

# 实验说明\_草稿2

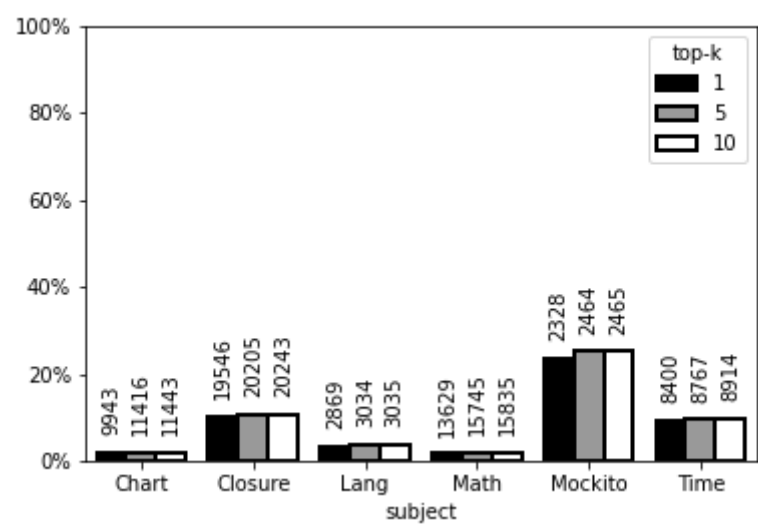
表 项目信息 [100%]

| subject | versions | LoC   | Functions | Predicates | Tests | Language |
|---------|----------|-------|-----------|------------|-------|----------|
| Chart   | 26       | 85074 | 7489      | 558336     | 1815  | Java     |
| Closure | 176      | 83721 | 9467      | 199186     | 3460  | Java     |
| Lang    | 65       | 18487 | 2220      | 84378      | 1823  | Java     |
| Math    | 106      | 50651 | 4819      | 845426     | 216   | Java     |
| Mockito | 38       | 8901  | 1317      | 9218       | 1141  | Java     |
| Time    | 27       | 27259 | 3881      | 93138      | 3917  | Java     |

注明，Math 因为时间关系仅计算 bug 相关的测试（相关由defects4j定义），各项结果求平均。

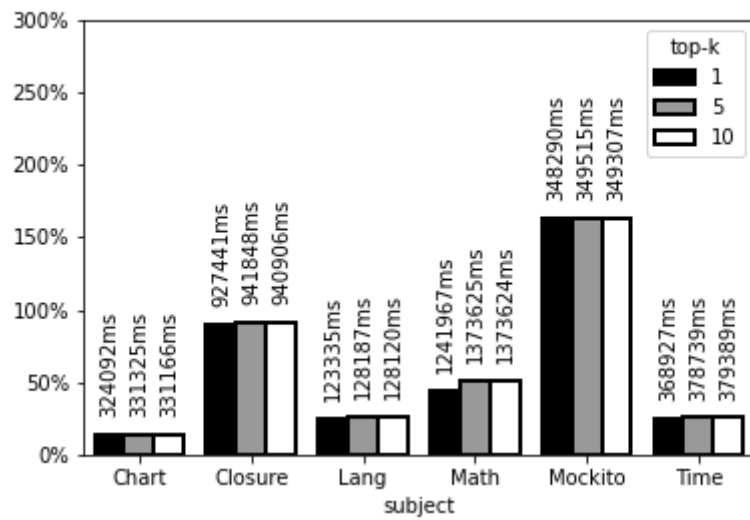
## 实验1 结果 [90%]

谓词 [100%]

