Maksim Tkachenko

CONTACT INFORMATION	E-ma Phone Home Linke Goog	il:maksim.tkatchenko@gmail.come:+65 82969128epage:www.mtkachenko.infoedIn:maksim-tkachenkode Scholar ID:AXrs46UAAAAJ	
INTEREST AREAS	Language Modelling, Text Mining, Natural Language Processing, Computational Linguistics, Machine Lear ing, Statistics, Probabilistic Graphical Models, Interpretability, Causal Inference		
Education		PhD, Singapore Management University, Singapore School of Information Systems, 2015 - 2019, Full-Ride Scholarship + Competitive Presidential Doctoral Fellowship Dissertation: Comparison Mining from Text Supervisor: Hady W. Lauw Dissertation Committee: Jiang Jing, Zheng Baihua, Xiaoli Li Specialist (Master's), St. Petersburg State University, Russia Faculty of Mathematics and Mechanics, Software Engineering Department, 2006 - 2011, Competitive Full-Ride Scholarship Thesis: Named Entity Recognition Technique Based on Wikipedia Supervisor: Boris Novikov	
SCIENTIFIC PUBLICATIONS (PEER-REVIEWED)	[1] [2]	Ween Jiann Lee, Maksim Tkachenko, and Hady W. Lauw. "Robust Bidirectional Poly-Matching". In: <i>IEEE Transactions on Knowledge and Data Engineering</i> , TKDE , <i>Journal Paper</i> , (2023). Alberto Poncelas, Maksim Tkachenko, and Ohnmar Htun. "Sakura at SemEval-2023 Task 2: Data Augmentation via Translation". In: <i>Proc. of International Workshop on Semantic Evaluation</i> , SemEval <i>Workshop Paper</i> (Accented for Publication). ACL, 2023	
	[3]	Chong Cher Chia, Maksim Tkachenko, and Hady W. Lauw. "Morphologically-Aware Vocabulary Re- duction of Word Embeddings". In: <i>Proc. of International Conference on Web Intelligence and Intelli-</i> <i>gent Agent Technology</i> , IEEE/WIC/ACM , <i>Full Paper</i> . 2022.	
	[4]	Ween Jiann Lee, Maksim Tkachenko, and Hady W. Lauw. "Robust BiPoly-Matching for Multi-Granular Entities". In: <i>Proc. of IEEE International Conference on Data Mining</i> , IEEE ICDM , <i>Short Paper</i> . 2021.	
	[5]	Maksim Tkachenko and Hady W Lauw. "CompareLDA: A Topic Model for Document Comparison". In: <i>Proc. of AAAI Conference on Artificial Intelligence</i> , AAAI , <i>Full Paper</i> . 2019.	
	[6]	Maksim Tkachenko, Chong Cher Chia, and Hady W. Lauw. "Searching for the X-Factor: Exploring Corpus Subjectivity for Word Embeddings". In: <i>Proc. of Annual Meeting of the Association for Computational Linguistics</i> , ACL, <i>Full Paper</i> . 2018.	
	[7]	Maksim Tkachenko and Hady W. Lauw. "Comparative Relation Generative Model". In: <i>IEEE Transactions on Knowledge and Data Engineering</i> , TKDE , <i>Journal Paper</i> , 29.4 (2017).	
	[8]	Maksim Tkachenko and Hady W. Lauw. "Plackett-Luce Regression Mixture Model for Heterogeneous Rankings". In: <i>Proc. of ACM International Conference on Information and Knowledge Management</i> , CIKM , <i>Full Paper</i> . 2016.	
	[9]	Maksim Tkachenko and Hady W. Lauw. "A Convolution Kernel Approach to Identifying Comparisons in Text". In: <i>Proc. of Annual Meeting of the Association for Computational Linguistics and Interna-</i> <i>tional Joint Conference on Natural Language Processing</i> , ACL-IJCNLP, <i>Full Paper</i> . 2015.	
	[10]	Maksim Tkachenko and Hady W. Lauw. "Generative Modeling of Entity Comparisons in Text". In: <i>Proc. of ACM International Conference on Information and Knowledge Management, CIKM, Full Paper.</i> 2014.	

- [11] Rinat Gareev, Maksim Tkachenko, Valery Solovyev, Andrey Simanovsky, and Vladimir Ivanov. "Introducing Baselines for Russian Named Entity Recognition". In: *Proc. of Computational Linguistics* and Intelligent Text Processing, CICLing, Full Paper. 2013.
- [12] Maksim Tkachenko and Andrey Simanovsky. "Named Entity Recognition: Exploring Features". In: *Proc. of Conference on Natural Language Processing*, **KONVENS**, *Full Paper*. 2012.
- [13] Maksim Tkachenko and Andrey Simanovsky. "Selecting Features for Domain-Independent Named Entity Recognition". In: Proc. of Conference on Natural Language Processing, KONVENS, Short Paper. 2012.
- [14] Maksim Tkachenko, Alexander Ulanov, and Andrey Simanovsky. "Classifying Wikipedia Entities into Fine-Grained Classes". In: Proc. of IEEE International Conference on Data Engineering (Workshops), ICDEW, Workshop Paper. 2011.

Employment

Rakuten Institute of Technology, Singapore

Research Scientist 09/2022 (Current Position)

I am conducting applied natural language processing research in the fields of machine translation, language modeling, and information retrieval, including analysis of large language models.

FedEx Express, Analytics Center of Excellence, Singapore

Data Science Advisor 06/2021 - 08/2022

I led data science efforts in optimizing customs clearance processes from the natural language processing perspective: inferring a valid tariff code from the shipment description, flagging controlled packages, etc.

