DevSearch
Code Search Made Better

DevSearch EPFL

Code search made better

Paste your code here

Our Project

DevSearch will be a tool to help developers find relevant code based on short snippets thus providing context and examples to API and language usage. The query space will be language agnostic to leverage the full extent of open source software repositories.

Links

Wiki Page

Connect

devsearch@googlegroups.com

DevSearch - Vision

Search through millions of open source repositories

Find code using both syntax and semantics

From code to features

Variable/function/method declarations

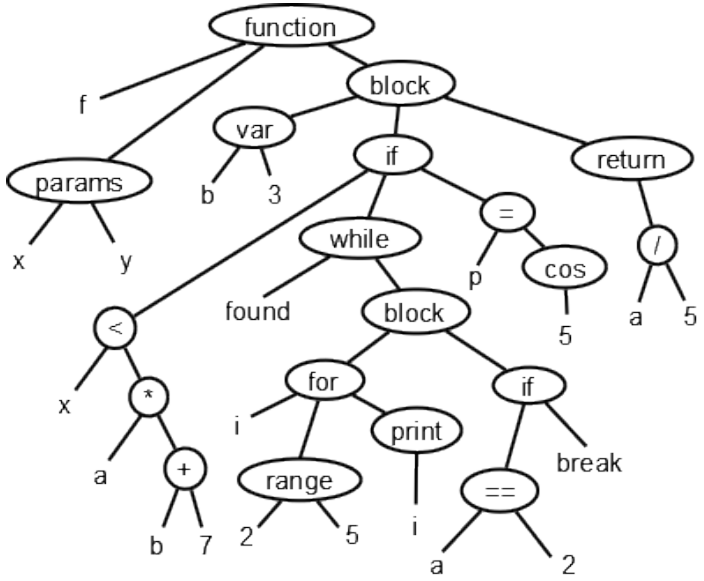
Class hierarchy

Control statements

Import statements

Semantic features

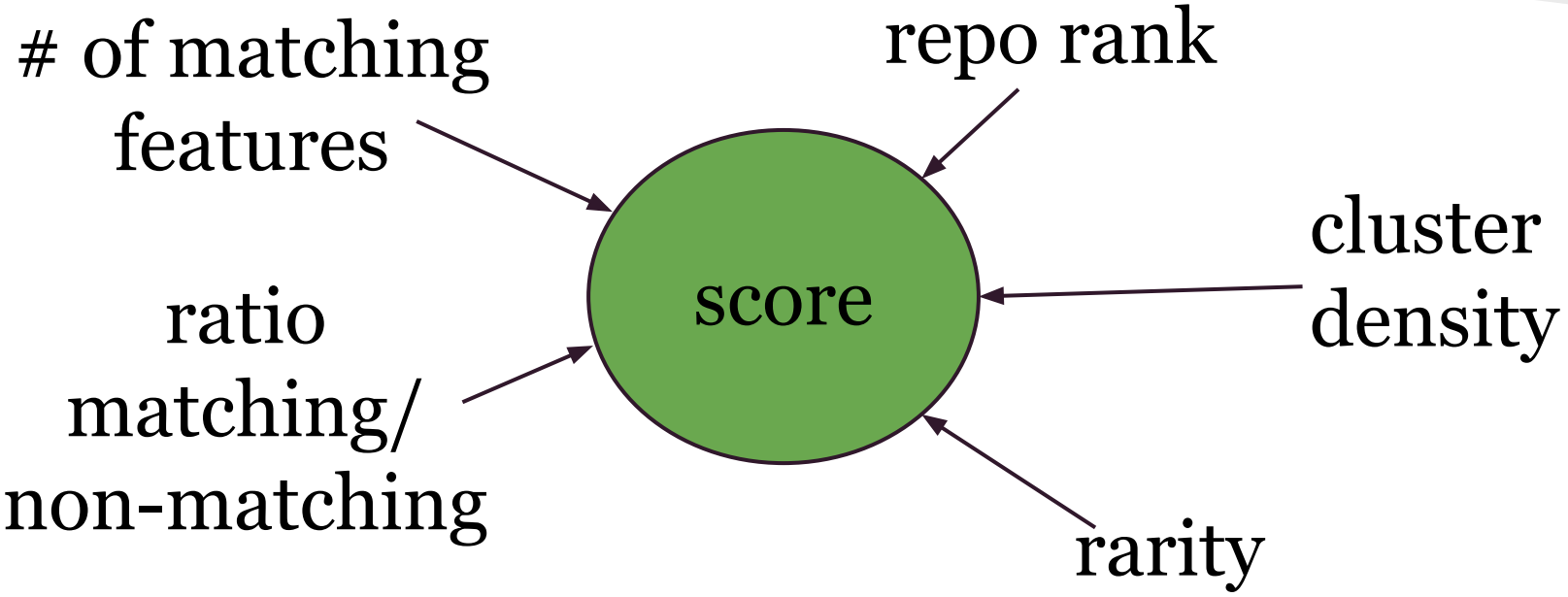
... and potentially many more



➔ language agnostic! 😊

DEMO

Score Function



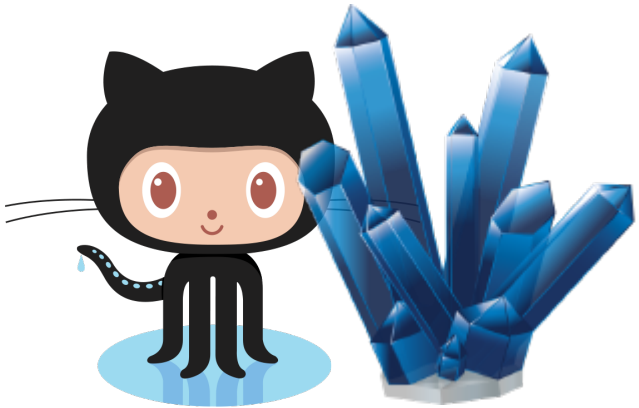
Our Data Set

Repositories: 200'000 (535 GB)

