

姓名：王国畅

学号：DZ1933026

软件工程导引

2019年11月17日

编译说明：

网络尚佳且github稳定的情况下：

```
cd codesim  
go get github.com/Benjamin15122/simcalc  
go get github.com/urfave/cli  
go build .
```

（如果go get安装时间过长并不是代码有错误，是go get的锅，已测试数次安装时间大约在5~20分钟）

网络不佳或github抽风的情况下：

```
cd $GOPATH/github.com  
git clone https://github.com/Benjamin15122/simcalc.git  
git clone https://github.com/urfave.git  
cd codesim  
go build .
```

补充说明：

codesim包是命令行工具，核心算法在simcalc包，Benjamin15122是本人。

由于运行时调用了shell命令获取llvm ir中间文件（没有找到合适的go包获取中间结果），在运行时会产生两个中间文件。程序中调用了rm命令清理该文件（os.Remove()不work），但可能由于权限问题无法正常清理，需要手动删除或给命令行工具授权。

代码相似度检测算法：

基本方向是needle算法，但由于golang没有适用的MaxWeightMaxFlow算法库，能力有限，最后也没实现一个可用的MaxWeightMaxFlow，故而打了大折扣：

对P1中的所有函数 f_i ，计算其与P2中 f_j 的最长公共子串 s ，计算 $s.len/f_i.len$ 作为重复率，依据 f_i 在P1中的长度占比计算重复率加权平均。