

# Ruey-Cheng Chen

---

Discipline of Computer Science & IT  
RMIT University  
GPO Box 2476  
Melbourne, VIC 3001  
Australia  
+61 3 9925 2711

[ruey-cheng.chen@rmit.edu.au](mailto:ruey-cheng.chen@rmit.edu.au)  
<http://rueycheng.com>  
[twitter:rueycheng](https://twitter.com/rueycheng)  
[github:rueycheng](https://github.com/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

- 2015–present **Research Fellow, RMIT University**  
With Mark Sanderson, Falk Scholer (School of CS&IT), and W. Bruce Croft (UMass Amherst) on question answering and summarization; with J. Shane Culpepper (School of CS&IT) and Roi Blanco (formerly Yahoo! Research) on search efficiency
- 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 (Institute for Creative Technologies) on topic modeling over narrative texts
- 2004–2008 **Research Assistant, Academia Sinica**  
With Lee-Feng Chien (Institute of Information Science) 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 probabilistic modeling, ranking, and theories. My typical daily work involves carefully crafting systems with the best known technologies, seeking further improvements with new ideas implemented as experiments/algorithms, and grinding away on research clusters. I have published 19 peer reviewed conference papers, 8 of them in prestigious venues such as ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Annual Meeting of the Association for Computational Linguistics (ACL), and ACM International Conference on Information and Knowledge Management (CIKM). I have an h-index of 6, with 207 citations to my research work according to Google Scholar, as of February 2017. One of my earlier work was awarded Best Paper in ROCLING 2009.

## AREA OF INTEREST

Question Answering, Summarization, Representation, Retrieval Models, Learning to Rank, Probabilistic Modeling (Information Theory/Statistical Inference), Machine Learning, Search Efficiency, and NLP Applications

## SERVICES

PC Member:

- The Web Conference (WWW) 2018
- ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR) 2016/2017
- ACM International Conference on Web Search and Data Mining (WSDM) 2017
- ACM International Conference on Information and Knowledge Management (CIKM) 2017
- European Conference on IR Research (ECIR) 2017
- Large-Scale Distributed System for Information Retrieval 2014/2015 (WSDM/CIKM workshop)
- Supporting Complex Search Tasks 2017 (CHIIR '17 workshop)

Reviewer:

- IEEE Transactions on Knowledge and Data Engineering (TKDE)

## Teaching

- 2013 **Guest Lecturer, National Taiwan University**  
Formal Language and Automata Theory, Fall 2013 (3 lectures)
- 2007 **Teaching Assistant, National Taiwan University**  
2008 Formal Language and Automata Theory, Fall 2011  
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.
- 2013 **SIGIR Travel Grant, ACM SIGIR**  
Part of the grant comes from Donal B. Crouch grant from the Singhal Family Foundation
- 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 visiting 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

## Publications

### REFEREED CONFERENCE/WORKSHOP PAPERS

- Chen, R.-C., Azzopardi, L., & Scholer, F. (2017). An empirical analysis of pruning techniques. In *Proceedings of CIKM '17*. to appear.
- Chen, R.-C., Yulianti, E., & Sanderson, M. (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). SIGIR '17. 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). SIGIR '17. ACM.

- 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). ADCS '16. 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). ECIR '16. Springer-Verlag.
- 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). ESAIR '15 (CIKM workshop). ACM.
- 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). ICTIR '15. 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). ACL '13. 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). SIGIR '13. 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). CIKM '12. ACM.
- 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). SIGMORPHON '12 (NAACL workshop). Association for Computational Linguistics.
- 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). AIRS '10. Springer Berlin / Heidelberg.
- 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). ROCLING '09. 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). ROCLING '09. Taichung, Taiwan.
- 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). AIRS '09. 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). CIKM '09. ACM.
- 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). WSA '05. Hualien, Taiwan.
- 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). SIGIR '04. 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*. RTCSA '00. IEEE Computer Society.

## REFEREED JOURNAL PAPERS

- Yulianti, E., Chen, R.-C., Scholer, F., Croft, W. B., & Sanderson, M. (2017). Document Summarization for Answering Non-Factoid Queries. *IEEE Transactions on Knowledge and Data Engineering*. in press.

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**

- 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*. TREC 2016.
- 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). NTCIR-12.
- 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*. TREC 2015.
- 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., Swanson, R., & Gordon, A. S. (2010). *An adaptation of topic modeling to sentences*. Unpublished. Retrieved from <https://arxiv.org/abs/1607.05818>

## **Skills**

### **COMPUTER / PROGRAMMING**

Linux/Unix: 13+ years  
C/C++, Java, Perl: 8+ years  
R, Python: 4+ years  
Golang: 1 year

### **TECHNICAL SOFTWARE**

CoreNLP and NLTK (NLP)  
Scikit-Learn, xgboost, RankLib, and Theano/Lasagne (Machine Learning)  
Indri, Terrier, Galagosearch, and Lucene (Search Systems)  
Familiarities with tools in specific areas such as question answering and summarization  
Been painstakingly hacking Indri for 6+ years (index optimization/retrieval experiments)  
Recently worked a lot on xgboost and Theano (ranking experiments)

### **LANGUAGE**

Taiwanese: native  
Mandarin Chinese: 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