**实验6 区块链大数据分析**

# 1. 目的

1. 掌握链表、二叉树和图等结构的基本知识和使用技术。
2. 培养对问题建模和抽象的能力。
3. 培养设计和使用新工具的能力。.
4. 培养自学能力。

# 2. 问题描述

区块链（Block chain）作为一种新兴技术，在金融科技、司法存证以及数据共享等领域有着广阔的应用前景。那么区块链到底是什么呢？区块链的定义并不统一，有的说它是分布式账本，有的说它是共享可编程数据库，还有人认为它是分布式的计算环境。但追溯到“区块链”这个词本身，它的含义就是一个如图1所示的数据结构。

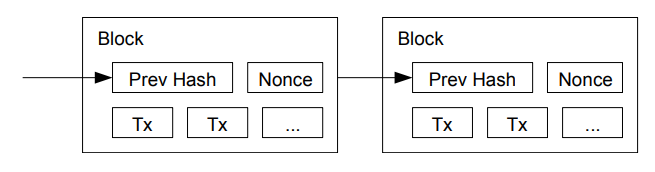


图1. 区块链数据结构示意

一个简单的区块格式如图1所示，一个区块主要由四部分组成：

1. 时间戳：记录该区块创建的时间。
2. Prev hash值：该区块的前一区块的hash值，该hash值由前一区块的内容生成，用来检验前一个区块是否被篡改。本实验中的hash值由prev hash和nonce共同生成，即：MD5(concat(prevhash, nonce))。
3. Nonce：随机数，满足一定条件的大整数，即比特币“挖矿”的成果。
4. Tx：属于该区块的交易记录，一般一个区块可以包含1000个交易左右。

每个区块的hash值相当于该区块的指纹。在上面的示例中，只要区块内容被改变，那么该区块生成的hash也会发生变化，该hash值与下一区块存放的prev hash不一致，这样就能感知到数据被篡改。以此类推，所有的区块被串起来了，形成一个链。

每条交易(Tx)的格式如下：

|  |  |  |
| --- | --- | --- |
| From | To | Amount |

From 和 To存储的是唯一的账号。每个账号是由数字和字母组成的字符串。Amount记录该笔交易的金额。每个区块中保存大约1000个交易记录。新增的区块只能添加到链尾。

请设计并实现一个区块链数据处理程序，以便对区块链中包含的大量交易记录进行分析，进而发现一些规律。

# 3. 功能要求

1. **数据初始化**

从指定文件（约100M）读入数据，初始化区块链。数据格式参见《数据集说明》。要求用尽量少的存储开销，在尽量短的时间内完成数据初始化。数据结构需记录每个区块的生成时间，**要求**一个区块中所有的交易按二叉树或B-树形式组织。

1. **数据查询**
   1. 查找指定账号在一个时间段内的所有转入或转出记录，返回总记录数，交易金额最大的前k条记录（k 为一个正整数，由查询输入）；
   2. 查询某个账号在某个时刻的金额（允许有负数）；
   3. 在某个时刻的福布斯富豪榜！输出在该时刻最有钱的前k个用户，k默认值为50，可以由用户修改k值。
2. **数据分析**
   1. 构建交易关系图。若账号A曾给B转账，则A到B有一个弧，弧上有权重，表示A给B累计转账的金额。
   2. 统计交易关系图的平均出度、入度。显示出度或入度最高的前k个帐号。
   3. 检查交易关系图中是否存在环，输出YES或NO。
   4. 给定一个账号A，求A到其他所有账号的最短路径。路径长度为路径上所有弧的权重之和。不存在路径的，不用输出。
3. **数据插入**

从文件中读入新交易记录（用户界面需支持输入文件路径），在现有的交易图上增加新的交易数据，然后重新执行功能2和功能3。

# 4. 代码检查

1. 良好的用户交互界面（方式不限，命令行/GUI/html），有着足够的提示信息。
2. 要求程序对输入错误有一定的鲁棒性。
3. 在完成每一个操作后，输出该操作的运行时间。
4. 结构清晰，变量命名合理，注释丰富。