Scaling out NLP Applications to 100+ Languages

A Lecture Tutorial for The Web Conference 2021

Linjun Shou[†] **Ming Gong**[†] **Jian Pei**[‡] **Xiubo Geng**[†] **Xingjie Zhou**[†] **Daxin Jiang**[†]

[†]Microsoft STCA NLP Group, Beijing, China

[‡]School of Computing Science, Simon Fraser University

{lisho,migon,xigeng,xingzhou,djiang}@microsoft.com jpei@cs.sfu.ca

ABSTRACT

Natural Language Processing models have achieved impressive performance, thanks to the recent deep learning approaches. However, large deep learning models typically rely on huge amounts of human labeled data. There are more than 7,000 languages spoken in the world [1]. Unfortunately, most languages have very limited linguistic resources. Language scaling is invaluable to the advance of social welfare, and thus has attracted intensive interest from industrial practitioners who want to deploy their applications/services to global markets. At the same time, due to the huge differences in the vocabulary, morphology and syntax among different languages, scaling out NLP applications to various languages presents grand challenges to machine learning, data mining, and natural language processing.

In this tutorial, we systematically survey the frontier of language scaling using as examples the research, techniques, and engineering behind a series of concrete NLP applications in Microsoft products and services that need to be scaled out to 100+ languages. We start with a clear problem description for language scaling and an intuitive discussion on the overall challenges,. Then, we explore the problem from data perspective, including the availability of different types of training data as well as the evaluation benchmark data sets for various tasks. A large part of the tutorial will focus on various approaches to language scaling, including cross-lingual models, data augmentation, and language knowledge transferring algorithms. We have applied different approaches for various applications and built a platform to integrate our approaches. Using this platform we will demonstrate several case studies that have been shipped to Microsoft products, and share with the audience our lessons and experience learned. Finally, we will discuss several important challenges in this area and future directions.

1 INTENDED AUDIENCE AND LEVEL

Our tutorial is designed to serve three categories of audience.

First, researchers working on cross-lingual models and algorithms will find our tutorial a systematic survey of the state-of-the-art methods as well as a stimulating discussion on the core challenges and promising directions. The tutorial can serve as fast-track introduction course bringing them to the frontier quickly and equipping them with practical ideas and tools. More importantly, the tutorial connects research and industry best practice and applications.

Second, general audience in the areas of Natural Language Processing (NLP), data mining, and machine learning can get an overall picture of the frontier of language scaling and various approaches, which transfer knowledge from rich-resource languages to low-resource languages. Moreover, researchers in other fields who need to tackle related problems can quickly understand the available techniques that they can borrow to solve their problems.

Third, industrial NLP practitioners will find our tutorial a comprehensive and in-depth reference to the advanced techniques and engineering practice for language scaling. This tutorial will serve as a bridge between the research frontier and the industrial best practice.

We do not assume that the audience has any deep background knowledge in Natural Language Processing, Deep Learning, or Reinforcement Learning. We will build on only some basic concepts in these areas and use sufficient examples to explain the ideas and intuitions.

The tutorial will be delivered in a lecture style. We will provide slides to attendees, with proper copyright permission granted if necessary. In the case that the tutorial is delivered online, we will make sure it will fit virtual online only format.

2 RELEVANCE

Since language scaling is a common challenge in almost every NLP application, it attracts intensive interest from industrial practitioners who want to deploy their applications/services to global markets. Moreover, due to the huge differences in vocabulary, morphology and syntax among different languages, scaling out NLP applications to various languages is a very interesting and challenging task and intrigues prosperous research efforts in machine learning, data mining, and natural language processing. Thanks to the recent progress in deep learning, pre-training, and transfer learning, many approaches for language scaling have been proposed in the past years. It is high time we provided a comprehensive survey and tutorial to facilitate further research and practical applications. The topic of our tutorial will be interesting to many people.

The Web is the largest information and knowledge source in the world with thousands of languages. Cross-lingual access, retrieval,

[1]https://www.ethnologue.com/

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00

https://doi.org/10.1145/1122445.1122456

extraction, and comprehension of Web content produce huge impact to human society. Moreover, the Web itself is also the largest corpus and the richest knowledge source for the training of cross-lingual models. Therefore, the Web Conference is a perfect venue for this tutorial.

