CV: Khoi-Nguyen Tran

CONTACT INFORMATION	Postdoctoral Researcher IBM Research 60 City Road Southbank, VIC 3006, Australia

Mobile: +61 424 316 544
Email: khndtran@au1.ibm.com
WWW: kndtran.com

OBJECTIVE

To develop and apply my research skills in industry by creating innovative solutions for business problems with real-world impact.

RESEARCH INTERESTS

Data Science, Machine/Deep Learning, Text Mining, Natural Language Processing, Knowledge Discovery and Data Mining, AI for Business

EMPLOYMENT

IBM Research – Australia

- Postdoctoral Researcher. Industry AI.   Jun 2016 – Current
  - Manager: Dr. Christopher J. Butler
  - Various research and development projects with IBM Watson Education.
  - Developed techniques for chunking documents and generating learning objectives from keyphrases and Bloom's verbs. Work recognised as division led expertise.
  - Transitioned and significantly improved past research components for the business teams. For example, fully parallelised a bottleneck component and wrapped components in scripts or microservice APIs for one-click execution and standardised interaction end-points.
  - Other research work included geolocation of Twitter tweets, detecting vandalism on Wikipedia, and predicting price of items from their images.
  - Publications: [3], [15]

Australian Federal Government

- Data Scientist. Research and Development Team.   May 2015 – Jun 2016
  - Worked on sensitive (but unclassified) and classified projects.

The Australian National University (ANU)

- Research Assistant & Co-founder. Cybercrime Observatory. Mar 2013 – May 2015
  - Supervisors & Co-founders: Dr. Mamoun Alazab, Prof. Roderic Broadhurst
  - Projects: Investigating Malicious Spam Emails, History of Cybercrime Activities
  - Developed a novel detection technique for malicious content (attachments or URLs) only using email text. High detection rates demonstrated for over 1 million emails with attachments and 21 million emails with URLs.
  - Publication: [7].

- Research Assistant. Research School of Computer Science.   Jan 2011 – May 2015
  - Supervisors: Prof. Peter Christen, Dr. Scott Sanner, Dr. Lexing Xie
  - Project: Detecting Abnormal Text Values
    * Researched techniques to automatically detect abnormal text values from large databases using only the distribution of text data within those databases.
  - Project: New Objective Functions for Social Collaborative Filtering
    * Developed a Facebook app to collect user data over 5 months from 100+ participants and their 37,000+ friends.
  - Project: Synthetic Data Generation and Corruption
    * Developed a user interface for a novel synthetic data generation method that mimics real-world errors through a variety of data corruption methods.
  - Publications: [1], [4], [16]

EDUCATION   The Australian National University (ANU)

- Ph.D. in Engineering and Computer Science   Feb 2010 – Jul 2015
  - Area of Study: Machine Learning Applications
  - Thesis Title: *Detecting Vandalism on Wikipedia across Multiple Languages*
  - Supervisors: Prof. Peter Christen, Dr. Scott Sanner, Dr. Lexing Xie
  - Publications: [6], [8], [9], [10], [11]

- Bachelor of Computer Science, with First Class Honours   Feb 2006 – Dec 2009
  - GPA: 6.75 / 7. Overall course average of High Distinction.
  - Honours Thesis Topic: *Semantic Sensor Composition*
    * Supervisor: Dr. Michael Compton
    * Publications: [2], [5], [12]
  - Individual Research Projects: *Detecting Network Anomalies*
    * Supervisor: Dr. Huidong (Warren) Jin
    * Publications: [13], [14]

SKILLS   Languages:
- Experienced: Java, Python, R, SQL, HTML, JavaScript, Unix Shell Scripting
- Familiar: Scala, CSS

Software:
- Experienced: Scikit-learn, JQuery, Flask, MySQL, dplyr / tidyverse, Eclipse, PDFBox, GitHub, ZenHub, IBM Cloud, LaTeX, Word/Excel/Powerpoint, Ubuntu Linux, Windows, macOS
- Familiar: Tensorflow, NumPy, Docker, Kubernetes, NLTK, Stanford NLP

ACADEMIC   The Australian National University (ANU)
AWARDS
- ANU Supplementary Scholarship, 2010–2014
- Australian Postgraduate Award (APA), 2010–2014
- ANU College of Engineering and Computer Science Dean's List, 2009
- ANU College of Engineering and Computer Science Dean's Prize, 2008
- Boyapati Prize for 2nd Year Computer Science and Mathematics, 2007
  - Awarded to students obtaining a High Distinction (highest grade possible) in two computer science and two mathematics courses in their 2nd year of study.
- Bachelor of Computer Science (Honours) Scholarship, 2006–2009

