**区块链数据共享系统中高效搜索技术研究**

**作者姓名 周恩愿**

**指导教师姓名、职称 刘雪峰 副教授**

**申请学位类别 工学硕士**

**区块链数据共享系统中高效搜索技术研究**

**作者姓名：**周恩愿

**一级学科：**网络空间安全

**二级学科（研究方向）：**选择学科

**学位类别：**工学硕士

**指导教师姓名、职称：**刘雪峰 副教授

**学　　院：**网络与信息安全学院

**提交日期：**选取日期

**西安电子科技大学**

**硕士学位论文**

**学　号　 18151213094**

**密　级　 公开**

**学校代码 10701**

**分类号** TN82

By

Zhang San

Supervisor: Li Si Title: Professor

February 2015

A thesis submitted to

XIDIAN UNIVERSITY

in partial fulfillment of the requirements

for the degree of selecting one

in selecting one

**Thesis/Dissertation Guide for Postgraduates**

**of XIDIAN UNIVERSITY**

**西安电子科技大学**

**学位论文独创性（或创新性）声明**

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名： 日 期：

**西安电子科技大学**

**关于论文使用授权的说明**

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在 年解密后适用本授权书。

本人签名： 导师签名：

日 期： 日 期：

摘要

随着信息技术的发展，网络中存储的数据越来越多，但数据对个人的价值有限，只有通过数据共享才能更好地利用数据的价值。由于隐私数据和单点失效的存在，中心化托管的数据共享技术很难得到数据所有者的信任，因此这就需要一种可靠的分布式基础设施来保证数据的安全共享。区块链技术作为可信的去中心化数据库已经在金融、供应链、医疗等越来越多的领域被广泛使用，大量有价值的数据被存放到区块链上，用户可以通过区块链来安全地共享数据。而随着区块链中数据的增加，对链上数据的搜索需求已经逐渐体现了出来，但现存的区块链系统并不支持对数据的快速检索。造成这种现象的主要原因是区块链系统在最初设计时没有考虑到丰富的查询需求，区块链底层的键值对数据库只支持快速写入而不支持快速检索，读取效率低已经成为限制区块链搜索功能发展的瓶颈。同时，只能追加的数据结构也使得如果想要完全检索必须从尾至头遍历整条区块链。

为了解决区块链数据共享系统中检索效率的问题，本文从区块链底层数据结构和存储模式出发，经过了大量的调研工作和基础理论研究后，提出并实现了一种在区块链数据共享系统中进行高效检索的技术方案。本文首先分析了区块链数据共享系统中进行快速检索存在的挑战，针对这些挑战探索相关的传统检索技术并思考其与区块链系统结合的可能性。具体的设计内容上，本方案创新性地将区块链的存储模式由单一的链上存储转变为了链上链下混合存储模式，并设计了一种分布式语义提取算法建立链上数据和链下数据的映射关系。单链检索部分，本方案在链上构建了一个名为默克尔语义搜索树的索引结构，通过结合了merkle树、B+树和帕特丽夏树等结构，提供了多关键字查询、范围查询、模糊查询等功能。在跨链检索部分，本方案首先对不同链上的交易格式进行整合，然后在每条链上抽取跨链委员会组成知识链，委员会中的节点们通过跨链共识对来自不同链上的实体三元组进行处理，构建一个跨链的可视化全局知识图谱，提供跨链检索服务。我们还同时为区块链中的轻节点提供了代理查询和结果验证的功能，轻节点需要连接全节点来进行代理查询，全节点在返回查询结果的同时需要提供对查询结果的验证证明，这样轻节点在自己本地即可对查询结果进行验证。

最后，根据上述整体架构，本文在以太坊平台上构建了数据共享系统中的高效检索系统，通过多个不同领域的数据集测试了我们方案的可行性，并与其他相关工作进行了整体性能对比测试。实验结果表明，本文方案在检索响应时间、检索准确率和召回率、检索相关度和安全性等方面具有较好的性能表现，对于数据共享系统中高效检索的研究具有重要意义。

**关 键 词**：区块链，数据共享，分布式检索，知识图谱

ABSTRACT

The Abstract is a brief description of a thesis or dissertation without notes or comments. It represents concisely the research purpose, content, method, result and conclusion of the thesis or dissertation with emphasis on its innovative findings and perspectives. The Abstract Part consists of both the Chinese abstract and the English abstract. The Chinese abstract should have the length of approximately 1000 Chinese characters for a master thesis and 1500 for a Ph.D. dissertation. The English abstract should be consistent with the Chinese one in content. The keywords of a thesis or dissertation should be listed below the main body of the abstract, separated by commas and a space. The number of the keywords is typically 3 to 5.

The format of the Chinese Abstract is what follows: Song Ti, Small 4, justified, 2 characters indented in the first line, line spacing at a fixed value of 20 pounds, and paragraph spacing section at 0 pound.

The format of the English Abstract is what follows: Times New Roman, Small 4, justified, not indented in the first line, line spacing at a fixed value of 20 pounds, and paragraph spacing section at 0 pound with a blank line between paragraphs.

**Keywords**: XXX, XXX, XXX, XXX, XXX

插图索引

图序号 插图示例 X

表格索引

表格序号 表格示例 X

符号对照表

符号 符号名称

XXX XXX

XXX XXX

XXX XXX

缩略语对照表

缩略语 英文全称 中文对照

XXX XXX XXX

XXX XXX XXX

XXX XXX XXX

目录

[摘要 I](#_Toc66179738)

[ABSTRACT III](#_Toc66179739)

[插图索引 V](#_Toc66179740)

[表格索引 VII](#_Toc66179741)

[符号对照表 IX](#_Toc66179742)

[缩略语对照表 XI](#_Toc66179743)

[第一章 绪论 1](#_Toc66179744)

[1.1 研究背景与意义 1](#_Toc66179745)

[1.2 国内外研究现状 2](#_Toc66179746)

[1.3 本文研究内容及主要贡献 4](#_Toc66179747)

[1.4 本文内容组织结构 5](#_Toc66179748)

[第二章 相关理论与技术 7](#_Toc66179749)

[2.1 区块链系统和结构 7](#_Toc66179750)

[2.1.1 区块链存储系统 7](#_Toc66179751)

[2.1.2 区块链数据结构 7](#_Toc66179752)

[2.1.3 区块链查询技术 8](#_Toc66179753)

[2.1.4 跨链技术 9](#_Toc66179754)

[2.2 信息检索技术 10](#_Toc66179755)

[2.2.1 TF-IDF算法 10](#_Toc66179756)

[2.2.2 倒排索引 11](#_Toc66179757)

[2.2.3 检索模型分类 11](#_Toc66179758)

[2.2.4 B+树 12](#_Toc66179759)

[2.2.5 pagerank算法 13](#_Toc66179760)

[2.2.6 RDF数据模型 14](#_Toc66179761)

[2.2.7 知识图谱构建技术 14](#_Toc66179762)

[2.3 本章小结 15](#_Toc66179763)

[2.4 题名页 16](#_Toc66179764)

[2.5 声明 17](#_Toc66179765)

[2.6 摘要 17](#_Toc66179766)

[2.7 插图索引 18](#_Toc66179767)

[2.8 表格索引 18](#_Toc66179768)

[2.9 符号对照表 18](#_Toc66179769)