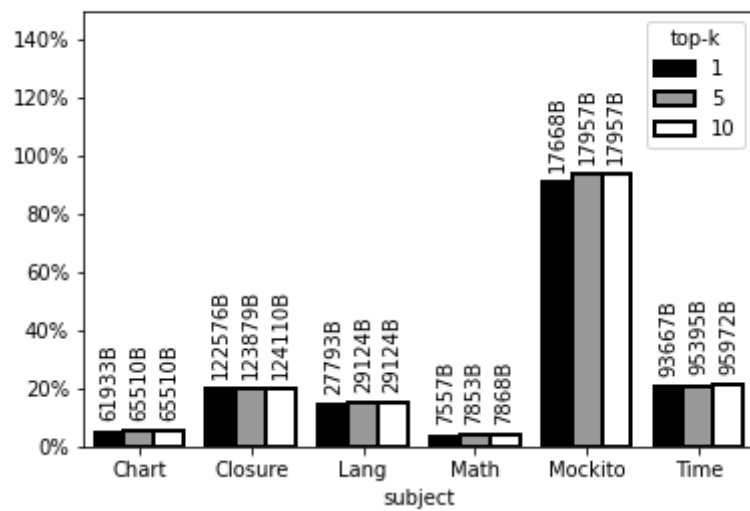


运行 [100%]

时间



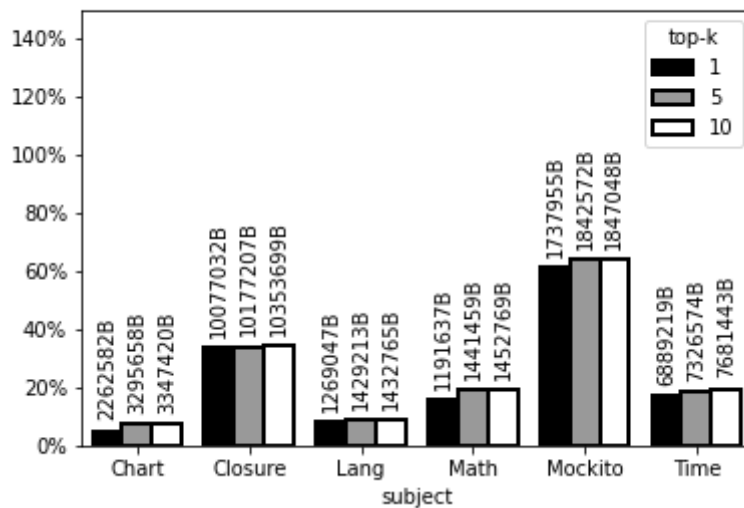
空间



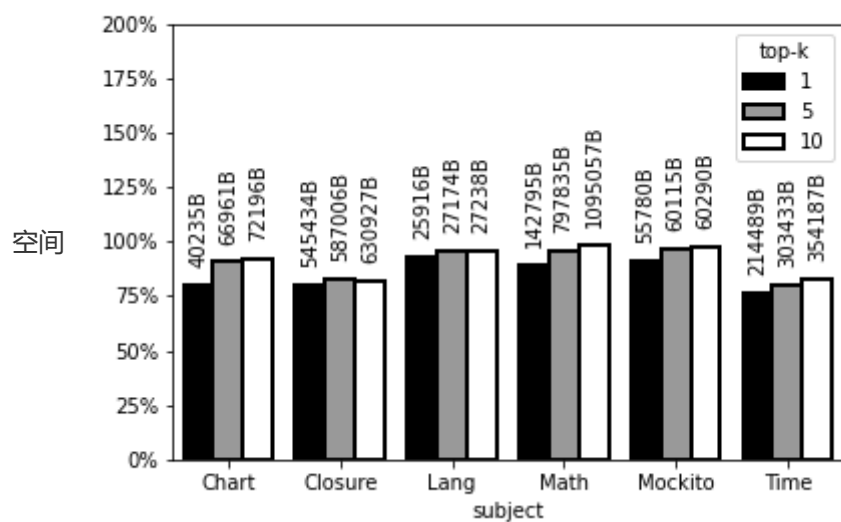
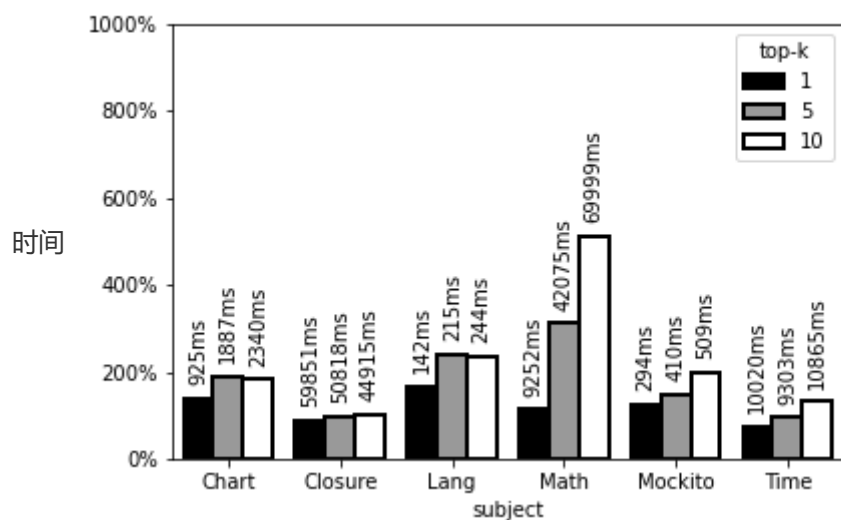
预处理 [80%]

