Title: Inferring Concept Hierarchy from Code Snippets and Documentation

Thursday, 1/16/20 | 12pm-1pm | West Hall, W106  
Speaker: Hui Chen  
Affiliation: City University of New York Brooklyn College

Abstract: Developers face challenges in maintaining large and complex code base, and often rely on information searches to retrieve useful information, such as, the usage of an API or a code snippet for maintenance tasks in hand. We posit that concept hierarchies of code and related documentation are useful to developers’ information search. In this talk, we discuss how to extract concept hierarchies from code snippets and related documentation via probabilistic models. Stack Overflow Web posts typically consist of code snippets and related technical discussion, and the Stack Overflow dataset is therefore ideal to test inference of concept hierarchies. We evaluate the model’s predictive power by identifying automatically tag synonyms and by predicting tags for unseen Web posts. For information search, we evaluate the search results via entropy-based metrics called document diversity and specificity. The research hints that the concept hierarchy as a modeling approach can create a representation of code snippets and their technical discussions useful for developers’ information search.

Bio: Dr. Hui Chen is an Assistant Professor with the Department of Computer and Information Science, CUNY Brooklyn College and on the faculty of the Computer Science Ph.D. Program, CUNY Graduate Center. His prior appointments were with Virginia State University and Chinese Academy of Sciences. He has co-authored 70 peer-reviewed research articles in wireless networks, system security, and software analytics. He served on program committees of computing and communications technical conferences and journal editorial boards. His teaching interest and experience are in software development, systems, and networks. He is a member of ACM and a senior member of IEEE. He received a Ph.D. in Computer Science from the University of Memphis.