[2.10 缩略语对照表 18](#_Toc66179770)

[2.11 目录 18](#_Toc66179771)

[2.12 正文 19](#_Toc66179772)

[2.12.1 绪论 19](#_Toc66179773)

[2.12.2 各章节 19](#_Toc66179774)

[2.12.3 结论 20](#_Toc66179775)

[2.13 参考文献 20](#_Toc66179776)

[2.14 致谢 21](#_Toc66179777)

[2.15 作者简介 22](#_Toc66179778)

[2.16 其他 22](#_Toc66179779)

[第三章 研究生学位论文的编辑、打印、装订要求 23](#_Toc66179780)

[3.1 学位论文封面的编辑和打印要求 23](#_Toc66179781)

[3.2 学位论文的版面设置要求 23](#_Toc66179782)

[3.3 学位论文的打印、装订要求 23](#_Toc66179783)

[3.4 其他说明 23](#_Toc66179784)

[第四章 图、表、公式示例 25](#_Toc66179785)

[参考文献 27](#_Toc66179786)

[致谢 31](#_Toc66179787)

[作者简介 33](#_Toc66179788)

# 绪论

## 研究背景与意义

随着社会和科技的发展，越来越多的人参与到网络空间活动中，导致了互联网中存储的数据量呈现出了指数级别的增长，而由于数据的所有者不同和存储方式的多样化，形成了大量的信息壁垒和数据孤岛。为了打通不同数据所有者之间的壁垒，学术界和工业界从各方面开展了对于海量数据共享的研究，想要找到一种安全高效的数据共享模式，但仍存在一些问题。

最近几年，随着比特币和以太坊等加密货币系统的成功，区块链技术作为其核心技术受到了越来越广泛的关注。作为一种分布式账本技术，区块链具有去中心化、不可篡改、安全可信、数据公开透明等特点，目前在供应链、农产品溯源、金融基础设置、版权存证等领域已经有了相关应用。

区块链作为一项分布式基础设施，有一个新的应用领域，分布式数据共享[1]。分布式数据共享在之前已经有了相关的尝试，例如Bittorrent协议和ipfs协议[2]，但是，缺乏一种稳定的激励机制，使得这种文件共享系统健康地运转下去，同时，多方写入的场景，也可能导致拜占庭问题的产生。利用区块链的特性，可以有效地解决当前分布式数据共享系统中存在的问题，但区块链本身仍存在数据搜索效率低、不支持复杂搜索等功能。

比特币白皮书中描述的区块链是一种按照时间顺序将区块以链表的形式连接起来的数据结构[3]，通过密码学技术保证了交易数据的不可篡改性，数据需要经过系统中多数节点的共识，才能够以追加的方式添加到区块链的尾部，如果恶意节点想要篡改已经经过共识并存储在区块链内的数据，需要掌握系统中的大部分算力，这在现实中往往是难以实现的，因此通过区块链技术可以解决分布式场景中的信任问题。

在区块链系统中，存储的数据并不是底层的原始数据，底层数据往往是带有一定格式的交易数据，这些数据一般不会以明文的形式存储到区块链上，需要经过散列运算，将得到的哈希值上链，比如比特币采用双sha256散列函数，可以将任意长度的原始交易记录存进区块。在保证安全性的同时，区块链这种只能追加的数据结构也导致了系统性能的下降，大多数区块链系统底层采用的是key-value数据库，通过牺牲读性能来换取写性能的提高[4]。然而，在实际应用中，区块链系统的交易吞吐量往往没有那么大，比特币的交易量只有每秒1笔左右，以太坊的交易写入量为每秒10笔左右[5]，而随着区块链中数据的增加，对链上数据的搜索需求已经逐渐体现了出来，key-value数据库的快速写入优势无法体现，而读取效率低已经成为限制区块链搜索功能发展的瓶颈。

现有的区块链系统在最初设计时，只是针对某种特定事务进行处理，没有考虑丰富查询的需求，数据要么存储在键值对中，要么存储在平面文件中，而这些存储方式都不支持复杂的查询操作。因此，如果实现了区块链上的快速、多关键词的搜索，用户就可以对分布式数据共享系统中的数据进行快速、个性化的搜索，可以更快地找到满足自己需求的数据位置，系统的可用性更强。同时可以利用智能合约对链上资产进行自动分配，系统的安全性更强。

除了区块链系统中存在的检索困难问题之外，另一个需要考虑的问题是不同区块链之间的跨链检索问题。由于不同的区块链系统是异构的，他们底层的数据结构和共识机制都不相同，这就造成了区块链与区块链之间的另一种“数据孤岛”。

跨异构链的信息检索问题可以归为两点：统一数据格式和建立信息关联。从搜索技术的发展史中可以了解到，知识图谱的两大关键技术，RDF（资源描述框架）和语义网（Semantic Web）很好地描述了非结构化数据之间的逻辑关系，RDF通过一种通用的三元组数据格式对非结构化数据进行简洁统一便于移植的形式化描述，语义网则以RDF数据格式为基础构建了一个能够描述知识实体或概念本身以及他们相互之间逻辑关系的智能网络。在这两种技术的基础之上构建的知识图谱，存储的全部都是统一格式的结构化数据，使不同来源的数据之间形成一种结构化状态，用户只需要通过这些关联关系，就可以快速搜索到自己需要的信息。因此，如果可以通过RDF和语义网这两种技术在不同结构的区块链系统中对数据进行统一的格式整合，并在相关数据间建立语义关联关系，就可以构建一个跨异构区块链的“知识图谱”，用户在使用该跨链检索系统进行检索操作时，不需要考虑底层的区块链结构，只需要关注搜索的内容本身，就可以得到自己需要的信息。

## 国内外研究现状

目前国内外对于区块链系统中的搜索模块的研究还处在起步阶段，主要的处理方法是对区块链存储结构的优化及对区块链索引系统的优化。绝大多数成熟的区块链系统采用的是以LevelDB为代表的key-value分布式数据库。比特币的早期版本采用BerkleyDB[6],之后改为LevelDB。以太坊采用的是LevelDB, Hyperledger Fabric在1.0版本更新后给用户提供了LevelDB和couchDB两种数据库可供用户选择[7]。LevelDB等key-value数据库提供的查询功能较为简单，因此区块链系统的设计者需要在LevelDB数据库的上层封装一个查询层，以提供基本的查询功能。

以太坊在将内存中的数据持久化到磁盘时，将格式化的交易数据通过递归长度前缀编码（RLP编码）转化成字节数组[8]，然后计算转化后字节数组的哈希值，最后将哈希值和编码后的交易数据以<Key,Value>对的形式存到LevelDB数据库中。进行查询操作时，通过解析用户提交的Key值来找到对应的Value字节数组，并将解码后的数据返回给用户。由于以太坊在最初设计之时没有考虑复杂的分析查询，因此没有在系统中加入索引结构，导致系统无法执行关系型查询，同时无法执行Top-K查询，K-NN查询等复杂的查询。

BigchainDB[9]是一种具有区块链特性的分布式数据库结构，底层数据库使用了mongoDB，并且设计了一套丰富的查询语言。但在严格意义上来说，BigchainDB本质上并不是一个传统的区块链系统，同时并不具有区块链的账本全复制特性，只是利用区块链的特性来保护数据的安全性。

