

MASTERS FINAL ORAL EXAMINATION

**Wednesday, April 13th
12:00p.m. @ 223 Atanasoff**

Carl Chapman

Faculty Advisor: Professor Kathryn Stolee

Usage And Refactoring Studies Of Python Regular Expressions

Though regular expressions provide a powerful search technique that is baked into every major language, is incorporated into a myriad of essential tools, and has been a fundamental aspect of Computer Science since the 1960's, no one has ever formally studied how they are used in practice, or how to apply refactoring principals to improve understandability and conformance to community standards. This thesis presents the original work of studying a sample of regexes taken from Python projects mined from GitHub, determining what features are used most often, defining some categories that illuminate common use cases, and identifying areas of significance for language and tool designers. Furthermore, this thesis defines an equivalence class model used to explore comprehension of regexes, identifying the most common and most understandable representations of semantically identical regexes, suggesting several refactorings and preferred representations. Opportunities for future work include the novel and rich field of regex refactoring, semantic search of regexes, and further fundamental research into regex usage and understandability.

IOWA STATE UNIVERSITY
Department of Computer Science