3 OUTLINE OF THE TUTORIAL AND LENGTH

The tutorial is planned for three and a half hours.

1. Introduction (20 minutes)
   (A) **Problem setting of language scaling**: given training data in resource rich languages, as well as relatively well trained models in those languages, the problem is to scale the model to hundreds or even thousands of target languages with low resources.
   (B) **Overview of NLP applications**: Several examples include word breaking, spelling correction, information extraction, search relevance, question answering, language understanding etc. They belong to several categories of tasks: classification, sequence labeling, and text generation.
   (C) **Challenges of language scaling**: (1) lack of training data; (2) maintenance cost; and (3) connection among different languages.
   (D) **Overview of approaches to language scaling**: (1) cross-/multi-lingual models; (2) data augmentation; (3) language knowledge transferring algorithms.
2. Data for Language Scaling Tasks (30 minutes)
   (A) **Training data**: (1) Monolingual corpus; (2) General cross-lingual data: bilingual dictionary, comparable data, parallel sentences; (3) Domain/Task specific data in target languages: unlabeled data, knowledge data, weakly supervised data, few shot examples.
   (B) **Evaluation data**: Benchmark evaluation sets, (1) for individual tasks, such as XNLI [12], MTOP [29], SNIPS [50], MultiATIS++ [64], WIKIANN [43], CoNLL-2003 [49], MLQA [28], XQuAD [3], FQuAD [15], XQA [35], TydiQA [10], Taboeba [4], X-stance [59], XPersona [33], entity linking [6, 57], summarization [52, 70], etc; and (2) for a collection of tasks, such as XGLUE [32], XTREME [22], etc.
   (C) **Discussion of data collection and usage**
3. Cross-Lingual Models (30 minutes)
   (A) **Cross-lingual word embedding**: (1) mapping-based methods: supervised, semi-supervised, and unsupervised methods [2, 16, 19, 40, 42, 60, 63]; (2) joint embedding methods [1, 7, 24, 39].
   (B) **Cross-lingual sentence embedding**: (1) Use parallel data: using decoding for alignment, explicit encoder alignment [5, 9, 20, 21, 25, 34, 38, 51, 56, 61, 66, 67].
   (C) **Multi-/Cross-lingual pre-trained contextual models**: (1) Use parallel data, such as XLM [27], Unicoder [23], InfoXLM [8] etc; (2) No parallel data, such as mBERT [14, 45], XLM-R [11], mBART [37], mT5 [65], etc.

Q&A Session 1. (15 minutes)

4. Data Augmentation for Language Scaling Tasks (40 minutes)
   (A) **Machine translation**: (1) translation for train; (2) translation for test [18, 36, 46, 58, 69].
   (B) **Task adaptation**: Construct pre-training data for specific tasks, such as leveraging Wikipedia pages or user behavior signals [31, 54, 62].
   (C) **Synthetic data generation**: (1) pre-train sequence-to-sequence models; (2) task-specific fine-tuning for generation model; and (3) filtering and selection of generated examples [26, 44, 47, 53].
5. Language Knowledge Transferring Algorithms (20 minutes)
   (A) **Knowledge distillation**: transferring knowledge from single or multiple teacher models to a student model [36, 62, 68].
   (B) **Meta learning**: consider each language as a task, and learn good initial model parameters [41, 62].
   (C) **Transfer learning**: learn language-agnostic part and language-specific part, and then fuse these two parts [30].
6. A Platform for Cross-Lingual Tasks (25 minutes)
   (A) **Architecture and components**: (1) data, model and application layers; (2) model training and inference; (3) model life-cycle management.
   (B) **Case studies**: several real cross-lingual applications in Microsoft Bing, Outlook, and Teams.
7. Summary: Challenges and Future Directions: (15 minutes)
   (A) **Challenges**: (1) crowd-sourcing in small languages (throughput and quality control); (2) trade-off for Return-Over-Investment; (3) language-specific features *vs.* language-agnostic features
   (B) **Future directions**: (1) Common representation of languages (syntax and semantics) by large pre-trained models; (2) Zero-shot and few-shot learning by data generation, transfer learning, and active learning; (3) Automatic language scaling for various NLP tasks with different availability of data and constraints of compliance.