Anh Dinh等人提出了一种名为Blockbench[10]的私有区块链评价框架,尝试通过用支持丰富查询功能的存储系统来替换传统区块链的LevelDB数据库。

Li等人提出了EtherQL[11],主要思想是将区块链中的数据拷贝到外部的数据库中作为区块链的搜索层，并在这一查询层上实现了丰富的搜索功能，可以支持Top-K查询、范围查询等复杂的搜索的功能。EtherQL方案不需要对原有区块链系统进行修改，是一种可以兼容多种成熟区块链系统的方案。

BlockchainDB[12]是一种构建在区块链之上的分片数据库系统，主要适用的场景是互不信任的多方实现数据读写的场景，设计了一种将区块链作为数据存储层，并将存储节点进行分片管理的系统。BlockchainDB提供了易用的查询接口和强大的数据一致性保证，提高了区块链存储系统的性能，并降低了使用的复杂度。

外联数据库解决方案的思想是为已有的区块链系统设计外部查询层，另一种对区块链数据进行快速搜索的方法是在区块链系统中内置索引结构。山东大学的李庆忠教授团队提出了一种支持快速搜索的区块链教育证书系统（ECBC）[13]，该方案扩展了以太坊系统中的MPT结构，设计了MPT-Chain索引结构，将学生的学历证书按照相同前缀进行字典存储，利用链上索引结构加快了对学生学历证书的搜索效率。 Ruan等人在2019年提出了LineageChain[14]，通过在区块链上建立了Merkle DAG结构和一个跳表结构，实现了对历史交易数据的快速溯源，并通过智能合约实现了用户存储和搜索的API。

vChain[15]实现了另一种区块链上的搜索方式，思路是将查询工作委托给全节点进行，全节点可通过爬虫或者本地建立数据库副本，从而进行快速查询，同时在区块中加入一种认证数据结构（ADS），全节点需要返回一个查询数据的加密证明，用户通过验证证明可以验证查询结果的完整性。优势是既保证了快速查询，同时又可以保证查询结果的正确性和完整性。但普通用户需要委托全节点进行查询，同时加密证明的验证也需要耗费时间。

华东师范大学的金澈清教授团队提出了“师大链”[16]，将链上数据与链下数据联系起来，为区块链上的交易数据赋予丰富的语义信息，可以对区块数据进行类SQL查询，同时加入了RDBMS对链上和链下的数据进行管理。

经过几年的研究，区块链的跨链技术已经有了比较成熟的解决方案，目前主流的技术有公证人机制，侧链/中继模式和哈希锁定等方式[31]，可以用于实现区块链之间的资产互操作，解决价值孤岛的问题。但是这些跨链方案只关注了跨链的资产交换问题，没有考虑到数据的交互问题，虽然解决的价值孤岛的问题，但数据孤岛和信息壁垒仍然存在，下一步需要考虑的是如何对不同链上的数据进行链接，解决跨链检索的问题。武汉科技大学的任惠[32]等人提出了一种基于知识图谱的区块链语义知识共享机制，设计并实现了一种去中心化的全局知识查询方案，虽然没有提及“跨链”的概念，但为跨异构链的数据检索方案提供了参考。

## 本文研究内容及主要贡献

为了解决目前区块链系统中存在的检索效率差，不支持复杂搜索的问题，本毕业设计重点分析现有区块链系统存储结构和索引方式中存在的问题，研究一种在区块链系统中进行快速搜索的方法，在实现对区块链系统中存储的数据进行快速搜索的同时支持较为复杂的分析查询，同时保证查询结果的完整性和可靠性。同时，针对跨区块链信息检索存在的问题，本毕业设计在单链检索的研究基础上研究对多条异构区块链进行信息检索的方法，提出了一种支持多链的信息检索方案，并以知识图谱的形式呈现给用户。

本文的主要内容和贡献如下：

1. 为了建立链下数据和链上数据的协同映射关系,首先对链下信息的语义信息进行提取，提出了一种改进的TF-IDF模型，称为去中心化语义提取算法,通过引入去中心化词频因子、词性因子和位置因子来提升语义特征提取的准确性，解决了传统关键字提取算法中缺乏语义理解的特点,并基于此构建了一个关键字的语义倒排索引,为链上索引结构的构建提供支撑；

2. 在链上构建一个名为默克尔语义搜索树（MST）的索引结构，结合了默克尔树，帕特丽夏树，B+树，哈希指针，AC自动机等技术，利用该索引结构可以对之前提取的链下数据的语义信息和元数据进行多功能的复杂搜索操作，包括多关键字搜索，范围搜索，模糊查询等。由于区块链的追加性质，MST的结构是动态构建的，随着创世区块的生成而生成，并随着新区快的生成而更新，整个操作由所有矿工通过共识机制来完成。针对不同的搜索功能采用了不同的技术手段实现，对于多关键字搜索，通过帕特丽夏树的前缀匹配和压缩特性来实现；B+树用于实现范围搜索功能，针对模糊查询，设计并实现了一个基于AC自动机的模糊态多关键字匹配算法；

3. 为了使得不包含所有账本信息的轻节点（瘦客户端）也可以对系统中的数据进行搜索和验证，设计一种代理查询的功能，即拥有所有账本信息的全节点为轻节点做代理查询和提供验证的操作。首先由轻节点发起查询请求，全节点接收到查询请求后在本地进行查询，然后将得到的查询结果和该查询结果在MST上对应的merkle proof一起返回给轻节点，轻节点收到查询结果和merkle proof后可以通过本地的区块链头部信息对查询结果进行验证。

4．为了提供跨链检索的功能，本文提出了一种基于RDF的去中心化知识抽取方案，可以将异构区块链中不同结构的数据抽取并整合成为实体三元组的形式；在不同的区块链间构建了一个全局的跨链索引结构，并设计了配套的索引生成共识方案；提供了跨链数据查询的通用接口，并在此基础上开发了可视化的知识图谱结构，用户无需考虑底层具体的区块链结构，只需要输入对应的搜索需求，就可以查询到相关的结果。

## 本文内容组织结构

本文章节共分五章，组织结构如下：

第一章，介绍研究的背景和意义，对国内外的研究现状进行了整理和分析，针对当前的研究现状分析了在区块链数据共享系统中进行高效检索存在的问题和挑战，最后介绍了本文的主要研究内容及主要贡献。

第二章，介绍了本文研究课题中涉及到的重要理论基础和关键技术，主要包括区块链系统及其结构，例如区块链的存储系统、数据结构、查询技术和跨链技术，以及传统的信息检索技术中使用的相关算法和数据模型。

第三章，针对区块链数据共享系统中的信息检索问题提出了一种链上链下混合架构中的高效搜索方案。首先介绍了方案的系统设计和整体架构，对存在的问题进行了定义并给出了系统需要体现的性能指标。然后对系统的整体工作流程进行了描述，将该方案中的系统整体上分为五个模块，前三个模块结合提供单链查询的功能，后两个模块在单链查询的基础上提供跨链查询和知识图谱可视化的功能，单链检索和跨链检索结合，共同完成了区块链数据共享系统中的高效检索方案。最后，对方案的可行性进行了具体分析，具体体现为检索效率的分析，并对方案的安全性进行了分析。

