

4. *Which Similarity Metric to Use for Software Documents? A study on Information Retrieval based Software Engineering Tasks.* **Md Masudur Rahman**, Saikat Chakraborty, Baishakhi Ray. The 40th International Conference on Software Engineering (ICSE '18 Companion), 2 pages, 2018. [Poster] [Link]
5. *Finding Similar Projects in GitHub using Word2Vec and WMD.* **Md Masudur Rahman**. Workshop on the Naturalness of Software (NL+SE 2016) at FSE 2016. [Slides]
6. *Topic Model based Privacy Protection in Personalized Web Search.* Wasi Ahmad, **Md Masudur Rahman**, Hongning Wang. The 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR2016), 4 pages, 2016. [Link]

RESEARCH
EXPERIENCE

Graduate Research Assistant
Supervisor: Baishakhi Ray

Summer 2016 - Summer 2018

Research Project

- **Impact of Similarity Measure on IR based Software Engineering Tasks**
Evaluated the performance of different similarity metrics (VSM, BM25, LSI, and embedding based WMD) on various SE documents w.r.t. two tasks: (i) similar project recommendation and (ii) bug localization. It has been observed that in general the context-aware models achieve better performance on textual artifacts. In contrast, simple keyword-based bag-of-words models perform better in code artifacts.
- **Code Intent Analysis**
Developed an automated technique to classify search queries to code vs. non-code leveraging Stack Overflow tags. A code score is calculated for each query based on how many stack-overflow tokens are present in it. A higher code score indicates the query is more likely to be code related.
- **Code Search on General-Purpose Web-Search Engine**
Explored whether a general purpose search engine like Google is an optimal choice for code-related searches.
- **Semantically Complex Query Image Retrieval using Scene Graph**
We proposed a solution to the complex query image retrieval by leveraging the state-of-the-art semantic query parser and detailed image captioning in the form of the scene graph. We confirm the effectiveness of our proposed model by evaluating on Visual Genome image dataset.

- **Code Completion using Language Model and Code Template**
We proposed an approach for code completion task leveraging language model and code template. Our preliminary result shows the effectiveness of using Neural Network based language models such as RNN, LSTM, Bidirectional LSTM over n-gram based language model.
- **Finding Similar Projects on GitHub**
Leveraged Word2Vec word embedding and Word Mover Distance (WMD) to measure the similarity of GitHub projects utilizing projects' source code and available textual description.
- **Privacy in Information Retrieval**
Proposed a client-centered approach to hide user intents that prevent search engine from building user profile by injecting cover queries with the original user-issued query in a controlled way. Each user's query is submitted with a set of cover queries which carry a similar amount of information to generalize user profile while still having some personalization.

TEACHING EXPERIENCE Graduate Teaching Assistant Fall 2015 - Spring 2017
Department of Computer Science, University of Virginia
Teaching Instructor Spring 2013 to Summer 2015
Department of Computer Science and Engineering, BRAC University

WORK EXPERIENCE • Lecturer February 2013 - August 2015
Department of Computer Science and Engineering
Brac University, Dhaka, Bangladesh

AWARDS & HONORS • Student Travel Awards: ICSE/MSR 2018, NL4SE@FSE 2016, SIGIR 2016
• Enrolled in Dean List for academic excellence in B.Sc for three academic years (2008-2013).
• Department Scholarship for all eight terms for good results in B.Sc (2008-2013).
• Education Board Scholarship, H.S.C (2007) and S.S.C (2005).

TECHNICAL SKILLS Python, Java, C/C++, Apache Lucene, Solr.

ONLINE PROFILE
LinkedIn: <https://www.linkedin.com/in/masud99r>
GitHub: <https://github.com/masud99r>