Ruey-Cheng Chen

Canva
30 Rupert St
Collingwood, VIC 3066
Australia

rcchen@canva.com
http://rueycheng.com
twitter:rueycheng
github:rueycheng

Education

2013   **PhD**, *National Taiwan University*
Computer Science and Information Engineering
Dissertation: "Information preservation and its application to natural language processing"

2003   **MS**, *National Taiwan University*
Computer Science and Information Engineering

2001   **BS**, *National Taiwan University*
Computer Science and Information Engineering

Experience

2022–present   **Staff Machine Learning Engineer**, *Canva*
On search and AI modeling

2022   **Senior Data Scientist**, *SEEK Ltd*
2018–2021   **Data Scientist**, *SEEK Ltd*
On search and recommender systems

2015–2018   **Research Fellow**, *RMIT University*
With Mark Sanderson, Falk Scholer (School of CS&IT), and W. Bruce Croft (UMass Amherst) on question answering and ad hoc retrieval; With J. Shane Culpepper (School of CS&IT) and Roi Blanco (Amazon) on search efficiency and learning to rank

2013–2014   **Postdoctoral Fellow**, *National Taiwan University / Academia Sinica*
Contract position on various data analytics projects

2010–2012   **Research Assistant**, *National Taiwan University*
With Jieh Hsiang (CS&IE) on natural language processing

2009–2010   **Visiting Scholar**, *University of Southern California*
With Andrew S. Gordon (ICT/USC) on topic modeling over narrative texts

2004–2008   **Research Assistant**, *Academia Sinica*
With Lee-Feng Chien (Google) on information retrieval and web mining

Research Profile

I am generally interested in research on Natural Language Processing and Information Retrieval, with a strong focus on ranking and theoretical models. From day to day I am mostly engaged in applied research activities such as model building and evaluation, with some occasional exposure on theoretical work. I have published 11 peer-reviewed papers in prestigious venues such as SIGIR, CIKM, WSDM, EMNLP, and ACL. I have an h-index of 12, with 620 citations to my research work according to Google Scholar (July 2023). One of my earlier work was awarded Best Paper in ROCLING 2009.

Area of Interest

Question Answering, Representation & Neural Models, Ranking, Search/Recommender Systems, Text Generation, and NLP Applications

PC Member/Reviewer:

- WebConf 2018/2022–2023, EMNLP 2020–2023, ACL 2023, NAACL 2021, KDD 2021–2023, SIGIR 2016–2023, WSDM 2017/2021–2023, CIKM 2017–2023, ECIR (European Conference on IR Research) 2017, ADCS (Australasian Document Computing Symposium) 2021–2022
- IEEE TKDE 2014/2018, ACM TOIS 2020
- Workshop: LSDS-IR 2014/2015, SCST 2017, SPIRE 2018

Teaching

2013   **Guest Lecturer**, *National Taiwan University*
Formal Language and Automata Theory, Fall 2013 (3 lectures)

2007
2008
2011   **Teaching Assistant**, *National Taiwan University*
Formal Language and Automata Theory, Fall 2011
Web Retrieval and Mining, Fall 2007/2008

Honors and Awards

2016   **Hugo Research Grant**, *Abbrevi8*
With Mark Sanderson, Falk Scholer, Damiano Spina, Zhifeng Bao, and W. Bruce Croft. "A learning to rank framework for entity centrality". A\$13,489.

2009   **Best Paper Award**, *ROCLING '09*
With Chia-Jung Lee, Yi-Chun Lin, Pei-Sen Liu, and Pu-Jen Cheng. "Query formulation by selecting good terms". In *Proceedings of the 21st Conference on Computational Linguistics and Speech Processing (ROCLING '09)*

2008   **Graduate Students Study Abroad Program**, *NSC Taiwan*
Travel grants for 1-year overseas visit to University of Southern California

2008   **Distinguished Teaching Assistant Award**, *National Taiwan University*
Web Retrieval and Mining, Fall 2007. Granted by Department of CS&IE

Invited Talks

2021   **Search and recommendation at SEEK**, *Lecture at RMIT University*

2021   **So you wanna practise science in industry**, *Lecture at Swinburne University of Technology*

2024   **NLP 101: An industry perspective**, *Talk at UNSW Sydney*

Publications

Refereed Conference Papers

Chen, R.-C., & Lee, C.-J. (2020). Incorporating Behavioral Hypotheses for Query Generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 3105–3110). Association for Computational Linguistics.