时间

空间



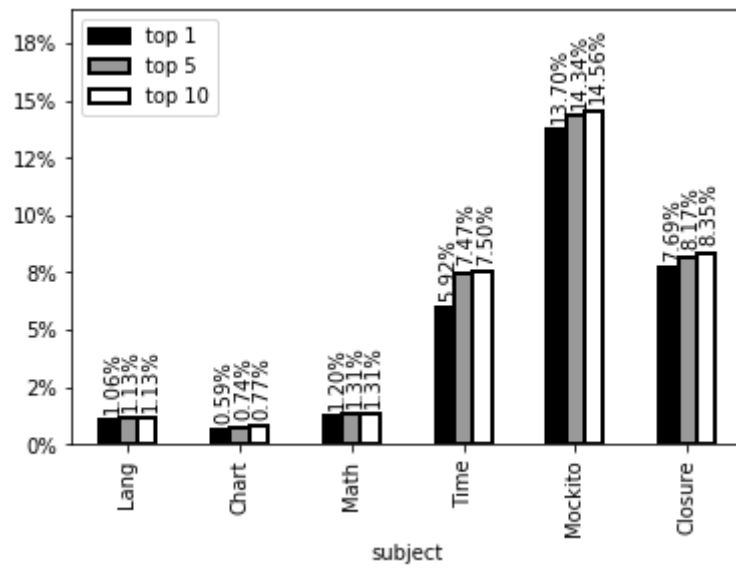
## 挖掘 [60%]



## 实验2 结果 [95%]

### RQ1 [100%]

| subject | top 1 | top 5 | top 10 |
|---------|-------|-------|--------|
| Lang    | 1.06  | 1.13  | 1.13   |
| Chart   | 0.59  | 0.74  | 0.77   |
| Math    | 1.20  | 1.31  | 1.31   |
| Time    | 5.92  | 7.47  | 7.50   |
| Mockito | 13.70 | 14.34 | 14.56  |
| Closure | 7.69  | 8.17  | 8.35   |



## RQ 2 [100%]

branch

| subject | branch_iter | branch_perd |
|---------|-------------|-------------|
| Lang    | 6.60        | 0.98        |
| Chart   | 45.64       | 1.75        |
| Math    | 18.74       | 1.39        |
| Time    | 65.22       | 8.26        |
| Mockito | 62.89       | 18.50       |
| Closure | 341.98      | 10.58       |

return

| subject | return_iter | return_perd |
|---------|-------------|-------------|
| Lang    | 5.74        | 1.33        |
| Chart   | 24.52       | 1.18        |
| Math    | 16.16       | 1.12        |
| Time    | 63.65       | 4.59        |
| Mockito | 45.57       | 14.78       |
| Closure | 282.53      | 9.43        |

scalar

| subject | scalar_iter | scalar_perd |
|---------|-------------|-------------|
| Lang    | 4.98        | 1.31        |
| Chart   | 17.32       | 0.69        |
| Math    | 17.74       | 1.37        |
| Time    | 42.70       | 8.44        |
| Mockito | 23.54       | 13.23       |
| Closure | 127.45      | 7.34        |

