**北 京 邮 电 大 学**

**本科毕业设计（论文）中期进展情况检查表**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 学院 | | | 计算机学院 | 专业 | | 网络工程 | | 班级 | 2008211314 |
| 学生姓名 | | | 马庆元 | 学号 | | 08211543 | | 班内序号 | 6 |
| 指导教师姓名 | | | 房鸣 | 职称 | | 副教授 | | | |
| 设计（论文）题目 | | | 基于控制流分析的代码反抄袭系统 | | | | | | |
| 目前已完成任务 | 主要内容: (毕业设计（论文）进展情况，字数一般不少于1000字)   1. 已经完成了反抄袭系统的输入数据库的schema的设计。并且构建了一个从配置文件，代码库和测试数据自动生成反抄袭系统运行所需的数据包的工具。 2. 已经完成了对GCC工具GCOV(GNU覆盖率测试工具)的修改，使其可以从GCC编译的包含覆盖率测试的二进制中提取该反作弊系统工作所需的代码控制流数据。 3. 已经设计并实现了控制流化简和预处理的算法，减少控制流信息中与相似度匹配无关的各种噪音。 4. 已经设计和实现算法，将控制流地的信息量化，生成适合相似度计算的中间数据。 5. 已经设计并实现了2个控制流相似度计算算法：BASIC和ADVANCED，BASIC利用一次类似LCA的动态规划和一次简单的距离计算来确定相似度，ADVANCED引入了更多信息，利用控制流的前向分支信息2次类似LCA的动态规划确定相似度。BASIC算法较快，ADVANCED算法在复杂的控制流上有更佳的效果。 6. 已经完成了系统的自动化运行控制工具，只要一个指令就可以执行完整的运行-分析-报告过程。 | | | | | | | | |
| 是否符合任务书要求进度 **是** | | | | | | | | |
| 尚需完成的任务 | 1. 设计一个从PHP+MySQL系统中自动生成反抄袭系统输入数据库的参考程序。 2. 设计匹配算法INSANE，利用匈牙利算法代替ADVANCED算法中的第二次类LCA算法来计算相似度。 3. 将代码文本相似检测功能加入ADVANCED和INSANE算法，当控制流匹配不能提供足够证据时启用文本匹配，同时使用文本匹配信息辅助控制流匹配（当2份代码使用相同的程序设计语言时）。 4. 设计报告可视化工具，提供更准确，更详细的信息。 | | | | | | | | |
| 能否按期完成设计（论文） **是** | | | | | | | | |
| 存在问题和解决办法 | 存  在  问  题 | 1. ADVANCED算法和即将设计的INSANE算法非常慢，对于一对100行左右的代码，ADVANCED算法需要大约2秒才能给出结果，预计INSANE算法会更慢。 2. 现有的报告模块只能给出代码的相似度，不能给出具体的抄袭位置。 3. 目前的反抄袭系统不能良好的对很简单的程序给出相似度，因为逻辑过于简单，几乎所有可以达到目标的实现的确采用了非常相似的算法。基于控制流的算法不能分辨“诚信的”代码间的区别从而误判。 | | | | | | | |
| 拟  采  取  的  办  法 | 1. 优化算法的代码，如果必要，加入分布式设计，将负载分散在集群中。 2. 设计一个图形化工具，引出更详细的匹配结果，预计会采用HTML+JS来动态呈现报告。 3. 将朴素的文本相似匹配加入ADVANCED和INSANE算法，一旦检测到程序过于简单就启用这个算法，只有发现文本过于相似才给出警告。 | | | | | | | |
| 指导教师签字 | |  | | | 日期 | | 2012年4月 19 日 | | |
| 检查小组意见 | | 负责人签字： 2012年4月 日 | | | | | | | |

注：可根据长度加页；一式二份，学院、学生各一份。