**SJTU公司**

**立项建议书**

项目名称：基于区块链的数据共享系统

项目组组号：4

项目组负责人： 蒋海波

联系电话：18321309062

电子邮箱：jhb19940603@163.com

**2018 年 9月**

1. 项目的必要性

信息技术与经济社会的交汇融合引发了数据迅猛增长，数据已成为国家基础性战略资源，大数据正日益对全球生产、流通、分配、消费活动以及经济运行机制、社会治理方式产生重要影响。大数据既为我们的生活带来了巨大便利，打开了未来的无限可能，同时也提出了全新的挑战。高质量的数据集收集不仅费时费力，还容易“数”财两空。因此，可靠可信的数据共享平台应运而生。

以医疗行业为例，医疗行业关系到人民群众的身体健康和生老病死，与人民群众切身利益密切相关，是社会高度关注的热点。如今医疗机构都会维护一个私有的数据库来存储病人的医疗数据，这些病历是重要的研究资料，包含个人信息，伤病信息，诊断信息，用药信息等。同时，一位病人可能会在多家医疗机构就诊同一或不同的病症，他的医疗数据也就被分散存储在不同的医疗机构中。为了医疗行业的发展和改进，也为了提升医疗服务质量，医疗行业有必要对病人的健康信息进行整合研究，但想要获得这些数据似乎不那么容易。

云计算看上去是一个不错的解决方法，由云提供商来整合和管理各个节点的数据，负责这些数据的安全。不过，节点无法完全信任云提供商，因为其不受控制，也就无法保证信息的隐私安全；云提供商的误操作也会引起隐私泄露或是存储数据的永久损坏。

因此，本项目是一个基于区块链的数据共享系统，确保数据只在加入了该区块链的节点之间流通，不会涉及第三方。系统维护了一个区块链网络和共用块链分类帐，它记录了数据查询操作的详细信息，通过跟踪分类账的历史就可以捕获共享数据的流量。此外，由块链提供的智能合约机制实现了数据查询操作的统一模式和应用程序接口。API负责验证每个数据查询请求，如果成功通过验证，则将其详细信息记录在智能合约上。本项目还使用了一种安全数据查询模型，用于在认证的参与者之间共享数据，并在数据传输过程中保护数据的隐私。

为了贴近应用及项目展示，文本后续部分在基于医疗数据共享的应用背景下进行详述。

1. 项目外部条件落实情况
2. 技术基础

区块链是一种按照时间顺序将数据区块以顺序相连的方式组合成的一种链式数据结构，并以密码学方式保证的不可篡改和不可伪造的分布式账本。区块链的一个显著特点是去中心化，不依赖第三方机构。

智能合约是一种旨在以信息化方式传播、验证或执行合同的计算机协议。智能合约允许在没有第三方的情况下进行可信交易，这些交易可追踪且不可逆转。

近年来随着区块链技术的火热，越来越多的人投入到区块链技术的研究和应用，也涌现了一批成功的区块链企业项目和优秀的区块链开源项目。智能合约的出现也大大丰富了区块链的应用。智能合约应用的代表以太坊搭建了目前最流行的智能合约框架。下图为一些有代表性的区块链项目。



图2-1

本项目旨在通过区块链技术将多个医疗机构的系统和数据库进行连接，实现医疗数据安全、快速和便捷地共享，帮助医疗机构提升医疗服务质量。

1. 市场基础

随着科技的发展，数据在各行各业中发挥着越来越重要的作用，数据共享的需求也越来越多。一些对保密性要求不高的数据共享服务已经十分成熟，各行各业的各种数据共享平台层出不穷。但对于保密性要求很高的数据共享服务仍处于起步阶段，例如银行客户数据，医疗数据等，这些数据基本还处于通过点对点甚至线下的方式来进行共享，并没有一套安全、高效、易用的数据共享服务平台。但这些数据共享的需求同样很高。以医疗数据为例。2016年我国各类医疗卫生机构数为983394家，诊疗人次数为79.32亿人次。其中基层医疗卫生机构为926518家，占比94.22%，诊疗人次数为43.67亿人次，仅占55.06%；医院数为29140，占比2.96%，诊疗人次数为。95亿人次，占比41.23%。从数据可看出，平均每家医院所持有的诊疗数据是基层医疗卫生机构的24倍。