第四章，介绍了本方案的实验设计与结果分析，实验设计将实验分为系统性能测试和对比试验测试，系统性能测试对方案的性能指标进行测试，通过调节参数对不同状态下系统的检索能力进行测试，对比试验通过将本文系统与国内外其他相关的检索系统进行检索能力对比，以体现本方案的效率优势。根据实验结果，验证了本文提出的链上链下混合架构中的高效检索方案在区块链数据共享系统中的有效性以及相较于其他方案在检索能力和开销成本上的优势。

第五章，总结与展望，总结了文本研究课题的成果，分析了研究过程与工作中存在的不足和需要改进的地方，并对未来的研究目标进行了展望。

# 相关理论与技术

本章重点介绍两个部分，第一部分为区块链系统中与信息检索相关的模块和技术，包括区块链存储系统、数据结构、区块链本身提供的查询技术，以及在不同区块链之间进行互操作的跨链技术。第二部分介绍传统的信息检索技术，包括在搜索引擎、数据库、知识图谱中常用的相关算法和技术。

## 区块链系统和结构

### 区块链存储系统

在区块链系统中，存储的数据并不是底层的原始数据，底层数据往往是带有一定格式的交易数据，这些数据一般不会以明文的形式存储到区块链上，需要经过散列(hash)运算，将得到的哈希值上链，比如比特币采用双sha256散列函数[17]，可以将任意长度的原始交易记录存进区块。大多数区块链系统底层采用的是key-value数据库，通过牺牲读性能来换取写性能的提高，像是比特币和以太坊，使用levelDB数据库[23]，hyperledger fabric使用levelDB和couchDB结合的方式[7]。而随着区块链中数据的增加，对链上数据的搜索需求已经逐渐体现了出来，key-value数据库的快速写入优势无法体现，而读取效率低已经成为限制区块链搜索功能发展的瓶颈。

链上和链下混合存储架构架构作为提高区块链系统扩展性的有效方法,已经提出了一些方案，主要的做法是将大部分数据存在链下的分布式对等网络中，例如Bittorrent网络[19]和Ipfs网络[20]，然后将数据的整体散列值或者数据分块后的散列值存放到区块链网络中，比如BTT[21]将Bittorrent网络中存储数据的MD5散列值存放到区块链上，Filecoin[22]将Ipfs网络中文件分块后形成的散列结构merkle-DAG存在区块链上。但他们都是将链下数据的散列值存到链上，保证链上链下数据的一致性，但链上数据不包含链下数据的语义信息等有意义的信息。SEBDB[16]将数据存在链下的数据库中而将数据库的SQL语句存放在链上，使得链上的查询可以对应到链下的数据，但仍没摆脱传统数据库的限制，链下存储的数据必须以格式化的形式存在MySQL等关系型数据库中，对非格式化的链下数据则无法处理。

### 区块链数据结构

区块链是由一个接一个的区块链接而成的数据结构，每个区块又包含区块头和区块身体两部分，而决定区块的前后关系的核心数据，存在于区块头中，是一种被称作哈希指针的东西，每个区块头部都有一个previousBlockHash，这是对前一个的区块的哈希值的记录，由于链式结构的特殊性，可以认为一个区块头部的previousBlockHash与之前的所有区块都有联系，同时包含时间戳信息，可以认为是这种结构实现了区块链强大的数据安全性保证，然而，这种只能追加的数据结构，意味着如果想要对区块链的所有内容进行检索，你需要从后往前遍历整条链，才能确保你的搜索结果是完整的。由于这种数据结构对只包含区块头部的轻节点并不友好，一些区块链系统对交易等必要数据的验证过程进行了优化，比如，基于UTXO模型，比特币在区块中加入了一个merkle hash tree (MHT)结构[3]，通过层层哈希的方式保证区块内的交易顺序不被篡改，merkle root被存在区块头中，轻节点可以利用merkle root验证区块内交易的真实性。比特币在创始之初就为轻节点提供了便捷的验证功能，这就是简单支付验证（SPV，Simplified Payment Verification）[3]，当想要验证一条交易是否在区块内的某个位置时，只需要让full node提供该交易在区块内部的具体位置和对应的merkle proof即可，事实上，我们只需要在的时间复杂度之内即可对交易在区块内的所在位置进行判断。我们的搜索方案采用了类似SPV的搜索结果验证过程，在第4节的验证部分，我们对我们方案的验证过程进行了介绍。

不同于比特币，以太坊是基于账户的模型，被称为交易驱动的状态机[23]，每个区块的变化都可以看做是一次状态的迁移，可以描述为：

(2-1)

其中，σ为以太坊的世界状态，γ为以太坊的状态转换函数，T为一个区块，因为任何计算机系统都可以被抽象成一个状态机，所以这个状态转换函数是区块链范式的基础，也是任何去中心化交易系统的基础。 以太坊将所有账户的状态以RLP编码的方式存储在区块中的一个state trie中，叫做Merkle Patricia Trie (MPT)[23]，MPT的叶子节点是每个以太坊账户的状态，分支信息则存放账户地址的压缩前缀，矿工根据区块内的交易信息来更新MPT中账户的状态，因此最新区块中保存着所有账户的最新状态，MPT的root hash存在区块头中，因此以太坊节点可以通过最新区块中的MPT对账户的当前状态进行搜索，并通过区块头进行验证。

### 区块链查询技术

区块链底层的数据库大多是以LevelDB为代表的key-value数据库，系统提供的查询功能较为简单，仅能提供针对key的查询、更新和写入等操作。目前的区块链系统大多没有设计针对数据的查询功能，以以太坊为例，接收到用户的查询请求后，首先解析用户查询请求中的Key值，调用LevelDB的Get方法获得Value字符串，然后将Value字符串解码并用Json数据封装后返回给客户端。在以太坊平台上，查询操作包括三种:(1)按照区块高度查找区块头(2)按照区块哈希值查找区块头(3)按照交易哈希值查找交易[4]。由于在设计之初没有考虑到复杂的分析型查询需求，目前的区块链系统大都不支持复杂的语义查询、K-NN查询等方式。

为了增加区块链系统中的搜索功能，最有效的方法是增加查询处理引擎，主流的方案有两个，其中一个是外联数据库方法，是将区块链区块中的数据同步到外部数据库中，通过外部数据库的查询接口提供查询功能。图2.1是外联数据库方法的结构。

