CloneAnalyzer设计文档

# 程序各文件设计说明

系统前部分克隆家系提取器由慈萌完成，在此基础上实现了可视化、度量提取、克隆检索、克隆进化分析模块。

MainForm：程序主窗口；

CloneGenealogyViz：构建单个克隆家系可视化视图；

GenealogyAllViz：构建克隆家系整体可视化视图；

ExtractMetric：调用python文件进行度量值信息的提取，按钮“ExtractMetrics”触发文件编码转换和提取度量值的操作，按钮“train”触发支持向量机训练的操作，按钮“result”显示支持向量机训练的结果。（其中extract2.py为修改后的提取度量值文件，原支持向量机内容为extract1.py）。

FCMfaster：模糊C均值聚类算法并显示结果；

（FCM为原来写的代码，执行速度较慢，没有在系统中使用；KFCM为加入核函数的FCM，聚类效果不佳，没有在系统中使用）。

FragmentSettingForm：从外界载入指定代码片段检索；

CloneClassSettingForm：指定当前系统中克隆群进行代码片段检索；

MetricSetting：指定度量值范围进行代码片段检索；

Statistic：统计聚类结果文件各数值；

其余各文件为原克隆家系提取器中的文件，没有太大改动，但存在调用关系，详细介绍见慈萌的工作。

各个模块类图和时序图的详细介绍见论文第三章内容（含各个类详细介绍）。

各个模块操作流程见论文5.3系统模块的测试。

# 系统使用说

使用Nicad进行克隆代码检测，得到最原始克隆代码数据（“实验数据”-“克隆家系提取器结果数据”-“wget-results”-“blocks”文件）。

使用系统首先添加测试系统源代码（File-SourceCode按钮）和克隆代码数据（File-Import-Folder按钮）。

GenerateCRD按钮：对选中数据生成CRD文件（“CRDFiles”文件）。

GenerateForAll按钮：对所有克隆代码数据生成CRD文件。

MAP按钮：设置相互映射的文件，生成“MAPFiles文件”。

MapAll按钮：对所有相邻版本进行映射。

BuildGenealogy按钮：生成克隆家系文件（“GenealogyFiles“文件）。

PreProcess按钮：预处理，生成添加度量值的CRD文件（“emCRDFiles”文件）。

注：以上各文件可以每次执行重新生成，也可作为克隆代码数据载入。

ExtractMetrics按钮：提取度量值，生成matrix文件（生成于系统目录下，对于大型测试系统，生成时间可能较长）。

Train按钮：支持向量机训练数据。

Result按钮：显示支持向量机训练结果。

VisualizeGenealogy按钮：对选中的单个克隆家系文件生成可视化视图。

VisualizeAll按钮：全部克隆家系整体可视化视图。

Cluster按钮：利用当前matrix文件中的样本进行聚类。

Statistics按钮：统计聚类结果文件中各数值（需要动态修改统计约束条件）。

CloneRetrieval按钮：克隆代码检索。

其余按钮没有实际意义，作为预留或测试按键。