SERVICE   Mentor to 3 people, 2017-2018
Volunteer / Lecturer / Talk, VietAI, 2018 (TBC)
Invited Talk/Lecture, International Vietnamese Academics Network, 2018 (TBC)
Invited Talk, RMIT Vietnam, 2017
Invited Participant, Future Shapers Forum, 2017
Delegate, Australia-Vietnam Young Leadership Dialogue, 2017
Volunteer, Australasian Data Mining Conference (AusDM), 2013
Volunteer, Open Source Developers' Conference (OSDC), 2011
Microsoft Student Ambassador, 2011–2012
Volunteer and President, ANU Computer Science Students' Association, 2007–2012

REFERENCES   **Dr. Chris Butler** (email: chris.butler@au1.ibm.com)
AVAILABLE TO - IBM Research
CONTACT   - 60 City Road, Southbank, VIC 3006, Australia
             - *He is my manager.*

Prof. Peter Christen (email: peter.christen@anu.edu.au)
- Professor, Research School of Computer Science
- The Australian National University, Canberra, ACT 2601, Australia
- *He was my primary supervisor and panel chair for my PhD research.*

Prof. Roderic Broadhurst (email: roderic.broadhurst@anu.edu.au)
- Professor, College of Arts and Social Science
- The Australian National University, Canberra, ACT 2601, Australia
- *He was my supervisor at the ANU Cybercrime Observatory.*

CITED
PUBLICATIONS

Full list available at Google Scholar:
http://scholar.google.com.au/citations?user=ihFcT5QAAAAJ

[1] P. Christen, R. W. Gayler, K.-N. Tran, J. Fisher, and D. Vatsalan. Automatic Discovery of Abnormal Values in Large Textual Databases. *Journal of Data and Information Quality (JDIQ)*, 2016.

[2] M. Compton, H. Neuhaus, K. Taylor, and K. Tran. Reasoning about Sensors and Compositions. In *Proceedings of the 2nd International Workshop on Semantic Sensor Networks*, Washington DC, USA, 2009.

[3] J. H. Lau, L. Chi, K.-N. Tran, and T. Cohn. End-to-end Network for Twitter Geolocation Prediction and Hashing. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP)*, 2017.

[4] J. Noel, S. Sanner, K.-N. Tran, P. Christen, L. Xie, E. V. Bonilla, E. Abbasnejad, and N. D. Penna. New Objective Functions for Social Collaborative Filtering. In *Proceedings of the 21st International World Wide Web Conference (WWW)*, Lyon, France, 2012.

[5] K.-N. Tran. Semantic Sensor Composition. Master's thesis, The Australian National University, 2009. Honours Thesis. The Australian National University.

[6] K.-N. Tran. *Detecting Vandalism on Wikipedia across Multiple Languages*. PhD thesis, The Australian National University, 2015.

[7] K.-N. Tran, M. Alazab, and R. Broadhurst. Towards a Feature Rich Model for Predicting Spam Emails containing Malicious Attachments and URLs. In *Proceedings of the 11th Australasian Data Mining Conference (AusDM)*, Canberra, Australia, 2013.

[8] K.-N. Tran and P. Christen. Cross Language Prediction of Vandalism on Wikipedia Using Article Views and Revisions. In *Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, Gold Coast, Australia, 2013.

[9] K.-N. Tran and P. Christen. Identifying Multilingual Wikipedia Articles based on Cross Language Similarity and Activity. In *Proceedings of the 22nd ACM Conference of Information and Knowledge Management (CIKM): Poster*, San Francisco, USA, 2013.

[10] K.-N. Tran and P. Christen. Cross-Language Learning from Bots and Users to detect Vandalism on Wikipedia. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2015.

[11] K.-N. Tran, P. Christen, S. Sanner, and L. Xie. Context-Aware Detection of Sneaky Vandalism on Wikipedia across Multiple Languages. In *Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, Ho Chi Minh City, Vietnam, 2015.

[12] K.-N. Tran, M. Compton, J. Wu, and R. Goré. Short Paper: Semantic Sensor Composition. In *Proceedings of the 3rd International Workshop on Semantic Sensor Networks*, Shanghai, China, 2009.

[13] K.-N. Tran and H. Jin. Fusion of Decision Tree and Gaussian Mixture Models for Heterogeneous Data Sets. In *Proceedings of the International Conference on Multimedia Technology (ICIMT)*, 2009.

[14] K.-N. Tran and H. Jin. Detecting Network Anomalies in Mixed-Attribute Data Sets. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD)*, 2010.

[15] K.-N. Tran, J. H. Lau, D. Contractor, U. Gupta, B. Sengupta, C. J. Butler, and M. Mohania. Document Chunking and Learning Objective Generation for Instruction Design. In *(under review)*, 2018.

[16] K.-N. Tran, D. Vatsalan, and P. Christen. GeCo - An Online Personal data Generator and Corruptor. In *Proceedings of the 22nd ACM Conference of Information and Knowledge Management (CIKM): Demo*, San Francisco, USA, 2013.