目前的普遍情况是如果一名病人有个久治不好的毛病，他去一家新医院就诊，那么除非这名病人携带了他以前在其他医院就诊的病历，否则医生只能通过询问病人或者可能需要再做一系列的检查才能了解病人的详细情况。这样的话病人在时间和金钱上的花费都会增加，而医生也要做很多重复的工作。

另一方面，对于某些病症（特别是新出现的病症），各个医院的相关医疗数据量可能都比较少，单家医院的数据量并不足以用来做例如医疗数据挖掘等方面的研究。

本项目给以上两个问题提供了一个比较好的解决方案。对于病人，只要病人去的都是通过同一区块链链接的医疗机构，那么他的医疗数据就能在这些医疗机构共享，病人不需要再带厚重的实体病历，也能节省许多不必要的检查和步骤；对于数据量少的医疗数据，通过多家医院联合提供数据，可以增加数据量，从而提高医疗数据挖掘的准确率。

1. 项目目标和创新点
2. **项目主要建设目标**

本项目基于区块链技术，构建基于区块链技术的分布式系统，实现有效地共享数据节点之间的数据的功能，并在不涉及任何第三方的情况下提供必要的隐私保护。

本项目通过搭建基于区块链的数据共享系统，可以在不涉及第三方的情况下，在数据节点之间可靠地共享节点数据库中的数据，从数据传输、阻止未授权访问等方面来确保数据的安全性、隐私性和完整性。

本项目的总体目标是：搭建基于区块链的数据共享系统，解决信息隐私保护和信息共享之间的矛盾问题，并使数据被合理安全地使用。

1. **项目创新点**

**1）本项目提出了一个新颖的数据共享系统，利用区块链技术来有效地共享数据节点之间的数据，提供必要的隐私保护，有利于当今世界敏感数据保护的推进。**

数据节点大多维护自己的数据库系统来存储和管理数据，不愿意将其数据的所有权交给第三方，这给数据的共享带来了很多困难。本项目利用区块链技术来有效地共享数据节点之间的数据，并在不涉及任何第三方的情况下提供必要的隐私保护。

**2）本项目自主提出了一个数据查询模型，通过验证请求权限，拒绝未经验证的请求，在数据传输过程中保护数据隐私。**

本项目提供了本地数据数据库及其相应的应用程序接口（API）的统一模式，来检查数据查询请求的有效性。基于统一模式和API，本项目提出了一个查询模型，用于安全地获取本项目中的共享数据。

**3）本项目搭建的数据信息共享系统和隐私保护模型，有利于保障国家信息安全和我国各行业信息技术的发展。**

众所周知，互联网技术的不断发展为我国的网络环境带来了严峻的挑战。加之我国各行业信息化如火如荼的推进，各行业信息的保护成为了发展过程中的重中之重。本项目针对数据共享系统和信息保护，在不同层面对各种类型的信息进行了分布式的安全管理。

**4）本项目提供了一套兼容性较强且高效的统一多种数据库的解决方案，用于数据节点使用的不同数据库的统一查询。**

不同数据节点搭建的数据库系统存在差异，而本项目为分布式的信息共享系统，不仅提供了统一的多种数据库的解决方案，而且将针对数据库性能进行优化。

5）**本项目的价值不仅体现于项目产品本身，后期项目的商业化应用将推动数据节点的信息共享，进而推动基于共享数据信息的研究发展。**

本项目为应用层提供了数据挖掘、特征工程等技术处理的可能，病例数据在合理合法、不侵犯隐私的前提下，可以被用于多方面研究。这样的病历数据在全面性和安全性上将得到保证。

1. 项目方案和可行性分析
2. 概要需求

本项目是一个基于区块链的数据共享系统，确保数据只在加入了该区块链的节点之间流通，不会涉及第三方。由区块链提供的智能合约机制实现了数据查询操作的统一模式和应用程序接口。API负责验证每个数据查询请求，如果成功通过验证，则将其详细信息记录在智能合约上。本项目还使用了一种安全数据查询模型，用于在认证的参与者之间共享数据，并在数据传输过程中保护数据的隐私。