Q&A Session 2. (15 minutes)

4 RELATED TUTORIALS

This is a newly developed tutorial. We have not given this tutorial in any forums.

There are a small number of related tutorials. Ruder *et al.* [48] provided a tutorial on unsupervised cross-lingual representation in ACL'19. The tutorial mainly focuses on weakly-supervised and unsupervised cross lingual word embedding in low resource settings where bilingual supervision may not be available. Various approaches, training conditions, robustness for distant language pairs, and applications are discussed. Our tutorial conducts a comprehensive survey on various approaches to language scaling for NLP applications, where cross-lingual word embedding is one of the approaches discussed in Section 3.A.

Other related tutorials [13, 17, 55] target at one specific cross-lingual task each, such as machine translation, entity linking, or cross-lingual parallel data mining. Our tutorial covers a broad range of NLP applications. A unique feature is that, to connect research and industry best practice, we will use as examples the research, techniques, and engineering in Microsoft products and services that need to be scaled out to 100+ languages, including word breaking, spelling correction, information extraction, search relevance, question answering, language understanding, etc.

5 TUTORIAL EXPERIENCE

We have rich and successful experience in delivering well accepted tutorials in premier conferences. For example, Pei gave 25 tutorials in conferences such as KDD (11 times) [2], WWW and SIGIR. He also presented several keynote speeches in some conferences and workshops. Jiang gave tutorials at SIGIR (twice), KDD and WWW. He also delivered keynote speeches and invited talks at many conferences and workshops.

6 SHORT BIOGRAPHIES

Daxin Jiang, Ph.D., Chief Scientist, Microsoft Software Technology Center Asia. Daxin Jiang has years of experience of Research and Engineering in Machine Learning, Data Mining, Natural Language Processing, and Bioinformatics. He received Ph.D. in Computer Science from the Statue University of New York at Buffalo in 2005. He has published extensively in prestigious conferences and journals, and served as a PC member of numerous conferences. He received Best Application Paper Award of SIGKDD'08 and Runner-up for Best Application Paper Award of SIGKDD'04. Daxin is leading an R&D group in Microsoft with 170+ applied scientists and engineers to develop NLP algorithms, applications and platforms, which support various Microsoft products, including Bing, Cortana, Teams, Outlook, and Microsoft Cognitive Services.

Address: 5 Danling Street, Hai Dian, Beijing, China, 100080. Email: djiang@microsoft.com. Tel: +86 (10) 5917 3321.

Jian Pei, Ph.D., Professor, School of Computing Science, Simon Fraser University. His expertise is in developing effective and efficient data analysis techniques for novel data intensive applications. He is a research leader in the general areas of data science, big data, data mining, and database systems. He is recognized as a fellow of Royal Society of Canada (RSC) (i.e., the national academy of Canada), the Canadian Academy of Engineering (CAE), ACM and IEEE. He is one of the most cited authors in data mining, database systems, and information retrieval. His research has generated remarkable impact substantially beyond academia. His algorithms have been adopted by industry in production and popular open source software suites. He is responsible for several commercial systems of record-breaking large scale. As a renowned professional leader, he has played important roles in many academic organizations and activities. He is the Chair of ACM SIGKDD and was the Editor-in-Chief of IEEE TKDE. He received many prestigious awards, including the 2017 ACM SIGKDD Innovation Award and the 2015 ACM SIGKDD Service Award. In his last leave-of-absence from the university, he took the executive roles of two Fortune Global 500 companies. He is a mentor of Creative Destruction Lab (CDL).

Address: 8888 University Drive, Burnaby, BC Canada, V5A 1S6. Email: jpei@cs.sfu.ca. Tel: +1 (778) 782 6851. Fax: +1 (778) 782 3045.