School of Computing and Information Systems, Singapore Management University

Research Scientist 12/2018 - 05/2021

Research Assistant 11/2013 - 07/2015

I led a team of engineers, research interns, and graduate students in developing a comprehensive ecommerce knowledge base. Our knowledge base contains over 47m products and 100m reviews from various e-commerce platforms. It serves as a backbone for multiple prototypes and research projects (e.g., SnappyBuyer, ThriftCity).

St. Petersburg State University, Russia

Research Engineer 07/2012 - 11/2013

Junior Researcher 11/2011 - 07/2012

I researched statistical methods for natural language processing, specifically information extraction and knowledge base population. During the appointment, I developed a suite of information extraction tools (e.g., named entity recognition, entity linking) for multiple languages (e.g., English, French).

Hewlett-Packard Labs, St. Petersburg, Russia

Research Engineer 11/2011 - 11/2013

Research Intern 12/2009 - 11/2011

This is a joint appointment with St. Petersburg State University.

I conducted applied and basic research in machine learning and natural language processing. The lab delivered AL/ML expertise to improve the business processes of internal and external customers.

Lanit-Tercom, Inc., R&D department, St. Petersburg, Russia

Software Engineer 09/2009 - 12/2010

I was responsible for the development and maintenance of the news search engine augmented with sentiment/opinion analysis. As a part of language processing efforts, I worked with a linguist to develop tools for low-level language processing of Russian (e.g., part-of-speech tagger, lemmatizer).

JetBrains, MPS Team, St. Petersburg, Russia

Software Developer 06/2009 - 09/2009

I was developing domain-specific languages.

Awards & Grants	• 2017-2018 & 2018-2019, Presidential Doctoral Fellow, Singapore Management University (competi- tive, the recipients of this fellowship are selected from the top 5% of PhD students across SMU).		
	• Development Technology of Information Extraction Systems for the Russian Language, 2013-2015. (with V. Ivanov, F. Nikolaev, M. Sidikov, A. Simanovsky, R. Gareev, R. Yamilov and V. Solovyev). The project is supported by Russian Foundation for Basic Research.		
	• SIGIR Student Travel Grants, 2016.		
Mentorship	 Preferred.AI Summer Internship (SMU), Co-Founder and Mentor, 2019-2020. Interns: Cao Wanyue, Steffi Tan Xin Rong, Goh Wan Xuan 		
	- Neural Network Lab, a visual language for designing and evaluating neural network models.		
	- SnappyBuyer, a platform that allows you to make better decisions when buying electronics.		
ACADEMIC SERVICE	• PC Member: AAAI (2020-2023), IJCAI (2020-2023), IJCAI-ECAI (2022), PAKDD (2020-2023), SEIM (2016-2018, 2020), SYRCoDIS 2014.		
	• Invited Conference Reviewer: DSAA 2017, PAKDD 2017, DSAA 2016, PAKDD 2015, CIKM 2014, ICDM 2014.		
	• Invited Journal Reviewer: Transactions of the Association for Computational Linguistics (TACL), Lan- guage Resources and Evaluation Journal, Journal of the Association for Information Science and Tech- nology (JASIST).		
PUBLIC TALKS	• <i>Evolution of Language Models: From N-Grams to Transformers</i> , Rakuten Institute of Technology x DataScience SG, Singapore, Apr 2023		
	• Panel on Research in Academia & Industry, Rakuten Technology Conference, Singapore, Nov 2022		
	• Sentiment-Infused Word Embeddings with SentiVec,		
	- VNU Hanoi, University of Engineering and Technology, Vietnam, Dec 2019		
	- VNU Ho Chi Minh University of Science, Vietnam, Dec 2019		
	- Data Science Meetup by SEEK Asia & Big Data Malaysia, Kuala Lumpur, Malaysia, Mar 2019		
	CompareLDA: A Topic Model for Document Comparison, Preferred.AI TechFest, Singapore, Aug 2019		
	Sentive: Sentiment-Infused word Embeddings, Data Science Consortium, Singapore, Jun 2018		
	• Learning User Frejerences from Multi-Modal Data. Frejerence Signal from Review Text,		
	- St. Petersburg Academic University, Russia, Mar 2018		
	- Computer Science Club of Stekrov Institute of Mathematics, St. Petersburg, Mar 2018		
	- Third University, St. Fetersburg, Russia, Wai 2018		
	- Avito Data Science Meetup, Moscow, Russia, Feb 2018		
	 Moscow Institute of Physics and Technology (MIPT) Russia Feb 2018 		
	Language & Artificial Intelligence Interdisciplinery Electronic Arts Survey National University of		
	Singapore (NUS), Sep 2018		
	• <i>Comparison Mining</i> , The Three Minute Thesis (3MT) Competition, Nanyang Technological University, Singapore, Jun 2017		

Research Prototypes	• CompareLDA (Comparative Latent Dirichlet Allocation) learns predictive topic distributions that com- ply with pairwise comparison observations.		
	• Plackett-Luce Regression Mixture Model finds salient preference groups within a population of rankers.		
	• Tree-SVM learns SVM-based classifiers over tree structures (constituency and dependency trees).		
	• SentiVec learns word embeddings and enriches them with specified lexical information.		
	• Venom is a focused crawler framework.		
	Simple-ML implements online learning algorithms for classification.		
LANGUAGES	English (fluent), Russian (native), Chinese (basic)		