Chen, R.-C., Ai, Q., Jayasinghe, G., & Croft, W. B. (2019). Correcting for Recency Bias in Job Recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2185–2188).

Gallagher, L., Chen, R.-C., Blanco, R., & Culpepper, J. S. (2019). Joint Optimization of Cascade Ranking Models. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining* (pp. 15–23). ACM.

Yulianti, E., Chen, R.-C., Scholer, F., Croft, W. B., & Sanderson, M. (2018a). Ranking Documents by Answer-Passage Quality. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 335–344). ACM.

Chen, R.-C., Azzopardi, L., & Scholer, F. (2017). An Empirical Analysis of Pruning Techniques. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management* (pp. 2023–2026). ACM.

Chen, R.-C., Yulianti, E., Sanderson, M., & Croft, W. B. (2017). On the Benefit of Incorporating External Features in a Neural Architecture for Answer Sentence Selection. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1017–1020). ACM.

Chen, R.-C., Gallagher, L., Blanco, R., & Culpepper, J. S. (2017). Efficient Cost-Aware Cascade Ranking for Multi-Stage Retrieval. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 445–454). ACM.

Yang, L., Ai, Q., Spina, D., Chen, R., Pang, L., Croft, W. B., ... Scholer, F. (2016). Beyond Factoid QA: Effective Methods for Non-Factoid Answer Sentence Retrieval. In *Proceedings of the 38th European Conference on IR Research* (pp. 115–128). Springer-Verlag.

Chen, R.-C., Lee, C.-J., & Croft, W. B. (2015). On Divergence Measures and Static Index Pruning. In *Proceedings of the 2015 International Conference on the Theory of Information Retrieval* (pp. 151–160). ACM.

Chen, R.-C. (2013a). An Improved MDL-Based Compression Algorithm for Unsupervised Word Segmentation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 166–170). Association for Computational Linguistics.

Chen, R.-C., & Lee, C.-J. (2013). An Information-Theoretic Account of Static Index Pruning. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 163–172). ACM.

Chen, R.-C., Lee, C.-J., Tsai, C.-M., & Hsiang, J. (2012). Information Preservation in Static Index Pruning. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (pp. 2487–2490). ACM.

Chen, R.-C., Tsai, C.-M., & Hsiang, J. (2010). Relevance Model Revisited: With Multiple Document Representations. In *Proceedings of the 6th Asia Information Retrieval Societies Conference* (pp. 37–48). Springer Berlin / Heidelberg.

Lee, C.-J., Lin, Y.-C., Chen, R.-C., & Cheng, P.-J. (2009). Selecting Effective Terms for Query Formulation. In *Proceedings of the 5th Asia Information Retrieval Symposium on Information Retrieval Technology* (pp. 168–180). Springer-Verlag.

Lee, C.-J., Chen, R.-C., Kao, S.-H., & Cheng, P.-J. (2009). A Term Dependency-Based Approach for Query Terms Ranking. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 1267–1276). ACM.

Cheng, P.-J., Teng, J.-W., Chen, R.-C., Wang, J.-H., Lu, W.-H., & Chien, L.-F. (2004). Translating Unknown Queries with Web Corpora for Cross-Language Information Retrieval. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 146–153). ACM.

Hsiung, P.-A., Wang, F., & Chen, R.-C. (2000). On the Verification of Wireless Transaction Protocol Using SGM and RED. In *Proceedings of the Seventh International Conference on Real-Time Systems and Applications*, IEEE Computer Society.

Refereed Journal Papers

Yulianti, E., Chen, R.-C., Scholer, F., Croft, W. B., & Sanderson, M. (2018b). Document Summarization for Answering Non-Factoid Queries. *IEEE Transactions on Knowledge and Data Engineering*, *30*(1), 15–28.