Linjun Shou, Senior Applied Scientist Manager, Microsoft Software Technology Center Asia. He has good publications on several prestigious international conferences such as ACL, EMNLP, COLING, SIGKDD, AAAI, WSDM, etc and served as the program committees on numerous conferences. His research interests include question answering, cross lingual transfer learning, representation learning, etc. Plenty of his research has been transferred to real Microsoft products such as Bing universal question answering, query understanding, document understanding, news/tweets ranking systems. Besides, he is also actively contributing to the academic community through open sourcing projects (e.g. NeuronBlocks) and benchmarks like XGLUE, CodeXGLUE.

Address: 5 Danling Street, Hai Dian, Beijing, China, 100080. Email: lisho@microsoft.com.

Ming Gong, Ph.D., Principal Applied Scientist Manager, Microsoft Software Technology Center Asia. She received Ph.D. on Graphics and Visual Computing in 2013 from Institute of Computing Technology, Chinese Academy of Sciences. She is leading an elite team with 10+ applied scientists and engineers to develop novel NLP technologies for AI applications. Her research interests include question answering, search intelligence, multilingual/cross-modal modeling, representation learning, etc. She published 30+ papers in top conferences and journals (e.g. ACL, COLING, SIGKDD, EMNLP, WSDM, AAAI, PR, CVIU), and also served as PC members of top NLP/AI conferences. Besides, she is actively contributing to the academic community by open-sourcing projects (e.g. NeuronBlocks) and benchmarks like XGLUE, CodeXGLUE. Many of the novel technologies have been transferred to Microsoft's global products and online services including Bing search, multilingual question answering services and document understanding platform.

Address: 5 Danling Street, Hai Dian, Beijing, China, 100080. Email: migon@microsoft.com.

Xiubo Geng, Ph.D., Senior Applied Scientist, Microsoft Software Technology Center Asia. She received Ph.D in Computer Science in 2011 from Institute of Computing Technology, Chinese Academy of Sciences. Her research interests include machine learning, search intelligence, question answering, multilingual modeling, reasoning, etc. She has good publications on top conferences (e.g. SIGIR, NeurIPS, WWW, EMNLP, IJCAI, etc.), and served as a PC member of several conferences.

Address: 5 Danling Street, Hai Dian, Beijing, China, 100080. Email: xigeng@microsoft.com.

Xinjie Zhou, Ph.D., Senior Software Engineer Lead, Microsoft Software Technology Center Asia. He received Ph.D. on natural language processing in 2017 from Peking University. His research interests include machine learning, natural language understanding, multilingual modeling, etc. He has good publications on top conferences including ACL, EMNLP, AAAI, etc.

Address: Bldg #25, 328 Xinghu Street, SIP, Suzhou, China, 215000. Email: xinjzhou@microsoft.com.

REFERENCES

[1] J. Alaux, E. Grave, Marco Cuturi, and Armand Joulin. 2019. Unsupervised Hyperalignment for Multilingual Word Embeddings. *ArXiv* abs/1811.01124 (2019).
[2] David Alvarez-Melis and T. Jaakkola. 2018. Gromov-Wasserstein Alignment of Word Embedding Spaces. In *EMNLP*.
[3] M. Artetxe, Sebastian Ruder, and Dani Yogatama. 2020. On the Cross-lingual Transferability of Monolingual Representations. In *ACL*.
[4] M. Artetxe and Holger Schwenk. 2019. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics* 7 (2019), 597–610.
[5] M. Artetxe and Holger Schwenk. 2019. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics* 7 (2019), 597–610.
[6] Muhao Chen and Y. Tian et al. 2018. Co-training Embeddings of Knowledge Graphs and Entity Descriptions for Cross-lingual Entity Alignment. In *IJCAI*.
[7] Xilun Chen and Claire Cardie. 2018. Unsupervised Multilingual Word Embeddings. In *EMNLP*.

