

# Regex Usage in

Carl Chapman and Kathryn T. Stolee  
Department of Computer Science  
Iowa State University  
{carl1978, kstolee}@iastate.edu

**Abstract**—Regular expressions are used frequently in many programming languages for form validation, ad-hoc file searches, and simple parsing. Given their popularity, many researchers have focused on making regular expressions easier to build, understand, and use. Yet, there does not exist a study of regular expression feature usage and diversity. In this paper, we explore how often regular expressions are used, which language features are most common, and how syntactically and semantically similar regular expressions are to one another. To do this, we scraped 2 million open source Python projects from GitHub and explored the regular expressions contained within. Our results indicate that **TODO**: high level results

## I. INTRODUCTION

Regular expressions are used extensively in many programming languages, for example, to search text files [3], in form validation, and for XYZ.

## II. MOTIVATION

Bugs related to regular expressions are common, resulting in tens of thousands of bug reports [4].

## III. RELATED WORK

### A. Research on Regular Expressions

Visual debugging of regular expressions [1]

Static analysis to reduce errors in building regular expressions by using a type system to identify errors like `PatternSyntaxExceptions` and `IndexOutOfBoundsExceptions` at compile time [4].

### B. Research on Regular Expressions

Visual debugging of regular expressions [1]

### C. Research that Depends on Regular Expression Usage

Regular expressions are used as queries in a data mining framework [2]

## IV. STUDY

### RECORDING REGEX USAGES

Using GHTorrent, we found the clone urls of

## V. RESULTS

### A. CONTEXT AND CORPUS ORIGIN

1) SATURATION :  
Although

...only

## VI. DISCUSSION

## VII. CONCLUSION

## ACKNOWLEDGMENT

This work is supported in part by NSF SHF-1218265, NSF SHF-EAGER-1446932, and the Harpole-Pentair endowment at Iowa State University.

## REFERENCES

- [1] F. Beck, S. Gulan, B. Biegel, S. Baltes, and D. Weiskopf. Regviz: Visual debugging of regular expressions. In *Companion Proceedings of the 36th International Conference on Software Engineering, ICSE Companion 2014*, pages 504–507, New York, NY, USA, 2014. ACM.
- [2] A. Begel, Y. P. Khoo, and T. Zimmermann. Codebook: Discovering and exploiting relationships in software repositories. In *Proceedings of the 32nd ACM/IEEE International Conference on Software Engineering - Volume 1, ICSE '10*, pages 125–134, New York, NY, USA, 2010. ACM.
- [3] C. L. A. Clarke and G. V. Cormack. On the use of regular expressions for searching text. *ACM Trans. Program. Lang. Syst.*, 19(3):413–426, May 1997.
- [4] E. Spishak, W. Dietl, and M. D. Ernst. A type system for regular expressions. In *Proceedings of the 14th Workshop on Formal Techniques for Java-like Programs, FTfJP '12*, pages 20–26, New York, NY, USA, 2012. ACM.