

Khushboo Thaker (Singhal)

Phone: 412-954-8721

Email: k.thaker@pitt.edu

Website: <http://kthaker.com/>

Education and Certifications

Ph.D. Intelligent Systems Program, University of Pittsburgh, PA – expected Fall 2023

MTech. Computer Science, DA-IICT, India

B. Eng. Information Technology, Saurashtra University, India

Current Project:

HELPeR (Jan 2020 – present)

(Graduate Research Scholar – NIH Project - <https://pitt.edu/~dah44/helper>)

- Proposed a novel approach to utilize domain specificity for document difficulty or knowledge level prediction (publication 3). We found our approach gives 5-point increase in f-score in predicting knowledge-level
- Health articles filled with specialized medical vocabulary, which makes these articles challenging for lay users to understand. I am currently working on building a dynamic patient knowledge model by using their activities on an online forum. This will help recommender system to prioritize health articles based on patients' knowledge-level, so each patient receives most relevant documents which they understand.

Open Corpus Personalized Learning (Jul 2017 – Dec 2019)

(Graduate Research Scholar – NSF Project - <https://bit.ly/2YpGLYN>)

- Remedial recommendation for student learning utilizing student knowledge state which improved the state-of-the-art text-based recommendation by 9 points in F1 score. (publication 2)
- Improved student modeling and adaptation by incorporating student reading behavior to existing student modeling frameworks, helps adaptation in online textbook learning (publication 5 & 7)
- Education content linking, suffers with the problem of vocabulary mismatch at key phrase level, we tried to overcome this by building domain-specific key phrase embeddings, utilizing textbook and Wikipedia resource (publication 6)

Previous Experience

PhD ML Intern – Facebook Search (Summer 2019)

- Click through bias is a challenge in Search ranking, which is partially solved by integrating human judgement. But human judgement is expensive to obtain. In this work I tried transfer learning to reweight user personalized records. The model we developed was as good as rule-based model for personalization, which was upper bound. The main purpose of the project was to remove exponentially growing rule-based overhead in search pipeline.

Project Associate (Oct 2012 – Aug 2016) (Prof. N. Balakrishnan)

[Information Systems Lab](#) - Supercomputer Education and Research Centre, Indian Institute of Science IISc

- Build a method on improving the accuracy of ensemble of Machine Learning (ML) classifiers using fusion through Neural Network. Results show a significant increase in performance on twitter sentiment classification task.
- Developed a Trend tracking application using Temporal Topic Analysis and change in sentiment.

Junior Research Fellow (May 2011 – Jul 2012) <http://sandhan.tdil-dc.gov.in/>

[Information Retrieval Lab](#) - Dhirubhai Ambani Institute of Information and Communication Technology

- Developed semi-supervised as well as unsupervised Gujarati language stemmer incorporated into the backend of a cross lingual search engine. This work includes hierarchical agglomerative clustering and rule-based learning.

M.Tech. Dissertation

SMS Query Processor for Information Retrieval and FAQ Retrieval, Guide: Dr. Prasenjit Majumder DA-IICT

- Developed a System aims to retrieve most relevant information based on query through noisy improvised text (for example SMS or online lingo).
- Proposed an Approximate String machine metric to handle user improvisation and displayed significant improvement in FAQ retrieval task

Selected Publications: [Google Scholar](#):

Conference

1. Meng, R., **Thaker, K.**, Zhang, L., Dong, Y., Yuan, X., Wang, T., & He, D. (2021). Bringing Structure into Summaries: a Faceted Summarization Dataset for Long Scientific Documents. ACL. 2021
2. **K. Thaker**, L. Zhang, D. He and P. Brusilovsky. "Recommending Remedial Recommendation using Student Knowledge State", Education Data Mining, 2020.
3. **K. Thaker**, Y. Lee and P. Brusilovsky and D. He. "Automated categorization of online health documents using Domain Specificity", AMIA 2020 (poster)
4. Yuan X, Wang T, Meng R, **Thaker K**, Brusilovsky P, He D, Trischler A. One Size Does Not Fit All: Generating and Evaluating Variable Number of Keyphrases. ACL. 2020
5. **K. Thaker**, P. Carvalho, and K. Koedinger. "Comprehension Factor Analysis: Modeling student's reading behavior: Accounting for reading practice in predicting students' learning in MOOCs." LAK. ACM, 2019.
6. **K. Thaker**, P. Brusilovsky, and D. He. "Concept Enhanced Content Representation for Linking Educational Resources.", IEEE/WIC/ACM International Conference on Web Intelligence (WI). IEEE, 2018.
7. **K. Thaker**, Y. Huang, P. Brusilovsky and D. He. Dynamic Knowledge Modeling with Heterogeneous Activities for Adaptive Textbooks. Education Data Mining, 2018

Journal

8. Chau, H., Labutov, I., **Thaker, K.**, He, D., & Brusilovsky, P. (2020). Automatic concept extraction for domain and student modeling in adaptive textbooks. International Journal of Artificial Intelligence in Education, 1-27.
9. H. Daqing, Z. Wang, **K. Thaker** and N. Zhou. "Translation and Expansion: Enabling Laypeople Access to the COVID-19 Academic Collection." Data and Information Management. 2020.
10. [in review JMIR Cancer] **K. Thaker**, Y. Chi, S. Birkhoff, H. Daqing, P. Brusilovsky, Y. Lee, H. Donovan. "Exploring Resource Sharing behaviors in an Online Ovarian Cancer Community"

Skill Set

- Machine Learning tools – PyTorch, Scikit-Learn, Lucene
- Programming – Python, Java, Bash, SQL

Achievements and Certification

- Provost Graduate Fellowship from School of Computing and Information - (Aug 2019 – Jul 2020)
- Graduate Fellowship from Kenneth P. Dietrich School of Arts and Sciences – (Aug 2017 – Jul 2018)
- Graduate Fellowship from Kenneth P. Dietrich School of Arts and Sciences – (Aug 2016 – Jul 2017)
- Among Top 2% in National Level Postgraduate entrance exam GATE and GATE scholarship (2011)

Extracurricular Activities

- Department Webmaster - May 2016 – Apr 2018
- Volunteered Gujarati and Hindi Corpus to IRSI (Information Retrieval Society of India)
- GitHub Profile - <https://github.com/khushsi>
- Publication Profile - [Google Scholar](#)
- Home Page - <http://kthaker.com/>