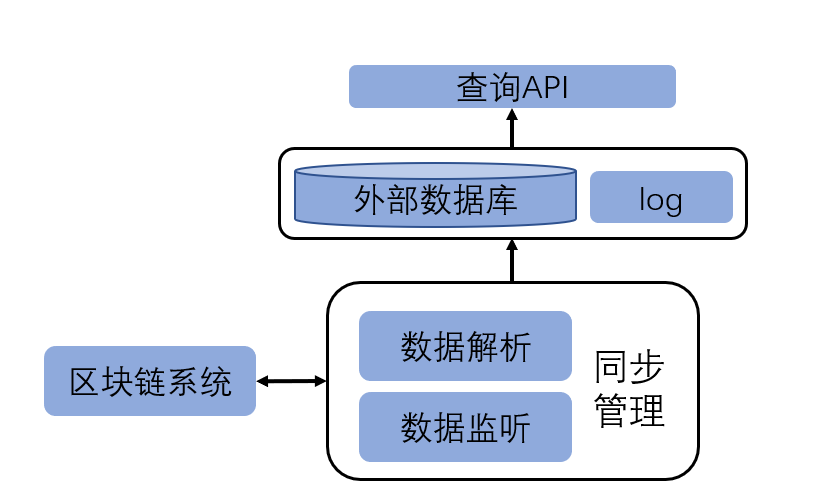


图2.1 外联数据库方法

内置索引的方式在区块链系统内部建立索引结构，通过矿工共识维护索引结构，构建一个查询层，图2.2是内置索引方法的结构。

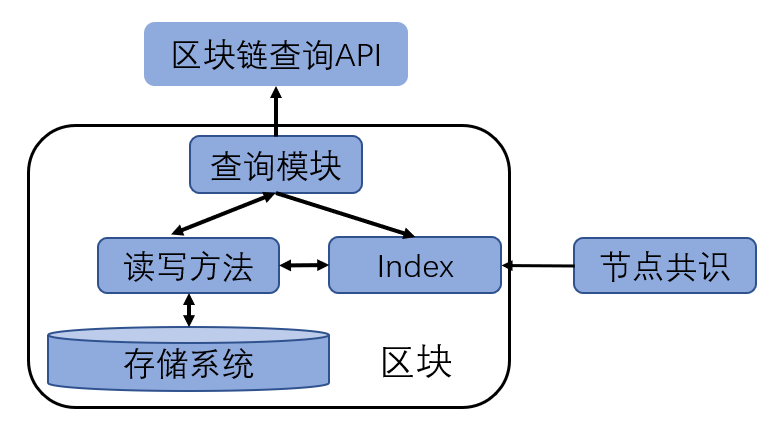


图2.2 内置索引结构

### 跨链技术

传统的区块链系统大多是独立存在的，链与链之间的共识协议不相同、数据账本不共享，这样会导致价值孤岛和数据孤岛的存在，但随着应用的发展更加广泛和深入，链链互联的需求越来越迫切。跨链技术就是为了解决区块链系统的价值孤岛和数据孤岛而产生的，为了实现区块链间的相互合作，必须通过跨链技术来实现不同链之间的互联互通。目前已经出现的跨链技术有公证人模式、中继链模式、侧链模式等，这些技术可以很好地解决跨链资产交易地原子性和安全性等问题。

## 信息检索技术

信息检索是用户进行信息查询的主要方法和手段，有广义和狭义之分，广义的信息检索包括信息的存储和检索，狭义的信息检索为信息的“查找”或“搜索”，目的都是针对用户的信息需求，通过信息检索的技术和方法返回满足用户需求的相关信息。在数据库出现之后，为了对数据库中存储的数据进行查询，人们在数据库管理系统（DBMS）中实现了数据库查询功能，特别是针对关系型数据库，利用结构化查询语言（SQL）可以快速查找到想要查询的数据。而随着互联网上海量非结构化数据的出现，结构化查询语言已经无法满足用户的搜索需求，因此全文索引技术[24]出现，全文索引技术是对大量非结构化数据进行搜索的关键技术，也是google，bing等搜索引擎的核心技术[25]。由于区块链系统中存放的数据是半结构化或非结构化的，所以本节主要介绍全文索引技术中使用的相关算法和技术，同时针对于跨链检索，本节还将介绍语义搜索和知识图谱中存在的相关技术。

### TF-IDF算法

TF-IDF算法是一种在信息检索和数据挖掘领域中常见的加权技术，是传统的统计算法，可以用于评估一个词在一个文档集或语料库中对于某一个文档的重要性[26]。TF-IDF表示一个词的重要程度随着它在文件中出现的次数增大而增大，同时随着在整个文档集中出现的频率增大而减小。TF-IDF的主要思想是如果一个词在某一篇文章中频繁出现而在其他文章中较少出现，则认为该词对于该文章在所有文章中进行分类很重要。

TF-IDF实际上是TF \* IDF，其中，TF(Term Frequency)指词条在数据中出现的频率，计算公式如下：

(2-2)

其中，表示关键字i在数据j中出现的次数，分母则表示数据j中的总词数，IDF(inverse document frequency)则用于表示词条的重要程度，计算公式为：

(2-3)

其中N表示数据总数，是包含关键字i的数据在N中出现的次数，为了避免分母为0，经常使用拉普拉斯平滑对进行处理。

TF-IDF算法具有简单易统计且结果具有代表性的优势，但只通过词频来衡量一个关键词的重要程度还不够全面，并且无法体现词的位置信息和词性等其他重要语义信息。

### 倒排索引

在无结构大数据的全文索引技术中，倒排索引是最为核心的一项技术[27]，与倒排索引相对的是正向索引（forward index），正向索引是指在搜索引擎中每个文件都对应一个文件的ID，而文件的内容则被表示为一系列关键字的集合，通过文件的ID可以搜索到文件对应的关键字。当用户对语料库中的文档进行搜索时，正向索引需要遍历索引库中的所有文档，找到包含用户输入关键词的文档，再根据文档中该关键词的重要性排名后将结果反馈给用户，这样的搜索模式是低效且成本巨大的，在海量的互联网数据中使用正向索引进行搜索并不可行。因此，将正向索引进行转置后得到的倒排索引，把文件ID到关键词组合的映射转变为了关键词到文件ID的映射，每个关键词都对应包含这个关键词的一系列文件，这样，通过搜索用户需求的关键词，就可以找到于其匹配的文档。

倒排索引主要由两部分组成，第一部分是单词词典，单词词典记录了文档集合中出现过的所有单词，并对每个单词记录了一个指向“倒排列表”的指针；第二部分是倒排列表，记录了包含单词词典中某个单词的所有文档和该词在文档中的位置信息，倒排列表顺序地存放在磁盘的一些文件中，这些文件被称为“倒排文件”。倒排索引的构建方法有转置法和归并法两种，转置法主要是通过将正向索引进行拆分后转换为倒排索引的方法，该方法是顺序执行的，因此不易用于并行处理操作；归并法则是每次将内存中的数据写入磁盘时，将包括词典在内的其他所有中间结果一并写入磁盘，生成临时倒排文件，然后对生成的多个临时倒排文件进行多路归并，输出得到最终的倒排文件。

### 检索模型分类

信息检索模型，主要是用于计算用户搜索需求和搜索结果内容相关度的理论基础和核心组件，经典的信息检索模型有布尔模型、向量模型和概率模型[28]。

布尔模型是基于集合论和布尔代数的检索模型，查询请求基本上是由and,not和or构成的布尔表达式，可以用于多关键字检索的场景中。通常，一种多关键字搜索中关键字组合的形式为<key1,key2,key3>,用and ,or等连接词进行连接，有时也可以称之为布尔搜索，用户可能会想搜索包含某几个确定关键字或者是包含几个关键字的其中一个或几个关键字的数据。比如“Blockchain”∧(“Retrieval”∨“Search”)，表示用户想要得到包含Blockchain并且包含 Retrieval与Search两者其中至少一个的搜索结果。

向量模型也叫做向量空间模型（Vector Space Model），它把每个文档表示成一个t维的向量，这个t维的向量可以是单词、词组、N-gram片段等，其中最常用的是单词。向量空间模型的定义如下:

