Versioned Lexical Search

Letter of Intent

Michael Haufe <mlhaufe@uwm.edu>
Senior Thesis, Summer 2014
Advisor: Ethan Munson, PhD <munson@uwm.edu>

Two important tools utilized by a programmer include a Search tool capable of locating relevant information across a codebase, and a Version Control System (VCS) to manage the changes to that codebase over time. The Search tool is of even greater utility if it is aware of the syntax of the underlying codebase, referred to as a Lexical Search. While VCSs and Search tools are prevalent (to a lesser extent Lexical Search), the use and existence of a unified tool remains exceptional.

The goal of this thesis is to create a Versioned Lexical Search system in order to provide a user the capability to perform a syntactically aware search over a codebase and across time through an IDE and VCS. This will enable the ability to not only find a relevant syntactic form in the codebase, but to also see how that syntactic form changed over time as the application evolved. For example: one could search for a class named: "Foo", and issue a query to locate the earliest instance of that class in the repository and track its evolution through branches and merges. This system will be realized by leveraging and integrating a number of existing and popular tools in order to minimize development time as well as reduce potential migration taxes towards adoption. More concretely, the company Semantics Designs <https://www.semanticdesigns.com> has developed a Lexical Search Tool which supports a number of programming languages. This tool will be integrated into Subversion (VCS) to enable the desired Versioned Lexical Search capability. Subversion will also be extended to expose an API to execute the desired queries. Additionally, the Eclipse IDE will be extended in order to provide a user friendly interface to execute queries and display results.