## RQ 3 [90%]

| subjects | s1     | s100  | s10000 | cg    | iterative |
|----------|--------|-------|--------|-------|-----------|
| Lang     | 129.98 | 7.71  | 5.53   | 4.66  | 1.35      |
| Chart    | 178.74 | 19.26 | 16.07  | 27.19 | 1.06      |
| Math     | 638.52 | 59.99 | 30.87  | 30.33 | 1.36      |
| Time     | 140.21 | 15.08 | 11.23  | 28.48 | 1.13      |
| Mockito  | 5.92   | 5.48  | 5.50   | 7.00  |           |
| Closure  | 8.37   | 5.48  | 4.99   | 6.82  |           |

## 实验环境 [80%]

系统

sample-cc 需要 Fedora 19/20 64-bit (x86\_64) mbs 需要 boost-option1.46.1 包含于 debian 系如 ubuntu1204

想完全自动化实验流程，目前不少步骤需要手工操作，比如不同环境之间交换文件。

每步操作保留原始输出文件（汇总前的文件）。

## 原实验对象结果

### EX1 [25%]

整理了原实验的脚本 [100%]

跑通流程 [5%]

收集数据 [1%]

EX2RQ1 [100%]

| 行标签        | 平均值项:top-1 | 平均值项:top-5 | 平均值项:top-10 |
|------------|------------|------------|-------------|
| apache-ant | 11.46      | 14.54      | 14.54       |
| bash       | 30.99      | 30.99      | 30.99       |
| derby      | 15.92      | 20.69      | 22.15       |
| grep       | 27.91      | 36.07      | 36.08       |
| gzip       | 32.06      | 32.20      | 35.50       |
| nanoxml    | 36.37      | 50.48      | 50.67       |
| sed        | 40.50      | 48.87      | 60.56       |
| siena      | 20.00      | 31.43      | 31.43       |
| space      | 45.30      | 49.20      | 51.98       |

EX2RQ2 [100%]

branch

| name       | branch_iter | branch_perd |
|------------|-------------|-------------|
| apache-ant | 490.75      | 21.87       |
| bash       | 239.00      | 27.96       |
| derby      | 4425.33     | 41.45       |
| grep       | 40.80       | 52.55       |
| gzip       | 32.69       | 47.20       |
| nanoxml    | 49.09       | 49.23       |
| sed        | 50.75       | 67.72       |
| siena      | 29.00       | 24.22       |
| space      | 80.06       | 69.05       |

return

| name       | return_iter | return_perd |
|------------|-------------|-------------|
| apache-ant | 385.75      | 20.63       |
| bash       | 225.67      | 23.38       |
| derby      | 1890.67     | 19.66       |
| grep       | 17.60       | 39.47       |
| gzip       | 17.23       | 36.75       |
| nanoxml    | 36.18       | 51.33       |
| sed        | 40.00       | 62.84       |
| siena      | 16.00       | 14.73       |
| space      | 58.35       | 70.22       |

scalar

| name       | scalar_iter | scalar_perd |
|------------|-------------|-------------|
| apache-ant | 190.75      | 12.30       |
| bash       | 238.17      | 31.38       |
| derby      | 1436.67     | 21.43       |
| grep       | 18.60       | 36.03       |
| gzip       | 24.54       | 32.24       |
| nanoxml    | 12.95       | 61.16       |
| sed        | 37.25       | 60.49       |
| siena      | 25.00       | 35.70       |
| space      | 58.09       | 45.54       |

