zheng, Xiaogang Ma, Peter Fox, and Heng Ji. 2015. Language and domain independent entity linking with quantified collective validation. In *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP2015)*.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph and text jointly embedding. In *EMNLP*. Citeseer, pages 1591–1601.

Zhiguo Wang, Haitao Mi, Wael Hamza, and Radu Florian. 2016a. Multi-perspective context matching for machine comprehension. *arXiv preprint arXiv:1612.04211* .

Zhiguo Wang, Haitao Mi, and Abraham Ittycheriah. 2016b. Sentence similarity learning by lexical decomposition and composition. *COLING* .

Daniel S. Weld, Raphael Hoffmann, and Fei Wu. 2009. Using Wikipedia to Bootstrap Open Information Extraction. In *ACM SIGMOD Record*.

UK WR143PS. 1989. Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters .

Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2016. Joint learning of the embedding of words and entities for named entity disambiguation. *arXiv:1601.01343* .

Z. Zheng, X. Si, F. Li, E. Chang, and X. Zhu. 2012. Entity disambiguation with freebase. In *Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence (WI'2012)*.