

# SURAJ NAIR

Department of Computer Science, University of Maryland College Park, MD 20740

+1(240) 917-7287 ◊ srnair@umd.edu

## EDUCATION

---

### PhD in Computer Science

*Aug 2016 - May 2023 (expected)*

University of Maryland, College Park

### Bachelor of Technology in Computer Engineering

*Jun 2010 - May 2014*

Veerмата Jijabai Technological Institute, Mumbai, India (CGPA: 9.0/10.0)

## WORK EXPERIENCE

---

### Amazon Inc., *Applied Scientist Intern*

May 2022 - Aug 2022

- Developed an open domain question answering pipeline as part of the Alexa Web-based Question Answering team.
- Built a fully differentiable end-to-end pipeline comprising a neural context reranker and answer generator

### Johns Hopkins University HLTCOE, *Visiting Researcher*

Jun 2021 - Aug 2021

- Built a cross-lingual generalization of the multi-vector dense retrieval model, ColBERT, as part of the SCALE 2021 workshop on CLIR.
- Explored two training approaches, zero-shot, and translate-train, using translated MS MARCO.

### Raytheon BBN Technologies, *Intern*

Jun 2020 - Aug 2020

- Developed a neural CLIR reranker using multilingual pretrained models (e.g., mBERT, XLM-R).
- Leveraged data from related tasks such as cross-lingual QA to train the CLIR reranker.

### University of Maryland, *Graduate Research Assistant*

August 2018 - May 2023

- Working with Prof. Doug Oard on building cross-language information retrieval systems.

### Samsung R&D Institute, Bangalore, India, *Senior Software Engineer*

Jul 2014 - Apr 2016

- Worked on detection and classification of car driving maneuvers using smartphone sensors.

## RESEARCH EXPERIENCE

---

### Cross-Language Neural Retrieval Models

Jun 2021 - Ongoing

*with Prof. Doug Oard*

- Explored cross-lingual transfer learning using monolingual retrieval collections (MS MARCO) to build dense and sparse neural retrieval models.
- Focused on Pareto-optimal systems that balance the effectiveness-efficiency tradeoffs of the neural models, which include retrieval effectiveness, indexing latency, and query latency.

### Neural CLIR Reranking

Jun 2020 - May 2021

*with Prof. Doug Oard*

- Used multilingual pretrained language models (e.g., mBERT, XLM-R) to build neural reranker.
- Developed a rationale training approach for reranking long documents to improve effectiveness.

with Prof. Doug Oard

- Improved CLIR effectiveness through a combination of statistical and neural translation approaches.
- Explored two fusion approaches, pre, and post-retrieval, to combine evidence from different systems.

## SELECTED PUBLICATIONS

---

- **Suraj Nair**, Eugene Yang, Dawn Lawrie, James Mayfield, and Douglas W. Oard. “*BLADE: Combining Vocabulary Pruning and Intermediate Pretraining for Scaleable Neural CLIR*”, To appear in SIGIR 2023.
- Dawn Lawrie, James Mayfield, Douglas Oard, Eugene Yang, **Suraj Nair**, and Petra Galuščáková. “*HC3: A Suite of Test Collections for CLIR Evaluation over Informal Text*”, To appear in SIGIR 2023.
- **Suraj Nair**, Eugene Yang, Dawn Lawrie, James Mayfield, and Douglas W. Oard. “*Learning a Sparse Representation Model for Neural CLIR*”, DESIRES 2022
- **Suraj Nair**, Eugene Yang, Dawn Lawrie, Kevin Duh, Paul McNamee, Kenton Murray, James Mayfield, and Douglas W. Oard. “*Transfer Learning Approaches for Building Cross-Language Dense Retrieval Models*”, ECIR 2022
- Eugene Yang, **Suraj Nair**, Ramraj Chandradevan, Rebecca Iglesias-Flores, and Douglas W. Oard. “*C3: Continued Pretraining with Contrastive Weak Supervision for Cross Language Ad-Hoc Retrieval*”, SIGIR 2022
- Yanda Chen, Chris Kedzie, **Suraj Nair**, Petra Galuščáková, Rui Zhang, Douglas W. Oard, and Kathleen McKeown. “*Cross-language Sentence Selection via Data Augmentation and Rationale Training*”, ACL 2021
- **Suraj Nair**, Petra Galuščáková, and Douglas W. Oard. “*Combining Contextualized and Non-contextualized Query Translations to Improve CLIR*”, SIGIR 2020
- **Suraj Nair**, Anton Ragni, Ondrej Klejch, Petra Galuscakova, and Douglas W. Oard. “*Experiments with Cross-Language Speech Retrieval for Lower-Resource Languages*”, AIRS 2019

## MANUSCRIPTS

---

- **Suraj Nair**, Petra Galuščáková, Douglas W. Oard, Le Zhang, Damianos Karakos, and Bonan Min. “*Rationale Training Based Neural Re-Ranking for Ad-hoc CLIR*”, under review
- Eugene Yang, **Suraj Nair**, Dawn Lawrie, James Mayfield, and Douglas W. Oard. “*Parameter-efficient Zero-shot Transfer for Cross-Language Dense Retrieval with Adapters*”, under review
- Petra Galuščáková, Douglas W. Oard, and **Suraj Nair**. “*Cross-language Information Retrieval*”, book chapter under review

## HONORS AND AWARDS

---

- Secured 1st place in the TREC 2020 Spotify Podcast retrieval task
- Secured 3rd place in the CLPsych 2017 shared task
- Receptient of Deans Fellowship, Department of Computer Science, University of Maryland, 2016-2018

## TECHNICAL SKILLS

---

**Programming Languages**

Python, C/C++, Java

**Deep Learning Toolkits**

PyTorch, Transformers, AllenNLP

**Information Retrieval Toolkits**

Indri, Lucene, Anserini