## EX2RQ3 [80%]

原数据取平均

| project    | outputs | s1       | s100    | s10000  | cg      | iterative |
|------------|---------|----------|---------|---------|---------|-----------|
| apache-ant | 1410033 | 8541052  | 1803368 | 970198  | 1456418 | 871358*   |
| bash       | 111338  | 683      | 623     | 646     | 112906  | 692*      |
| derby      | 2109886 | 12084760 | 4399369 | 3381398 | 2387106 | 0*        |
| grep       | 4367    | 51601    | 37821   | 34564   | 4979    | 1749*     |
| gzip       | 13205   | 46985    | 32844   | 30784   | 13910   | 11235     |
| nanoxml    | 43570   | 71052    | 63389   | 62486   | 97128   | 55221*    |
| sed        | 2439    | 11915    | 9163    | 8601    | 2967    | 1809*     |
| space      | 7310    | 13027    | 12886   | 12353   | 8247    | 7851      |
| siena      | 14701   | 15024    | 18577   | 16242   | 20860   | 0.00*     |

原数据归一化

| project    | outputs | s1    | s100 | s10000 | cg   | iterative |
|------------|---------|-------|------|--------|------|-----------|
| apache-ant | 1.00    | 6.06  | 1.28 | 0.69   | 1.03 | 0.62*     |
| bash       | 1.00    | 0.01  | 0.01 | 0.01   | 1.01 | 0.01*     |
| derby      | 1.00    | 5.73  | 2.09 | 1.60   | 1.13 | 0.00*     |
| grep       | 1.00    | 11.82 | 8.66 | 7.92   | 1.14 | 0.40*     |
| gzip       | 1.00    | 3.56  | 2.49 | 2.33   | 1.05 | 0.85*     |
| nanoxml    | 1.00    | 1.63  | 1.45 | 1.43   | 2.23 | 1.27      |
| sed        | 1.00    | 4.89  | 3.76 | 3.53   | 1.22 | 0.74*     |
| space      | 1.00    | 1.78  | 1.76 | 1.69   | 1.13 | 1.07      |
| siena      | 1.00    | 1.02  | 1.26 | 1.10   | 1.41 | 0*        |

备注：

- iterative 列数据有问题，原脚本也不能复现，而且有代码改动，需要数据更新。
- bash 项目似乎有问题，这个就虚拟机里原来的脚本，好奇怪。
- 原实验是挑的部分版本跑的，没看出按照什么挑的。
- 汇总数据按照原始数据取平均值。
- 时间本身容易波动。

## 两次实验汇总



## 实验2RQ1 [100%]

| subject    | top 1 | top 5 | top 10 |
|------------|-------|-------|--------|
| Lang       | 1.06  | 1.13  | 1.13   |
| Chart      | 0.59  | 0.74  | 0.77   |
| Math       | 1.20  | 1.31  | 1.31   |
| Time       | 5.92  | 7.47  | 7.50   |
| Mockito    | 13.70 | 14.34 | 14.56  |
| Closure    | 7.69  | 8.17  | 8.35   |
| -          |       |       |        |
| apache-ant | 11.46 | 14.54 | 14.54  |
| bash       | 30.99 | 30.99 | 30.99  |
| derby      | 15.92 | 20.69 | 22.15  |
| grep       | 27.91 | 36.07 | 36.08  |
| gzip       | 32.06 | 32.20 | 35.50  |
| nanoxml    | 36.37 | 50.48 | 50.67  |
| sed        | 40.50 | 48.87 | 60.56  |
| siena      | 20.00 | 31.43 | 31.43  |
| space      | 45.30 | 49.20 | 51.98  |

## 实验说明

关于实验对象

关于实验算法，整个实验包含两个算法.

- 1.
- 2.

关于计时：

关于项目差异：

关于绘图样式：： `plt.imshow(x,aspect='auto')`

