**基于多克隆实例差异分析的代码模板提取工具**

摘要

1. 引言
   1. 研究背景

代码克隆是指代码库中出现了相同或相似的代码片段。这些代码片段可能是在不同的位置出现的一段代码、一个函数或一个类。代码克隆是在软件开发过程中广泛存在的一种现象，它的出现可能是由于多种因素引起的，比如开发人员的复制粘贴、使用模板或框架等。

虽然克隆代码可以提高软件系统的开发效率和重用性，但同时也带来了一些潜在的问题。首先，克隆代码带来的代码重复可能导致代码的维护和更新变得更加困难。如果克隆代码需要进行修改，那么这些修改也必须在所有的克隆代码中进行，这会增加代码的维护成本。其次，如果没有对克隆代码进行仔细的审查，导致这些代码里存在一些缺陷和错误没有被发现，复用这些代码会导致错误在整个系统中传播。再次，某些情况下对多个的克隆片段进行不同的修改，可能会产生出人意料的bug。此外，代码克隆还会降低代码可读性、影响代码演化过程等。

代码克隆分为文本相似和功能相似两种。文本相似又分为三种类型，类型一是除了注释、布局等，代码的内容基本完全相同；类型二是代码的结构基本相同，但是自定义的变量名、类型等可能有所不同；类型三是在前两种类型的基础上，可能还存在部分代码的删除和增加。类型一到三的辨别的难易程度也逐渐增加。功能相似的代码克隆指的是几个代码片段具有语义上的相似性，他们之间的代码可能不存在复用的关系，但是他们所执行的功能确实一样的，比如使用for循环和使用递归去实现阶乘。通常功能相似的代码克隆也被称为类型四的克隆。

随着软件规模的不断扩大和复杂度的增加，代码克隆问题变得日益突出。为了解决代码克隆带来的问题，越来越多的研究人员开始关注代码克隆分析。代码克隆分析旨在找出软件系统中的克隆代码，并得到有关克隆代码的信息。目前在学术界，代码克隆检测方式有很多种，比如基于Token序列匹配、抽象语法树AST、程序依赖图PDG、深度学习等。基于对克隆检测结果的分析，可以进一步提取出一些代码模板，代码模板是一种克隆代码的特殊形式，它包含了一些可变和不可变的代码片段，并且通常会被多次复制和粘贴。代码模板可以用于自动生成代码，提高软件开发的效率和质量。

* 1. 研究意义

本研究旨在提出一种基于多克隆实例差异分析的代码模板提取工具，以便在软件开发过程中自动化地提取和应用代码模板，从而提高软件开发效率和质量。同时提供代码推荐功能，可以

* 1. 研究内容

本研究主要在克隆检测工具的基础上，根据检测工具检测出的克隆方法组，提取出方法级克隆的模板。克隆方法组称为一个克隆组，而其中的每个方法都被称为一个克隆实例，本研究首先对克隆组内的克隆实例进行antlr词法分析，对

1.4 预期成果

1.5 主要工作

1.6 文章结构