1. 项目技术架构

图4-1为本项目的技术架构图。

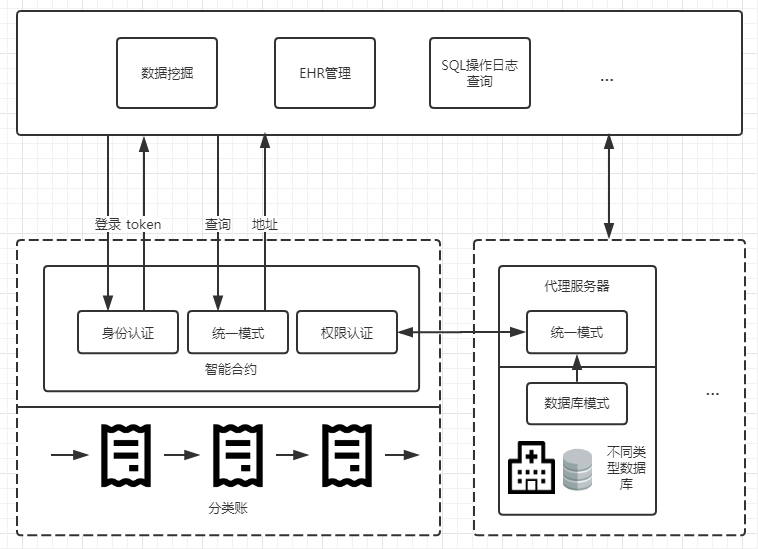


图4-1

本系统的结构分为三层：分布式数据层、数据采集层和应用层。本次项目开发的重点集中在分布式数据层和数据采集层。项目的执行周期内，除项目的基础开发外，还将对项目系统进行功能上的完善和性能上的优化。

1）分布式数据层

分布式数据层负责管理共享的医疗数据数据库，并处理数据查询请求。具体而言，分布式数据层建立在区块链网络上，由该节点的医疗机构维护。该层由数据管理层和数据存储层两部分组成。数据管理层包含一个智能合约机制和一个区块链分类帐，负责管理有关医疗数据查询操作的接口。医疗机构可以通过使用该接口，来请求远程的医疗数据数据库的访问，同时，在请求之前，数据管理层负责检查医疗机构的权限，来识别查询请求的有效性。

2）数据采集层

数据采集层用于采集共享的医疗数据。数据采集层包含了一个主要部件，称为医疗数据聚合器。当用户初始化了一组特定的医疗数据属性的查询请求，医疗数据聚合器会尝试在智能合约中注册该请求，来获得包含所需医疗数据属性的医疗数据数据库的访问地址。然后，医疗数据聚合器将通过建立连接来从远程医疗数据数据库获得数据。接收到数据后，该层就将数据集本地合并为结果数据集，并将其返回到应用层。

3）应用层

应用层是利用数据挖掘、差异隐私保护、特征工程等技术处理接收到的电子健康档案，不泄露病人信息的隐私。

1. 开发手段

|  |  |
| --- | --- |
| 开发方法 | 面向对象方法 |
| 建模工具 | Processon |
| 编程语言 | Solidity，javascript，  Java |
| 编程工具和框架 | 以太坊web3.js库，nodejs，npm |

1. 可行性分析
2. 技术

虽然区块链技术还不成熟，但已经有一些成功的商业和开源项目出现，可以用于借鉴。而且本项目所使用的主要技术之一——智能合约，以太坊已经提供了一个完整的javasrcipt库来支持与智能合约交互。

1. 市场

数据在各行各业发挥着越来越重要的作用，如何获取大量可靠的数据是关键。数据共享是当前最为有效的方式，而中心化的数据共享平台是目前较为常见的一种实现方式。但是中心化的数据共享平台无法满足保密性要求高的数据共享需求。以医疗数据的共享为例，目前国内医疗数据共享的发展还比较缓慢，尽管有一些国家推动的医疗数据共享大平台，但还处于起步阶段，并且中心化的平台服务模式不能满足一些医疗机构特别是私立医疗机构对病人信息安全性和保密性的要求。根据上面第1到4章的说明，本项目由于拥有去中心化的优势，能较好的解决这些医疗机构的顾虑。