给定一个文档 ，D符合以下两条约定：

（1）各个特征项)互异（既没有重复）；

（2）各个特征项无先后顺序关系（即不考虑文档的内部结构）

 在以上两个约定下，可以把特征项看成一个n维坐标系，而权重为相应的坐标值，因此，一 个 本就表示为维空间中的一个向量，我们称 为文本D的向量或向量空间模型，如图2.3所示：

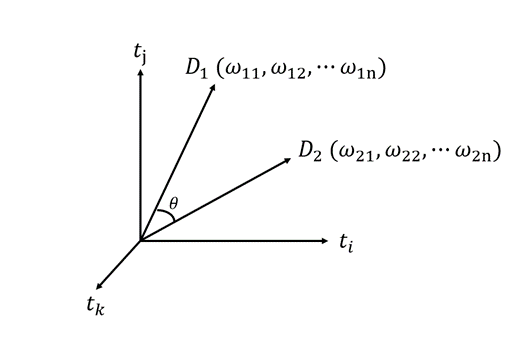


图2.3 向量空间模型

TF-IDF算法是向量空间模型中最常用的方法。

概率模型是基于概率假设的检索模型，它的基本原理为：给定一个用户的查询请求和语料库中的一篇文档d\_j,概率模型会尽可能的给出用户可以搜索到相关文档d\_j的概率。概率检索模型与贝叶斯分类的思想非常接近，但是存在着本质区别，概率检索模型的检索目标不是分类，而是根据查询请求和文档相关度的概率大小为文档进行排序。

### B+树

B+树和二叉树、平衡二叉树一样，是一个经典的数据结构。B+树由B树和索引顺序访问方法演化而来，常用于在MySQL等关系型数据库中构建索引。B+树的定义是为磁盘或者其他存储设备设计的一种平衡查找树，是B树的扩展，可以进行有效的插入、删除和搜索操作[29]。B+树通常用于数据库和操作系统的文件系统中，其特点是能够保证其中数据是稳定有序的，B+树的插入和修改操作可以保持在比较稳定的O(log n)时间复杂度。B+树通常具有几个特性:数据只存储在叶子节点上，且所有的叶子节点位于同一层，叶子节点以单链表的形式连接从而提升查询的效率；内部节点只存储键值，且由于主存储器的大小有限，主存储器中只存储内部节点，叶子节点被存储在辅助存储器中。图2.4为B+树的结构，其中最下面一层为叶子节点。

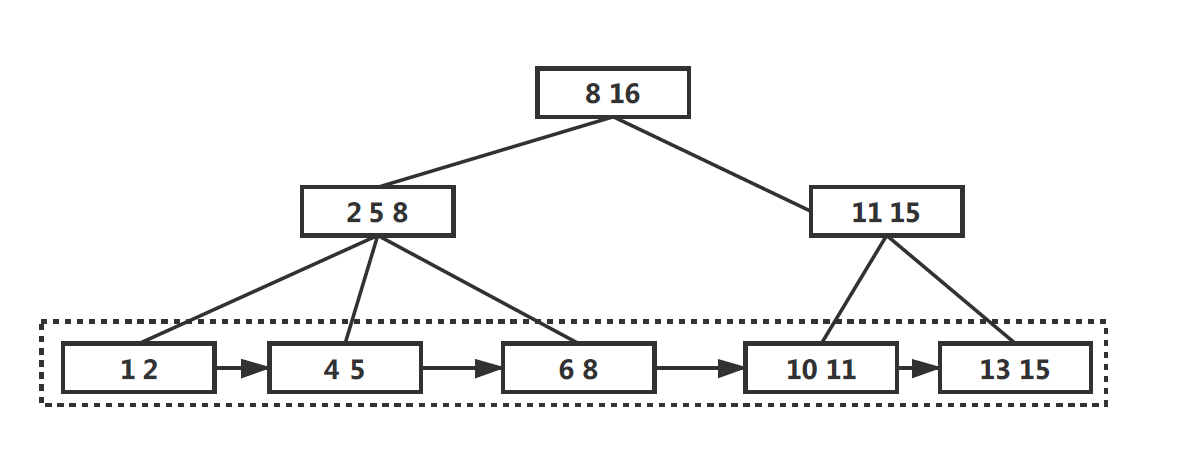


图2.4 B+树的结构

### pagerank算法

Pagerank是一种根据网页之间的超链接来计算网页排名的技术，是Google公司用于体现网页之间相关性和重要性的关键算法，也经常被用作评价搜索引擎优化程度的指标。Pagerank算法由谷歌创始人Lawrence Edward Page和Sergey Brin在斯坦福大学共同发明[30]。Pagerank的核心思想是，如果一个网页被很多其他网页链接，则该网页的重要性比较大，pagerank值较高，同时，如果一个pagerank值很高的网页链接到另一个网页，则该网页的pagerank值也会提高。

Pagerank算法主要分为两步：

1.给每个网页设置一个初始的PR值，代表网页的被访问概率，假设网页的初始PR值为，其中N为网页的总数量，在一般的情况下，所有网页的PR值总和为1。

2.通过投票算法不断迭代，更新每个网页的PR值，举例来说，A、B、C三个网页都有进入D网页的链接，则D网页的PR值为A、B、C三个网页PR值的总和，计算公式如下：

(2-2)

继续进行计算，A网页除了链接到D网页之外，还同时链接到了B网页和C网页，因此用户通过A网页访问B、C、D三个网页的概率均为1/3，同理，如果B网页同时包含对A、D两个网页的链接，则进入这两个网页的概率为1/2，C如果只能链接到D网页，则C所有的跳转网页都进入了D网页，这样A对D网页的投票量为1/3，B为1/2，C为1，因此D网页的PR值计算公式变为:

PR(D)=PR(A)/3+PR(B)/2+PR(C) (2-3)

经过总结，我们可以得出一个网页的PR值计算公式应该为：

PR(w) = (2-4)

其中，是所有链接到网页w的网页集合，网页n是属于集合的一个网页，L(n)是网页n的出度（对外链接的数量）。

### RDF数据模型

RDF（Resource Description Framework）是一种资源描述框架[33]，其中资源可以是任何文本内容，网络上的文档等，RDF是Web中进行数据交换的标准模型，可以在底层数据格式不同的情况下进行数据合并，有利于在网络中形成人机可读，并且可以由机器自动识别的文档。RDF扩展了Web的链接结构，使用URI(统一资源标识符)来定义和命名事物之间的关系，并通过一种简单的二元关系模型来表示事物之间的语义关系，即RDF三元组。RDF三元组常用来表示实体之间的关系或者实体某个属性的值，结构为<实体，属性，值>或<实体，关系，实体>。RDF三元组互相连接可以构成RDF图，一个RDF图的实例如图2.5所示:

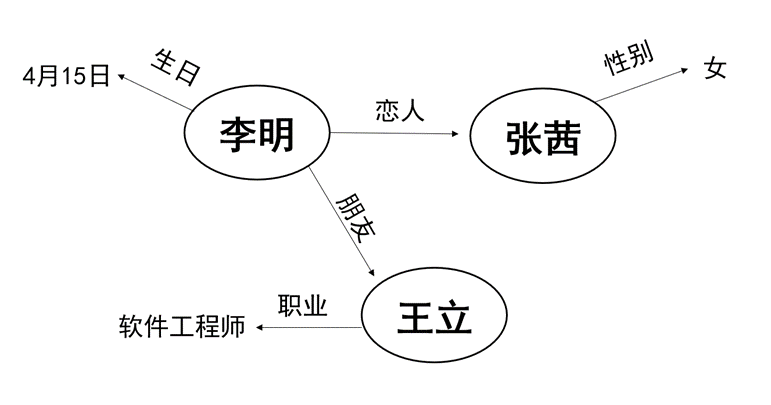


图2.5 RDF图实例

### 知识图谱构建技术

知识图谱实质上是通过关系（边）将语义实体链接起来的语义知识网，以图的形式呈现，并可以对存储的知识进行推理和进一步分析。知识图谱通过将不同类型的数据整合为人们易于使用的结构化数据，目前已经被广泛应用在语义搜索、智能问答、情报分析、金融风险控制等领域。知识图谱构建技术主要分为知识抽取、知识表示、知识融合和知识推理四大核心步骤[34]。

知识抽取主要的研究内容是如何自动或半自动地从不同异构的数据中抽取实体、属性、关系等知识要素。知识表示通过互相关联的三元组等形式对抽取出来的知识进行较为合理的表示。知识融合主要是将多个不同知识库中的知识进行整合，形成一个统一的知识库，主要解决的是实体对齐的问题，将不同知识库对实体的描述进行统一化的整合，获得对一个实体较为完整全面的描述。知识推理则是根据已有的数据模型和数据，挖掘出隐含的知识或识别出错误知识，对现有的知识图谱进行扩充和纠正。除了这四个核心步骤之外，还有知识存储、质量评估等步骤，知识图谱的构建是一个系统化的工程，涉及到各种信息检索或自然语言处理的知识，目前的技术仍不算成熟，还有很多挑战需要研究人员去解决。

## 本章小结

本章根据第一章介绍的区块链信息检索的研究背景和基本现状，从区块链系统和传统的检索技术两方面介绍了本文相关的理论知识。首先介绍了区块链系统中与查询相关的结构和技术，体现了区块链查询模式的特殊性，随后分类介绍了传统信息检索中使用的不同技术，并分析了它们在实现查询功能中的优劣势，以探索其与区块链结合的可能性。这些相关知识为本文后续方案的设计提供了坚实的理论和技术基础。

# 链上链下混合存储架构中高效搜索技术方案

## 学位论文封面的编辑和打印要求

学位论文的封面由研究生院按国家规定统一制定印刷，封面内容必须打印，不得手写。

## 学位论文的版面设置要求

（1）行间距：固定值20磅（题名页除外）。

（2）字符间距：标准。

（3）页眉设置：单面页码页眉标题为章节题目，每一章节的起始页必须在单面页码，双面页码页眉标题统一为“西安电子科技大学博/硕士学位论文”，页眉标题居中排列，字体为宋体，字号为五号。页眉文字下添加双横线，双横线宽度为0.5磅，距正文距离为：上下各1磅，左右各4磅。

（4）页码设置：学位论文的前置部分和主体部分分开设置页码，前置部分的页码用罗马数字标识，字体为Times New Roman，字号为小五号；主体部分的页码用阿拉伯数字标识，字体为宋体，字号为小五号。页码统一居于页面底端中部，不加任何修饰。

（5）页面设置：为了便于装订，要求每页纸的四周留有足够的空白边缘，其中页边距为上3厘米、下2厘米；内侧2.5厘米、外侧2.5厘米；装订线为0.5厘米；页眉2厘米，页脚1.75厘米。

## 学位论文的打印、装订要求

（1）打印：学位论文必须用A4纸页面排版，双面打印；

（2）装订：依次按照中文题名页、英文题名页、声明、摘要、插图索引、表格索引、符号对照表、缩略语对照表、目录、正文、附录（可选）、参考文献、致谢、作者简介的顺序，用学校统一印制的学位论文封面装订成册。盲审论文必须删除致谢部分的文字内容（致谢标题须保留）以及封面和研究成果中的作者和指导教师姓名，研究成果列表中应体现作者的排序，如第一作者、第一发明人等。

## 其他说明

本规定由研究生院负责解释，从申请2015年9月毕业和授位的研究生开始执行，其它有关规定同时废止。研究生毕业论文撰写要求参照学位论文撰写要求执行。

# 图、表、公式示例

图：包括曲线图、示意图、流程图、框图等。图序号一律用阿拉伯数字分章依序编码，如：图1.3、图2.11。

每一个图应有简短确切的图名，连同图序号置于图的正下方。图名称、图中的内容字号为五号，中文字体为宋体，英文字体为Times New Roman，行距一般为单倍行距。图中坐标上标注的符号和缩略词必须与正文保持一致。引用图应在图题右上角标出文献来源；曲线图的纵横坐标必须标注“量、标准规定符号、单位”，这三者只有在不必要标明（如无量纲等）的情况下方可省略。

图与正文之间一般应空一行。



插图示例

公式：正文中的公式、算式、方程式等必须编排序号，序号一律用阿拉伯数字分章依序编码，如：(3-32)、 (6-21)。

对于较长的公式，另起行居中横排，只可在符号处（如：+、-、\*、/、< >等）转行。公式序号标注于该式所在行（当有续行时，应标注于最后一行）的最右边。连续性的公式在“=”处排列整齐。大于999的整数或多于三位的小数，一律用半个阿拉伯数字符的小间隔分开；小于1的数应将0置于小数点之前。公式的行距一般为单倍行距。

公式与正文之间一般应空一行。



表：包括分类项目和数据，一般要求分类项目由左至右横排，数据从上到下竖列。

分类项目横排中必须标明符号或单位，竖列的数据栏中不要出现“同上”、“同左”等词语，一律要填写具体的数字或文字。表序号一律用阿拉伯数字分章依序编码，如：表2.5、表10.3。

每一个表格应有简短确切的题名，连同表序号置于表的正上方。表名称、表中的内容居中排列，字号为五号，中文字体为宋体，英文字体为Times New Roman，行距一般与正文保持一致。表格线统一用单线条，磅值为0.5磅。

表格与正文之间一般应空一行。

表格示例

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 馈电方式  电性能参数 | 探针 | 环形缝隙 | 探针和缝隙 | | 缝隙和CPW | |
| 探针 | 缝隙 | 缝隙 | CPW |
| 谐振频率 | 9.5 GHz | 8.8 GHz | 9.4 GHz | 9.8 GHz | 9.2 GHz | 9.3 GHz |
| 带宽  （|S11|<-10 dB） | 7.3% | 4.5% | 6.9% | 6.8% | 4.9% | 5.3% |
| 隔离度  （带内最差） | -16.5 dB | -17 dB | -31 dB | | -22 dB | |
| 方向图 | 不对称 | 对称 | 不对称 | 对称 | 对称 | 对称 |
| 交叉极化电平 | 高 | 低 | 高 | 低 | 低 | 低 |

计量单位：学位论文中出现的计量单位一律采用国务院1984年2月27日发布的《中华人民共和国法定计量单位》标准。

参考文献

普通图书示例：

1. 广西壮族自治区林业厅. 广西自然保护区[M]. 北京: 中国林业出版社, 1993.
2. 蒋有绪, 郭泉水, 马娟等. 中国森林群落分类及其群落学特征[M]. 北京: 科学出版社, 1998.
3. 唐绪军. 报业经济与报业经营[M] .北京: 新华出版社, 1999: 117-121.
4. 赵凯华, 罗蔚茵. 新概念物理教程: 力学[M]. 北京: 高等教育出版社, 1995.
5. 汪昂. (增补)本草备要[M]. 石印本. 上海: 同文书局, 1912 .
6. CRAWFPRD W, GORMAN M. Future libraries: dreams, madness, & reality[M]. Chicago: American Library Association, 1995.
7. International Federation of Library Association and Institutions. Names of persons: national usages for entry in catalogues[M]. 3rd ed. London: IFLA International Office for UBC, 1977.
8. O’BRIEN J A. Introduction to information systems[M]. 7th ed. Burr Ridge, III.: Irwin, 1994.
9. ROOD H J. Logic and structured design for computer programmers[M]. 3rd ed. Watertown: Brooks/Cole Thomson Learning, 2001.