[2] https://www.youtube.com/playlist?list=PL8n-erTbIhTMvdXs657kBOp2pXFJVyAnB

[8] Z. Chi and L. Dong et al. 2020. InfoXLM: An Information-Theoretic Framework for Cross-Lingual Language Model Pre-Training. *ArXiv* abs/2007.07834 (2020).
[9] M. Chidambaram, Yinfei Yang, Daniel Matthew Cer, Steve Yuan, Yun-Hsuan Sung, B. Strope, and R. Kurzweil. 2019. Learning Cross-Lingual Sentence Representations via a Multi-task Dual-Encoder Model. In *RepL4NLP@ACL*.
[10] J. Clark, E. Choi, M. Collins, D. Garrette, T. Kwiatkowski, V. Nikolaev, and J. Palomaki. 2020. TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages. *arXiv preprint arXiv:2003.05002* (2020).
[11] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116* (2019).
[12] Alexis Conneau, Guillaume Lample, Ruty Rinott, Adina Williams, Samuel R Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. XNLI: Evaluating cross-lingual sentence representations. *arXiv preprint arXiv:1809.05053* (2018).
[13] Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. 2020. Multilingual Neural Machine Translation. In *COLING*.
[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL: HLT*. 4171–4186.
[15] Martin d'Hoffschmidt, Maxime Vidal, Wacim Belblidia, and Tom Brendl'e. 2020. FQuAD: French Question Answering Dataset. In *EMNLP*.
[16] Zi-Yi Dou, Zhi-Hao Zhou, and Shujian Huang. 2018. Unsupervised Bilingual Lexicon Induction via Latent Variable Models. In *EMNLP*.
[17] Ahmed El-Kishky, Philipp Koehn, and Holger Schwenk. 2020. Mining the Web for Cross-lingual Parallel Data. In *SigIR*.
[18] Y. Fang, S. Wang, Z. Gan, S. Sun, and JJ. Liu. 2020. FILTER: An Enhanced Fusion Method for Cross-lingual Language Understanding. *ArXiv* abs/2009.05166 (2020).
[19] Manaal Faruqui and Chris Dyer. 2014. Improving Vector Space Word Representations Using Multilingual Correlation. In *EACL*.
[20] S. Gouws, Y. Bengio, and G. S. Corrado. 2015. BilBOWA: Fast Bilingual Distributed Representations without Word Alignments. *ArXiv* abs/1410.2455 (2015).
[21] Mandy Guo and Q. Shen et al. 2018. Effective Parallel Corpus Mining using Bilingual Sentence Embeddings. In *WMT*.
[22] J. Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and M. Johnson. 2020. XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalization. *ArXiv* abs/2003.11080 (2020).
[23] Haoyang Huang, Yaobo Liang, Nan Duan, Ming Gong, Linjun Shou, Daxin Jiang, and Ming Zhou. 2019. Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks. *arXiv preprint arXiv:1909.00964* (2019).
[24] Yunsu Kim, Jiahui Geng, and H. Ney. 2018. Improving Unsupervised Word-by-Word Translation with Language Model and Denoising Autoencoder. In *EMNLP*.
[25] A. Klementiev, Ivan Titov, and Binod Bhattarai. 2012. Inducing Crosslingual Distributed Representations of Words. In *COLING*.
[26] Varun Kumar, A. Choudhary, and Eunah Cho. 2020. Data Augmentation using Pre-trained Transformer Models. *ArXiv* abs/2003.02245 (2020).
[27] Guillaume Lample and Alexis Conneau. 2019. Cross-lingual Language Model Pretraining. *ArXiv* abs/1901.07291 (2019).
[28] Patrick Lewis, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2019. MLQA: Evaluating cross-lingual extractive question answering. *arXiv preprint arXiv:1910.07475* (2019).
[29] Haoran Li and A. Arora et al. 2020. MTOP: A Comprehensive Multilingual Task-Oriented Semantic Parsing Benchmark. *ArXiv* abs/2008.09335 (2020).
