**Title:** Accessing the Quality of Software Development Tutorials Available on the Web

**Friday, 10/25/19 | 12pm-1pm | West Hall, W105**

**Speaker:** Manziba Akanda Nishi  
**Affiliation:** VCU

**Abstract:** Both expert and novice software developers frequently access software development resources available on the Web in order to lookup or learn new APIs, tools and techniques. Software quality is affected negatively when developers fail to find high-quality information relevant to their problem. While there is a substantial amount of freely available resources that can be accessed online, some of the available resources contain information that suffers from error proneness, copyright infringement, security concerns, and incompatible versions. Use of such toxic information can have a strong negative effect on developer’s efficacy. This dissertation focuses specifically on software tutorials, aiming to automatically evaluate the quality of such documents available on the Web. In order to achieve this goal, we present two contributions: 1) scalable detection of duplicated code snippets; 2) automatic identification of valid version ranges.

Software tutorials consist of a combination of source code snippets and natural language text. The code snippets in a tutorial can originate from different sources, perhaps carrying stringent licensing requirements or known security vulnerabilities. Developers, typically unaware of this, can reuse these code snippets in their project. First, in this thesis, we present our work on a Web-scale code clone search technique that is able to detect duplicate code snippets between large scale document and source code corpora in order to trace toxic code snippets.

As software libraries and APIs evolve over time, existing software development tutorials can become outdated. It is difficult for software developers and especially novices to determine the expected version of the software implicit in a specific tutorial in order to
decide whether the tutorial is applicable to their software development environment. To overcome this challenge, in this thesis we present a novel technique for automatic identification of the valid version range of software development tutorials on the Web.

**Bio:** Manziba Akanda Nishi is a Ph.D. student at Department of Computer Science, Virginia Commonwealth University. She is working as research assistant in Software Improvement (SWIM) Lab. Her research interest includes the area of software engineering, machine learning, natural language processing, data mining. She completed her B.Sc. and M.S. in Computer Science and Engineering from University of Dhaka.