

Some data from the database:
nProjectsObserved: 5001.
nRefreshRepoFailures: 3.
nGetRewinderFailures: 5.

REPRESENTING A CORPUS OF REGEXES

1. How can we choose a small list of actual regexes to best represent all regexes within a corpus by usage?
 - a.) What regexes are most frequently used?
 - b.) What features are most frequently used?
 - c.) What regexes are most cloned?
 - d.) What behavioral clustering can be observed?
 - e.) What syntactic clustering can be observed?
 - f.) Various synthesis of a-e.

CONTEXT AND CORPUS