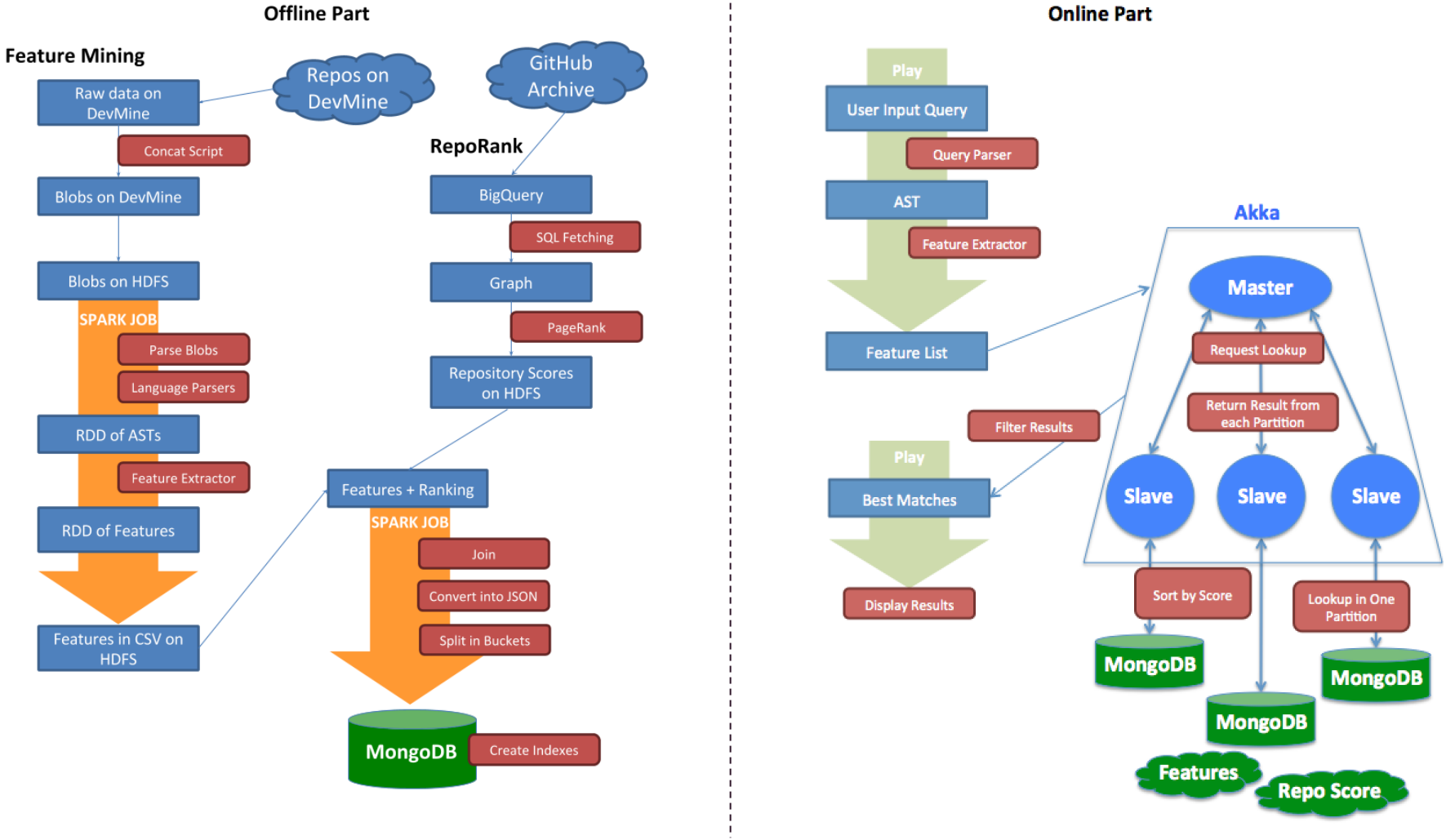
Source files: 70'000'000

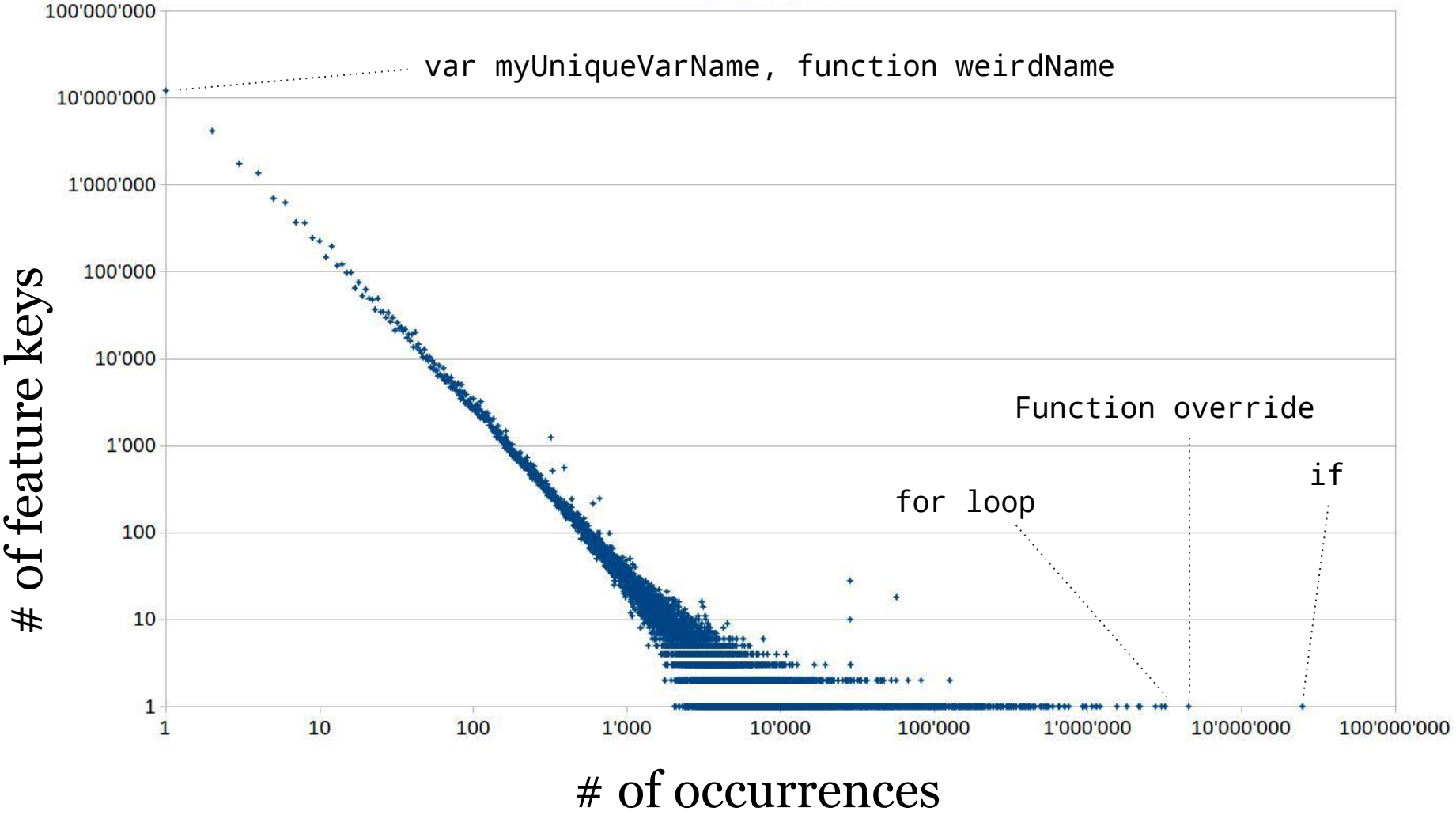
Features: 440'000'000

Users: 115'000



Technologies


Offline Part
Feature Mining
Raw data on DevMine
Repos on DevMine
GitHub Archive
Concat Script
RepoRank
Blobs on DevMine
BigQuery
SQL Fetching
Blobs on HDFS
Graph
SPARK JOB
PageRank
Parse Blobs
Repository Scores on HDFS
Language Parsers
RDD of ASTs
Features + Ranking
Feature Extractor
SPARK JOB
RDD of Features
Join
Convert into JSON
Features in CSV on HDFS
Split in Buckets
MongoDB
Create Indexes
Online Part
Play
User Input Query
Query Parser
AST
Feature Extractor
Akka
Master
Feature List
Request Lookup
Return Result from each Partition
Filter Results
Play
Slave
Slave
Slave
Best Matches
Sort by Score
Lookup in One Partition
Display Results
MongoDB
MongoDB
MongoDB
Features
Repo Score


var myUniqueVarName, function weirdName
Function override
for loop
if
100'000'000
10'000'000
1'000'000
100'000
10'000
1'000
100
10
1
1
10
100
1'000
10'000
100'000
1'000'000
10'000'000
100'000'000
# of feature keys
# of occurrences

DEMO

Thank you