Advanced Topic Review Abstract –
Automatically Generated Natural Language Descriptions of Code
Type – Research-oriented
Shiraj Pokharel
John Bent

CSC 8350: Advanced Software Engineering
Department of Computer Science, Georgia State University
Spring, 2017

Natural language descriptions of code are traditionally written manually, making it a time consuming, imprecise, and unstandardized practice. Additionally, the author of the description must have an intimate knowledge of the code being described and must update the document any time a change is made. There are many shortcomings of manually generated natural language descriptions of code, many of them stemming from human error. Several prominent documentation generators, such as Javadoc and Doxygen, have been created to alleviate some of these deficiencies, but these tools still require manually-written metadata. More recently, the possibility of automatically generating natural language descriptions of code has been researched with a handful of different aims, including general documentation or summarization, privacy policy generation, commit messages, bug reports, and test case documentation. Completely automated documentation generation would unburden developers from writing these documents and allow them to use their time more effectively. Another important function of automated generation is to provide a method of standardizing documentation, which would further provide a more efficient use of time for developers. This review is meant to explore the different current approaches being taken towards automatic generation of natural language descriptions, and the merits of each technique will be evaluated with respect to the applications for which it is intended.

Main References:

Other References: