

学校编码: 10384

分类号_____密级_____

学 号: 22320051302486

UDC_____

厦 门 大 学

硕 士 学 位 论 文

电子邮件调查分析系统的设计与实现

The Design and Implementation of Email Investigation and
Analysis System

曾 春 溪

指导教师姓名: 吴 顺 祥 教授

专 业 名 称: 系 统 工 程

论文提交日期: 2008 年 4 月

论文答辩时间: 2008 年 5 月

学位授予日期: 2008 年 月

答辩委员会主席: _____

评 阅 人: _____

2008 年 5 月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

兹呈交的学位论文,是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果,均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人(签名):

年 月 日

厦门大学博硕士学位论文摘要库

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

- 1、保密（ ），在 年解密后适用本授权书。
- 2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

厦门大学博硕士学位论文摘要库

摘 要

随着信息网络的高速发展，电子邮件作为一种快捷便利的通信手段，已经深入普及到人们的日常工作与生活中，其间蕴含着丰富的个人信息，是进行计算机调查取证的重要途径，能为案件侦破提供有力的线索。电子邮件客户端为邮箱用户提供了友好的管理界面，受到网民的普遍青睐，使用率高，其保存的邮件数据文件也是调查取证的重要对象，挖掘分析其中的有用线索是计算机取证的重要手段。

本文研究的电子邮件调查分析系统主要是针对 Microsoft Outlook 和 Outlook Express 这两种常用的电子邮件客户端软件，通过解析客户端下保存的邮件数据文件，即从 Pst、Dbx 两种格式的邮件存储复合文档中提取相关邮件的收发帐户邮件地址、主题、发送时间、邮件内容和附件等信息，然后根据收发件人邮件地址进行统计归类，运用网络分析、可视化等技术绘制邮件时间图和网络图，进而为分析收、发件人之间隐藏的内部关系网提供有力依据。

本文首先分析了电子邮件调查分析的重要意义，介绍了国内外电子邮件调查分析的现状；接着，详细阐述了本系统开发的关键技术：基于 COM 技术的 Dbx 邮件文件解析、基于 OLE 自动化技术的 Pst 邮件文件解析，以及在绘制邮件时间关系图和网络关系图过程中用到的数据库访问技术、图形绘制基本理论和基于遗传算法的图自动布局算法等；然后，介绍了系统的总体目标、运行环境、功能需求，以及详细功能设计和具体实现的基本情况等；最后，对本系统的研究开发情况进行了总结，并就开发中的不足之处提出了下步的努力方向。

本文研究的电子邮件调查分析系统经过反复测试和完善，目前性能良好、运行稳定，达到了系统开发的目标，它的深入研究和使用的将我国计算机调查取证方式的发展起到积极的作用。

关键词： 电子邮件；文件解析；COM；OLE 自动化；ADO；遗传算法

厦门大学博硕士学位论文摘要库

ABSTRACT

Along with the rapid development of information network, E-mail, as an important and quick means of communication, has been deeply into people's daily work and life. The rich personal information contained inside is significant for investigation and evidence collection, providing important clues to handle cases. E-mail clients offer a friendly user interface, generally favored by netizens. The mail data files preserved become an important target of computer forensics, from which we can mine out and analyze the useable clues.

The Email Investigation & Analysis System mentioned in this paper is mainly for Microsoft Outlook and Outlook Express, which are two commonly used e-mail clients. Through parsing of data files kept by mail-client, that is, extract the address, subject, date, mail-content and attachment of related mails from the Pst and Dbx format compound documents, then make statistic and classification according to sender-address or receiver-address. Using social networking analysis, visualization, and other technique to draw mail time-line layout and network layout, we can discover the hidden internal relation network between senders and receivers, and get powerful evidence for cases investigation.

This paper described the significance of e-mail survey firstly. Then, introduced in detail the key technologies of the system development: Dbx mail file parsing based on COM technology; Pst mail file parsing based on OLE Automation technology; database access technology used in the system; basic theory of graph drawing and Automating layout algorithm based on GA (Genetic Algorithm). The overall goal of the system, running environment, functional requirements, as well as designing details and other features are also described then. Finally, summarize the system development and on the lack propose the next step to make efforts.

The system discussed in the paper has gone through repeated testing and perfecting. At the moment it is in good performance, operation stability and achieves the objectives of the system development. So the system will have some active action for the development of computer forensics in our country, and has some value to spread.

Keywords: E-mail; File Parsing; COM; OLE Automation; ADO; Genetic Algorithm

厦门大学博硕士学位论文摘要库

目 录

第一章 绪 论	1
1.1 研究背景及其意义	1
1.2 电子邮件调查分析研究现状	1
1.2.1 国外研究现状	2
1.2.2 国内研究现状	2
1.3 电子邮件格式简介	3
1.3.1 电子邮件语法结构	3
1.3.2 电子邮件编码方式	5
1.4 本文研究内容与组织	6
第二章 基于 COM 技术的 Dbx 邮件文件解析	8
2.1 Outlook Express 及其 Dbx 数据文件	8
2.1.1 Outlook Express 简介	8
2.1.2 Dbx 数据文件介绍	9
2.1.3 Dbx 文件结构简介	10
2.2 COM 组件技术	11
2.2.1 COM 组件的特性	11
2.2.2 COM 的组成	12
2.2.3 COM 组件编程基础知识	15
2.3 基于 COM 技术的 Dbx 邮件文件解析	18
2.3.1 问题提出	18
2.3.2 使用的主要 COM 接口	19
2.3.3 解析流程	19
2.4 本章小结	20
第三章 基于 OLE 自动化技术的 Pst 邮件文件解析	21
3.1 Outlook 及其 Pst 数据文件	21
3.1.1 Outlook 简介	21
3.1.2 Pst 数据文件介绍	22
3.1.3 Pst 文件结构简介	22
3.2 OLE 自动化技术	23
3.2.1 OLE 自动化概述	24
3.2.2 OLE 自动化编程基础知识	24
3.3 基于 OLE 自动化技术的 Pst 邮件文件解析	27
3.3.1 问题提出	27

3.3.2 使用的类型库和主要类.....	28
3.3.3 解析流程.....	29
3.4 本章小结	30
第四章 邮件逻辑关系图的实现技术研究	31
4.1 数据库访问技术.....	31
4.1.1 数据库访问技术比较.....	31
4.1.2 ADO 对象模型.....	32
4.1.3 VC 中使用 ADO 方式操作 ACCESS 数据库.....	33
4.2 图形绘制知识简介	37
4.2.1 图的基本概念.....	37
4.2.2 图的布局问题.....	37
4.2.3 图形编程基础知识.....	38
4.3 基于遗传算法的图自动布局算法.....	39
4.3.1 遗传算法简介.....	40
4.3.2 基于遗传算法的图自动布局算法.....	43
4.3.3 算法分析与结论.....	46
4.4 本章小结	46
第五章 电子邮件调查分析系统的设计与实现	47
5.1 系统概述	47
5.1.1 系统总体目标.....	47
5.1.2 系统运行环境.....	47
5.1.3 系统功能描述.....	47
5.2 系统开发环境	49
5.3 系统总体结构	49
5.4 邮件文件解析模块的设计与实现.....	50
5.4.1 整体结构.....	50
5.4.2 数据库设计.....	51
5.4.3 Dbx 邮件文件解析的程序实现.....	51
5.4.4 Pst 邮件文件解析的程序实现.....	54
5.5 关键字查询模块的设计与实现.....	58
5.5.1 整体结构.....	58
5.5.2 界面设计.....	58
5.6 邮件逻辑图绘制模块的设计与实现.....	60
5.6.1 整体结构.....	61
5.6.2 数据预处理.....	61
5.6.3 邮件时间关系图的实现.....	62
5.6.4 邮件网络关系图的实现.....	65

5.7 本章小结	68
第六章 总结与展望	69
6.1 论文工作总结	69
6.2 研究工作展望	69
参考文献	71
攻读硕士学位期间发表的学术论文	75
致 谢	76

厦门大学博硕士论文摘要库

厦门大学博硕士学位论文摘要库

Contents

Chapter 1 Introduction	1
1.1 Background and Significance of the Research	1
1.2 Researching Status of Email Investigation and Analysis	1
1.2.1 International Status	2
1.2.2 Domestic Status	2
1.3 Introduction to the Format of Email	3
1.3.1 Grammar Structure of the Email	3
1.3.2 Encoding Methods of the Email	5
1.4 Study Contents and Structure of this Thesis	6
Chapter 2 Dbx Mail File Parsing Based on COM Technology	8
2.1 Outlook Express and the Corresponding Dbx-Format Data File	8
2.1.1 Brief Introduction to Outlook Express	8
2.1.2 Introduction to Dbx-Format Data File	9
2.1.3 Grammar Structure of the Dbx-Format File	10
2.2 COM Component Technology	11
2.2.1 Characters of COM Component	11
2.2.2 Constitution of COM Component	12
2.2.3 Basic Knowledge of COM Component Programming	15
2.3 Dbx Mail File Parsing Based on COM Technology	18
2.3.1 Problem Indicate	18
2.3.2 Main COM Interfaces in Use	19
2.3.3 Parsing Steps	19
2.4 Chapter Summary	20
Chapter 3 Pst Mail File Parsing Based on OLE Automation Technology	21
3.1 Outlook and the Corresponding Pst-Format Data File	21
3.1.1 Brief Introduction to Outlook	21
3.1.2 Introduction to Pst-Format Data File	22
3.1.3 Grammar Structure of the Pst-Format File	22
3.2 OLE Automation Technology	23
3.2.1 Introduction to OLE Automation	24
3.2.2 Basic Knowledge of OLE Automation Programming	24
3.3 Pst Mail File Parsing Based on OLE Automation Technology	27

3.3.1 Problem Indicate	27
3.3.2 Type Library and Main Classes in Use	28
3.3.3 Parsing Steps	29
3.4 Chapter Summary	30
Chapter 4 Research on Implementation Technology of Mail Logic Relational Graph	31
4.1 Database Access Technology	31
4.1.1 Comparison of Database Access Technology	31
4.1.2 ADO Object Model	32
4.1.3 Using ADO Method to Operate Access Database in VC	33
4.2 Brief Introduction to Knowledge of Graph Drawing	37
4.2.1 Basic Conception of Graph	37
4.2.2 Layout Problem of Graph	37
4.2.3 Basic Knowledge of Graphic Programming	38
4.3 Graph Automating Layout Algorithm Based on GA	39
4.3.1 Brief Introduction to Genetic Algorithm	40
4.3.2 Graph Automating Layout Algorithm Based on GA	43
4.3.3 Algorithm Analysis and Conclusion	46
4.4 Chapter Summary	46
Chapter 5 Design and Implementation of the Email Investigation and Analysis System	47
5.1 Summary of the System	47
5.1.1 Main Target of the System	47
5.1.2 Running Environment of the System	47
5.1.3 Function Describing of the System	47
5.2 Development Environment of the System	49
5.3 Main Structure of the System	49
5.4 Design and Implementation of Mail File Parsing Module	50
5.4.1 Unitary Structure	50
5.4.2 Design of Database	51
5.4.3 Programming Implementation of Dbx Mail File Parsing	51
5.4.4 Programming Implementation of Pst Mail File Parsing	54
5.5 Design and Implementation of Keyword Querying Module	58
5.5.1 Unitary Structure	58
5.5.2 Design of Running Interface	58
5.6 Design and Implementation of Mail Logic Diagram Drawing Module	60
5.6.1 Unitary Structure	61
5.6.2 Data Preprocessing	61

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库