论文集、会议录示例：

1. 中国力学学会. 第3届全国实验流体力学学术会议论文集[C]. 天津: 南开大学出版社,1990.
2. ROSENTHALL E M. Proceedings of the Fifth Canadian Mathematical Congress, University of Montreal, 1961[C]. Toronto: University of Toronto Press, 1963.
3. GANZHA V G, MAYR E W, VOROZHTSOV E V. Computer algebra in scientific computing: CASC 2000: proceedings of the Third Workshop on Computer Algebra in Scientific Computing, Samarkand, October 5-9, 2000[C]. Berlin: Springer, c2000.

科技报告示例：

1. U. S. Department of Transportation Federal Highway Administration. Guidelines for bandling excavated acid-producing materials, PB 91-194001[R]. Springfield:U. S.Department of Commerce National Information Service, 1990.
2. World Health Organization. Factors regulating the immune response: report of WHO Scientific Group[R]. Geneva: WHO, 1970.

学位论文示例：

1. 张志祥. 间断动力系统的随机扰动及其在守恒律方程中的应用[D]. 北京: 北京大学数学学院, 1998.
2. CALMS R B. Infrared spectroscopic studies on solid oxygen[D]. Berkeley: Univ. of California. 1965.

专利文献示例：

1. 刘加林. 多功能一次性压舌板: 中国, 92214985. 2[P]. 1993-04-14.
2. 河北绿洲生态环境科技有限公司.一种荒漠化地区生态植被综合培育种植方法: 中国, 01129210.5[P/OL]. 2001-10-24[2002-05-28]. http://211.152.9.47/sipoasp/zlijs/hyjs-yx- new.asp? recid=01129210.5&leixin.
3. KOSEKI A, MOMOSE H, KAWAHITO M, et al. Compiler: US, 828402[P/OL]. 2002-05-25 [2002-05-28].http://FF&p=1&u=netahtml/PTO/search-bool.html&r=5&f=G&1=50&co1=AND&d=PGOl&sl=IBM. AS. &OS=AN/IBM&RS=AN/IBM.

专著中析出的文献示例：

1. 国家标准局信息分类编码研究所. GB/T 2659-1986世界各国和地区名称代码[S]//全国文献工作标准化技术委员会. 文献工作国家标准汇编: 3.北京: 中国标准出版社, 1988:59-92.
2. 韩吉人. 论职工教育的特点[G]//中国职工教育研究会. 职工教育研究论文集. 北京: 人民教育出版社, 1985:90-99.
3. BUSECK P R, NORD G L, Jr. , VEBLEN D R. Subsolidus phenomena in pyroxenes[M]// PREWITT C T. Pyroxense. Washington, D. C. : Mineralogical Society of America, c1980: 117-211.
4. FOURNEY M E. Advances in holographic photoelasticity [C]// American Society of Mechanical Engineers. Applied Mechanics Division. Symposium on Applications of Holography in Mechanics, August 23-25, 1971, University of Southern California, Los Angeles, California. New York: ASME, c1971 : 17-38.
5. MARTIN G.. Control of electronic resources inAustralia[M]//PATTLE L W, COX BJ. Electronic resources: selection and bibliographic control. New York: The Haworth Press, 1996: 85-96.

期刊中析出的文献示例：

1. 李炳穆. 理想的图书馆员和信息专家的素质与形象[J]. 图书情报工作, 2000(2):5-8.
2. 陶仁骥. 密码学与数学[J]. 自然杂志, 1984,7(7):527.
3. 习亚洲地质图编目组. 亚洲地层与地质历史概述[J]. 地质学报, 1978, 3: 194-208.
4. DES MARAIS D J, STRAUSS H, SUMMONS R E, et al. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment [J]. Nature, 1992, 359: 605-609.
5. HEWITT J A. Technical services in 1983[J]. Library Resource Services, 1984, 28(3): 205-218.

报纸中析出的文献示例：

1. 丁文祥. 数字革命与竞争国际化[N]. 中国青年报, 2000-11-20(15).
2. 张田勤. 罪犯DNA库与生命伦理学计划[N]. 大众科技报, 2000-11-12(7).

电子文献(包括专著或连续出版物中析出的电子文献)示例：

1. 江向东. 互联网环境下的信息处理与图书管理系统解决方案[J/OL]. 情报学报, 1999, 18(2): 4[2000-01-18].
2. 萧钮. 出版业信息化迈人快车道[EB/OL]. (2001-12-19) [2002-04-15]
3. CHRISTINE M. Plant physiology: plant biology in the Genome Era[J/OL]. Science, 1998, 281:331-332[1998-09-23].
4. METCALF S W. The Tort Hall air emission study[C/OL]//The International Congress on Hazardous Waste, Atlanta Marriott Marquis Hotel, Atlanta,Georgia, June 5-8, 1995: impact on human and ecological health[1998-09-22].
5. TURCOTTE D L. Fractals and chaos in geology and geophysics[M/OL]. New York: Cambridge University Press, 1992[1998-09-231.
6. Scitor Corporation. Project scheduler[CP/DK]. Sunnyvale, Calif. : Scitor Corporation, c1983.

致谢

本论文是在导师的悉心指导下完成的，从论文的选题到论文的撰写，无不渗透着导师的心血，……值此论文完稿之际，谨对导师的辛勤培育以及谆谆教诲表示最衷心的感谢!

作者简介

##### 基本情况

张三，男，陕西西安人，1982年8月出生，西安电子科技大学XX学院XX专业2008级硕士研究生。

##### 教育背景

2001.08～2005.07西安电子科技大学，本科，专业：电子信息工程

2008.08～ 西安电子科技大学，硕士研究生，专业：电磁场与微波技术

##### 攻读硕士学位期间的研究成果

###### 发表学术论文

1. XXX, XXX, XXX. Rapid development technique for drip irrigation emitters[J]. RP Journal,UK.,2003,9(2): 104-110.(SCI: 672CZ, EI: 03187452127)
2. XXX, XXX, XXX. 基于快速成型制造的滴管快速制造技术研究[J]. 西安交通大学学报, 2001, 15(9): 935-939. (EI: 02226959521)

###### 申请（授权）专利

1. XXX, XXX, XXX等. 专利名称: 国别,专利号[P]. 出版日期.

###### 参与科研项目及获奖

1. XXX项目, 项目名称, 起止时间, 完成情况, 作者贡献。
2. XXX, XXX, XXX等. 科研项目名称. 陕西省科技进步三等奖, 获奖日期.