1. 优势

以医疗数据的共享为例，由于区块链去中心化的特效，医疗数据由医疗机构自己维护，只分享给同一区块链的合作医疗机构，能大大提高医疗数据的安全性。对于系统可靠性也有较大提升，中心化的平台一旦发生故障，所有的医疗机构都获取不到数据。而如果区块链中的某一家医疗机构的系统发生故障，其他医疗机构除了获取不到这家医疗机构的数据，其他不受影响。

1. 劣势

以医疗数据的共享为例，本系统的部署对于医疗机构自身的系统及机器性能有一定的要求：医疗机构的系统需要进行一定的改动以支持数据的共享和查询；对于医疗机构本身维护的数据库的访问量会远比以前要高，医疗机构自身的维护成本可能会增加。

1. 计划进度

|  |  |  |  |
| --- | --- | --- | --- |
| 迭代 | 时间 | 任务 | 风险 |
| 1 | 10.8-10.21 | 1.完成概要需求分析，撰写文档  2.学习并搭建基于区块链的分布式数据库  3.测试区块链网络 | 没有搭建区块链网络的经验，难度较大 |
| 2 | 10.22-11.4 | 1.细化数据共享与获取需求，撰写文档  2.实现数据共享，数据获取，提供基础的可视化界面  3.测试功能和界面 | 无法按进度实现功能 |
| 3 | 11.5-11.18 | 1.细化权限认证与数据处理需求，撰写文档  2.实现用户权限认证，数据清洗、特征工程  3.测试用户权限认证功能，评估数据模型 | 数据结构复杂，难以整合 |
| 4 | 11.18-12.2 | 1.集成各模块  2.细化需求，整理文档  3.学习部署Hadoop分布式数据库 | 无法按进度实现功能 |
| 5 | 12.3-12.16 | 1.拆分数据，部署Hadoop分布式数据库  2.设计性能优化方案并评审  3.实现性能优化方案  4.进行性能对比实验并继续优化 | 技术难度较大  无法优化到理想状态 |
| 6 | 12.17-1.9 | 1.评估性能对比结果  2.部署数据挖掘应用  3.进行数据挖掘  4.撰写相应文档 | 技术难度较大 |

1. 项目预期成果

本项目的预期成果为：

1. 立项建议书
2. 6份迭代计划
3. 6份迭代报告
4. 软件需求规约文档
5. 软件架构文档
6. 源代码
7. 5份软件测试计划
8. 5份软件测试报告
9. 区块链部署手册
10. 项目总结报告
11. 项目社会经济效益

本项目是基于区块链的数据共享系统，区块链去中心化的优点能保证每个节点的数据能在共享时，不被第三方泄露，以此维护链上每一节点的利益和数据的隐私。不单单是医疗行业，本项目能应用于多个行业领域，与多产业进行融合，带来大数据创新应用，数据挖掘用以研究分析数据，指导行业发展前景，提升行业效率，减少并预防风险，更好的服务于人民的物质和精神生活。

就医疗行业而言，以2016年的统计结果为例，医疗卫生机构诊疗人次达79.32亿次；医疗卫生机构观察室留观病例数达5077.00万人；医疗卫生机构门诊健康检查人数45290.13万人。其中产生了大量数据，能为今后医师们的诊断或研究提供强有力的支持，提升医疗行业的服务质量。同时，据国家统计局提供的数据，在全国总共983394个医疗机构中，基层医疗卫生机构有926518个，占比94.2%。接入共享医疗数据系统后，他们一边能获得更高级专业的数据知识，一遍为大型医疗机构提供源源不断的数据来源，一同为改进医疗质量贡献力量。

|  |
| --- |
| 导师意见    签章：  年 月 日 |
| 授课教师意见：  签章：  年 月 日 |