Day, W.-Y., Chi, C.-Y., Chen, R.-C., & Cheng, P.-J. (2010). Sampling the Web as Training Data for Text Classification. *International Journal of Digital Library Systems (IJDLS)*, *1*(4), 24–42.

Other Publications

Benham, R., Gallagher, L., Mackenzie, J., Damessie, T. T., Chen, R.-C., Scholer, F., ... Culpepper, J. S. (2017). RMIT at the TREC CORE Track. In *Proceedings of The Twenty-Sixth Text REtrieval Conference*.

Gallagher, L., Mackenzie, J., Benham, R., Chen, R.-C., Scholer, F., & Culpepper, J. S. (2017). RMIT at the NTCIR-13 We Want Web Task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies* (pp. 402–406).

Mackenzie, J., Chen, R.-C., & Culpepper, J. S. (2016). RMIT at the TREC 2016 LiveQA Track. In *Proceedings of The Twenty-Fifth Text REtrieval Conference.*

Yulianti, E., Chen, R.-C., Scholer, F., & Sanderson, M. (2016). Using Semantic and Context Features for Answer Summary Extraction. In *Proceedings of the 21st Australasian Document Computing Symposium* (pp. 81–84). ACM.

Ong, K., Chen, R.-C., & Scholer, F. (2016). RMIT at the NTCIR-12 MobileClick-2: iUnit Ranking and Summarization Subtasks. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies* (pp. 147–150).

Chen, R., Culpepper, J. S., Damessie, T. T., Jones, T., Mourad, A., Ong, K., … Yulianti, E. (2015). RMIT at the TREC 2015 LiveQA Track. In *Proceedings of The Twenty-Fourth Text REtrieval Conference.*

Chen, R.-C., Spina, D., Croft, W. B., Sanderson, M., & Scholer, F. (2015). Harnessing Semantics for Answer Sentence Retrieval. In *Proceedings of the Eighth Workshop on Exploiting Semantic Annotations in Information Retrieval* (pp. 21–27). ACM.

Chen, R.-C. (2014). *Incremental Learning for Fully Unsupervised Word Segmentation Using Penalized Likelihood and Model Selection.* Unpublished. Retrieved from https://arxiv.org/abs/1607.05822

Chen, R.-C. (2013b). *Information Preservation and Its Application to Natural Language Processing* (PhD dissertation, National Taiwan University).

Chen, R.-C., Tsai, C.-M., & Hsiang, J. (2012). A Regularized Compression Method to Unsupervised Word Segmentation. In *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology* (pp. 26–34). Association for Computational Linguistics.

Chen, R.-C., Swanson, R., & Gordon, A. S. (2010). *An Adaptation of Topic Modeling to Sentences.* Unpublished. Retrieved from https://arxiv.org/abs/1607.05818

Lee, C.-J., Lin, Y.-C., Chen, R.-C., Liu, P.-S., & Cheng, P.-J. (2009). Query Formulation by Selecting Good Terms. In *Proceedings of the 21st Conference on Computational Linguistics and Speech Processing* (pp. 69–84). Taichung, Taiwan.

Day, W.-Y., Chi, C.-Y., Chen, R.-C., Cheng, P.-J., & Liu, P.-S. (2009). Web Mining for Unsupervised Classification. In *Proceedings of the 21st Conference on Computational Linguistics and Speech Processing* (pp. 53–68). Taichung, Taiwan.

Liao, S.-P., Cheng, P.-J., Chen, R.-C., & Chien, L.-F. (2005). LiveImage: Organizing Web Images by Relevant Concepts. In *Proceedings of the Workshop on the Sciences of the Artificial* (pp. 210–220). Hualien, Taiwan.

Skills

Computer / Programming

Linux/Unix: 15+ years
Python: main language (8+ years)
C/C++, Java: proficient (4-8 years)
Perl, R, Golang: experienced

Technical Software

CoreNLP and NLTK (NLP)
Scikit-Learn, LightGBM, TensorFlow and PyTorch (Machine Learning)
Indri and Lucene/Solr (Search Systems)

Language

Taiwanese (native), Mandarin (native), English

Hobbies

Puzzles: Sometimes just hang out on Project Euler
Bass guitar: Played with full enthusiasm since 2000; been on and off in recent years