[30] Juntao Li, Ruidan He, Hai Ye, H. Ng, Lidong Bing, and Rui Yan. 2020. Unsupervised Domain Adaptation of a Pretrained Cross-Lingual Language Model. In *IJCAI*.
[31] Shining Liang, Linjun Shou, Jian Pei, Ming Gong, WanLi Zuo, and Daxin Jiang. 2020. CalibreNet: Calibration Networks for Multilingual Sequence Labeling. *ArXiv* abs/2011.05723 (2020).
[32] Y. Liang, N. Duan, Y. Gong, N. Wu, F. Guo, W. Qi, Ming Gong, Linjun Shou, Daxin Jiang, G. Cao, X. Fan, B. Zhang, R. Agrawal, E. Cui, S. Wei, T. Bharti, Y. Qiao, J. Chen, W. Wu, S. Liu, F. Yang, R. Majumder, and M. Zhou. 2020. XGLUE: A New Benchmark Dataset for Cross-lingual Pre-training, Understanding and Generation. *arXiv* abs/2004.01401 (2020).
[33] Zhaojiang Lin, Zihan Liu, Genta Indra Winata, Samuel Cahyawijaya, Andrea Madotto, Yejin Bang, Etsuko Ishii, and Pascale Fung. 2020. XPersona: Evaluating Multilingual Personalized Chatbot. *ArXiv* abs/2003.07568 (2020).
[34] Robert Litschko and Goran Glavas et al. 2019. Evaluating Resource-Lean Cross-Lingual Embedding Models in Unsupervised Retrieval. *SIGIR* (2019).
[35] Jiahua Liu, Yankai Lin, Z. Liu, and Maosong Sun. 2019. XQA: A Cross-lingual Open-domain Question Answering Dataset. In *ACL*.
[36] Junhao Liu, Linjun Shou, Jian Pei, Ming Gong, Min Yang, and Daxin Jiang. 2020. Cross-lingual Machine Reading Comprehension with Language Branch Knowledge Distillation. *ArXiv* abs/2010.14271 (2020).
[37] Yinhan Liu and Jiatao Gu et al. 2020. Multilingual Denoising Pre-training for Neural Machine Translation. *ArXiv* abs/2001.08210 (2020).
[38] B. McCann, James Bradbury, Caiming Xiong, and R. Socher. 2017. Learned in Translation: Contextualized Word Vectors. In *NIPS*.
[39] T. Mohiuddin and Shafiq R. Joty. 2019. Revisiting Adversarial Autoencoder for Unsupervised Word Translation with Cycle Consistency and Improved Training. *ArXiv* abs/1904.04116 (2019).
[40] Ndapandula Nakashole. 2018. NORMA: Neighborhood Sensitive Maps for Multilingual Word Embeddings. In *EMNLP*.
[41] Farhad Nooralahzadeh and Giannis Bekoulis et al. 2020. Zero-Shot Cross-Lingual Transfer with Meta Learning. *ArXiv* abs/2003.02739 (2020).
[42] Aitor Ormazabal, M. Artetxe, Gorka Labaka, A. Soroa, and Eneko Agirre. 2019. Analyzing the Limitations of Cross-lingual Word Embedding Mappings. *ArXiv* abs/1906.05407 (2019).
[43] Xiaoman Pan and Boliang Zhang et al. 2017. Cross-lingual Name Tagging and Linking for 282 Languages. In *ACL*.
[44] Baolin Peng, Chenguang Zhu, Michael Zeng, and Jianfeng Gao. 2020. Data Augmentation for Spoken Language Understanding via Pretrained Models. *ArXiv* abs/2004.13952 (2020).
[45] Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is Multilingual BERT? *arXiv preprint arXiv:1906.01502* (2019).
[46] L. Qin, Minheng Ni, Y. Zhang, and W. Che. 2020. CoSDA-ML: Multi-Lingual Code-Switching Data Augmentation for Zero-Shot Cross-Lingual NLP. In *IJCAI*.
[47] Jun Quan and Deyi Xiong. 2019. Effective Data Augmentation Approaches to End-to-End Task-Oriented Dialogue. In *2019 IALP*. IEEE, 47–52.
[48] Sebastian Ruder, Anders Søgaard, and I. Vulic. 2019. Unsupervised Cross-Lingual Representation Learning. In *ACL*.
[49] E. T. K. Sang and F. D. Meulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. *ArXiv* cs.CL/0306050 (2003).
