

# Xin QIAN

xinq@umd.edu · (412) 983 - 3408 · [Homepage](#) · [GitHub.com/xeniaqian94](#)

## EDUCATION

---

**University of Maryland, College Park** 08/2018 - Present

*Doctor of Philosophy in Information Studies; Advised by Prof. Joel Chan*

- Research interest: **natural language processing, human-computer interaction and dialog system.**

**Carnegie Mellon University, School of Computer Science** 08/2016 - 07/2018

*Master of Science in Language Technologies; Research-oriented degree; GPA: 3.69/4.00*

- Courses: Deep Learning, Machine Learning for Text Mining, Algorithm for Natural Language Processing, Search Engines, Machine Translation, Large-Scale Multimedia Analysis, Deep Reinforcement Learning, Neural Network for NLP

**Zhejiang University, Chu Kochen Honors College** 10/2012 - 07/2016

*Bachelor of Engineering in Software Engineering; GPA: 3.95/4.00;*

## PROFESSIONAL EXPERIENCE

---

**Human-Computer Interaction Lab, UMD** College Park, MD

*Research and Teaching Assistant*

*10/2016 - 04/2018*

- **Computational analogy:** Developed a pipeline of weak annotation, training sequence tagger and unsupervised analogy matching for paper abstracts. Designed as an information extraction [Snorkel](#) application with subjective ground truth.

**iDST (DAMO Academy), Alibaba Group** China

*Algorithm Engineering Intern*

*05/2018 - 08/2018*

- **Shopkeeper bot:** Implemented an attribute-extraction module for the chatbot at [2.taobao.com](#). Trained models from weakly supervised consumer dialog data. Experimented on the dialog state tracking task with WOZ 2.0 dataset. [\[Media\]](#)

**Language Technologies Institute, Carnegie Mellon University** Pittsburgh, PA

*Research and Teaching Assistant*

*10/2016 - 04/2018*

- **Multimodal Neural Machine Translation:** Implemented the early fusion multimodal model with two optimization techniques, MLE and advantage actor-critic algorithm. Evaluated on the WMT18 shared task corpus. [\[Report\]](#)
- **Duolingo second language acquisition:** Trained a BiLSTM-CRF-based sequence tagger with syntactic, lemma-based, and user meta-features to predict word-level mistakes. [\[Group report\]](#)
- **Reinforcement LeToR:** Modeled ad-hoc web search as sequential document selection with REINFORCE algorithm on the LETOR 3.0 dataset. [\[Code\]](#)
- **Crowdsourcing Pipeline:** Parsed and indexed 50 million ClueWeb09 web pages with 1K TREC Web/MQ track queries, then rendered and deployed Indri sentence retrieval snippets on CrowdFlower, with random sampling. [\[Interface\]](#)
- **Data Cleaning and Evaluation:** Processed 20K session data for turker qualification based on mouse movement tracking and user-based collaborative filtering, then trained and evaluated a LeToR model with massive click labels.

**Risk and Regulatory Services Innovation Center, Sponsored by PwC** Pittsburgh, PA

*Research Assistant*

*10/2016 - 04/2018*

- **Topic Tracking with Active Learning:** Community monitoring through training text classifiers with Naive Bayes/SVM, filtering with embedding space matching, geo-fencing with NER, and cost-sensitive annotation sampling.
- **Learning to Rank for Conversation:** Measured LeToR effectiveness on conversations with textual/temporal/metadata feature engineering. Collected a 10K annotation pool sampled from field-based relevance models (BM25F, etc.).

**Data Systems Group, University of Waterloo** Canada

*Research Assistant; GPA: 88.2/100*

*10/2015 - 05/2016*

- **Push notification:** Built a RESTful client system YoGosling, and integrated into Anserini, an open-source project.
- **A/B Testing vs. Interleaving:** Explored the two evaluation strategies on TREC Microblog track.

## PUBLICATION

---

**Xin Qian**, Jimmy Lin, and Adam Roegiest. Interleaved evaluation for retrospective summarization and prospective notification on document streams. In *SIGIR'16*, pages 175-184. ACM. Oral presentation. [\[PDF\]](#)

## SKILLS

---

**Languages:** Java, Python, HTML/CSS/Javascript, Linux/Unix, SQL, C/C++

**Tools:** PyTorch, pandas, scikit-learn, Git, Spark, Hadoop, AWS, CrowdFlower, Django, D3.js, seaborn