[50] S. Schuster, S. Gupta, Rushin Shah, and M. Lewis. 2019. Cross-lingual Transfer Learning for Multilingual Task Oriented Dialog. *ArXiv* abs/1810.13327 (2019).
[51] Holger Schwenk and M. Douze. 2017. Learning Joint Multilingual Sentence Representations with Neural Machine Translation. In *Rep4NLP@ACL*.
[52] T. Scialom, P. Dray, S. Lamprier, B. Piwowarski, and J. Staiano. 2020. MLSUM: The Multilingual Summarization Corpus. *ArXiv* abs/2004.14900 (2020).
[53] Siamak Shakeri, Noah Constant, Mihir Kale, and Linting Xue. 2020. Multilingual Synthetic Question and Answer Generation for Cross-Lingual Reading Comprehension. *ArXiv* abs/2010.12008 (2020).
[54] Linjun Shou, S. Bo, Fei-Xiang Cheng, Ming Gong, J. Pei, and Daxin Jiang. 2020. Mining Implicit Relevance Feedback from User Behavior for Web Question Answering. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery Data Mining* (2020).
[55] A. Sil, Heng Ji, D. Roth, and S. Cucerzan. 2018. Multi-lingual Entity Discovery and Linking. In *ACL*.
[56] K. Singla, Dogan Can, and Shrikanth S. Narayanan. 2018. A Multi-task Approach to Learning Multilingual Representations. In *ACL*.
[57] Zequn Sun, Wei Hu, and C. Li. 2017. Cross-Lingual Entity Alignment via Joint Attribute-Preserving Embedding. *ArXiv* abs/1708.05045 (2017).
[58] Ferhan Türe and Elizabeth Boschee. 2016. Learning to Translate for Multilingual Question Answering. *ArXiv* abs/1609.08210 (2016).
[59] Jannis Vamvas and Rico Sennrich. 2020. X -stance: A Multilingual Multi-Target Dataset for Stance Detection. *ArXiv* abs/2003.08385 (2020).
[60] Ivan Vulic and A. Korhonen. 2016. On the Role of Seed Lexicons in Learning Bilingual Word Embeddings. In *ACL*.
[61] Ivan Vulic and Marie-Francine Moens. 2015. Monolingual and Cross-Lingual Information Retrieval Models Based on (Bilingual) Word Embeddings. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2015).
[62] Q. Wu, Zijia Lin, Börje F. Karlsson, B. Huang, and Jianguang Lou. 2020. UniTrans : Unifying Model Transfer and Data Transfer for Cross-Lingual Named Entity Recognition with Unlabeled Data. In *IJCAI*.
[63] Chao Xing, D. Wang, Chao Liu, and Yiye Lin. 2015. Normalized Word Embedding and Orthogonal Transform for Bilingual Word Translation. In *HLT-NAACL*.
[64] Weijia Xu, B. Haider, and Saab Mansour. 2020. End-to-End Slot Alignment and Recognition for Cross-Lingual NLU. *ArXiv* abs/2004.14353 (2020).
[65] Linting Xue and Noah Constant et al. 2020. mT5: A massively multilingual pre-trained text-to-text transformer. *ArXiv* abs/2010.11934 (2020).
[66] Yinfei Yang and Daniel Matthew Cer et al. 2020. Multilingual Universal Sentence Encoder for Semantic Retrieval. In *ACL*.
[67] Yinfei Yang and G. Ábrego et al. 2019. Improving Multilingual Sentence Embedding using Bi-directional Dual Encoder with Additive Margin Softmax. *ArXiv* abs/1902.08564 (2019).
[68] Ze Yang, Linjun Shou, Ming Gong, Wutao Lin, and Daxin Jiang. 2020. Model Compression with Two-stage Multi-teacher Knowledge Distillation for Web Question Answering System. *WSDM* (2020).
[69] Fei Yuan, Linjun Shou, X. Bai, Ming Gong, Yaobo Liang, N. Duan, Y. Fu, and Daxin Jiang. 2020. Enhancing Answer Boundary Detection for Multilingual Machine Reading Comprehension. In *ACL*.
[70] Junnan Zhu, Q. Wang, Yining Wang, Y. Zhou, Jiajun Zhang, Shaonan Wang, and C. Zong. 2019. NCLS: Neural Cross-Lingual Summarization. *ArXiv* abs/1909.00156 (2019).