第一图书网, tushu007.com

<<信息内容安全管理及应用>>

图书基本信息

书名：<<信息内容安全管理及应用>>

13位ISBN编号：9787111299547

10位ISBN编号：711129954X

出版时间：2010-7

出版时间：李翔、李生红、刘功申、 等 机械工业出版社 (2010-07出版)

作者：李翔 等 著

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：http://www.tushu007.com

前言

近年来发生了很多安全事件，例如美国9·11事件、伦敦公交系统连环爆炸案、巴厘岛恐怖袭击、印度孟买恐怖袭击等。
灾难的发生促使大众开始重新审视社会各个方面的安全性和可靠性。
在这种环境下，计算机被认为是解决此类安全问题的一个有力工具，例如，它被广泛用来收集和分析情报。
美国政府在9·11事件后，建立了全球联网的指纹系统及日趋严格的出入境管理体系，以期建筑严密的恐怖袭击防控网络，尽管由于对恐怖活动的规律性还缺乏清晰的认识，这些网络暂时还未发挥出预警和防范恐怖袭击事件的作用。
就计算机本身而言，无论从硬件到软件，还是从操作系统到数据管理系统，都存在严重的安全问题。
网络所带来的计算机安全问题则更为严重。
网络互连在方便信息传送的同时，也给连网计算机所保护的信息带来了威胁。
除了基于网络和软硬件的安全问题以外，近几年来，互联网还暴露了其他的一些安全隐患，尤其是一些对于整个社会都起到负面影响的安全问题。
最为引人注目的是，自2005以来爆发的多起“人肉搜索”等网络暴力事件，把互联网中内容安全问题暴露在公众眼前。
事实上，网络“暴力”由来已久，互联网上公开的信息及越来越强大的搜索功能，使原本隐在角落的信息被“曝光”到大众视野内，一些本不构成隐私的信息在互联网上任意传播，并在引发网络上的语言暴力后，造成了严重的后果。
还有数字信息的知识产权问题。
由于数字信息复制及网络传播非常便利，造成信息自身具有的知识产权被有意或无意地侵犯。
尽管在欧洲发生了几起因有意或无意的共享了具有知识产权歌曲而引发的多起诉讼和巨额的罚金，但法律毕竟是版权侵权的最后防范手段。
目前，已经出现了在组织内部（局域网范围内）防范信息泄露的技术手段，尽管在整个互联网领域此类技术还很缺乏，但我们有理由相信计算机技术将能够起到更为重要的作用。
以上是一些计算机安全中的新型问题，大多是公共或私有信息的内容所带来的风险。
这些风险中，有些是商业风险，有些是个人或者组织的危机，有些是社会的安全风险。
相比于传统的信息安全问题，例如通信安全、计算机安全等与计算机网络和软硬件设备关系紧密的安全问题不同，对此类风险的评估及加强安全的防护是新的一类信息安全问题，我们把它称为“信息内容安全”，或称为“内容安全”。
本书是对此类问题的分析及相关技术的总结和介绍。

内容概要

《信息内容安全管理及应用》从信息处理的基本理论开始讲解，通过几个具有代表性的信息内容安全应用实例，系统地介绍信息内容安全在目前的发展和现实水平。
《信息内容安全管理及应用》共9章，主要内容包括互联网信息内容获取、文本特征的抽取、音频和视频特征抽取、信息处理模型和方法、分类算法、信息过滤、数字水印和舆情系统等。
《信息内容安全管理及应用》可作为高等院校信息安全相关专业信息内容安全课程的教材，也可作为从事信息内容安全工作的科技人员、工程技术人员以及其他相关部门人员的参考资料。

书籍目录

出版说明 前言 第1章 绪论 1 1.1 信息内容安全概述 1 1.2 信息内容安全威胁 2 1.3 信息内容安全特点及其与相关学科的联系 2 1.4 信息内容安全研究现状 3 1.4.1 政府部门主导的项目 3 1.4.2 科研院所或公司的项目与产品 4 1.5 信息内容安全研究的意义 4 1.6 本章小结 5 1.7 习题 5 第2章 网络信息内容的获取 6 2.1 互联网信息类型 6 2.1.1 网络媒体信息 6 2.1.2 网络通信信息 8 2.2 网络媒体信息获取原理 8 2.2.1 网络媒体信息获取理想流程 8 2.2.2 网络媒体信息获取的分类 11 2.2.3 网络媒体信息获取的技术难点 13 2.3 网络媒体信息获取方法 13 2.3.1 需身份认证静态媒体发布信息获取 13 2.3.2 内嵌脚本语言片段的动态网页信息获取 17 2.3.3 基于浏览器模拟实现网络媒体信息获取 20 2.4 网络通信信息获取方案 24 2.5 本章小结 25 2.6 习题 25 第3章 文本信息的特征抽取和选择 26 3.1 文本特征的抽取和选择概述 26 3.2 语义特征的抽取 27 3.2.1 词级别语义特征 27 3.2.2 亚词级别语义特征 29 3.2.3 语义与语用级别语义特征 30 3.2.4 汉语的语义特征抽取 30 3.3 特征子集选择 31 3.3.1 停用词过滤 32 3.3.2 文档频率阈值法 33 3.3.3 TF-IDF 34 3.3.4 信噪比 34 3.3.5 信息增益 35 3.3.6 卡方统计 36 3.4 特征重构 36 3.4.1 词干 36 3.4.2 知识库 37 3.4.3 潜在语义索引 37 3.5 向量生成 40 3.5.1 局部系数 40 3.5.2 全局系数 41 3.5.3 规范化系数 41 3.5.4 几种常见的组合方式 41 3.6 本章小结 42 3.7 习题 42 第4章 音频信息特征抽取 43 4.1 数字音频技术概述 43 4.2 人类的听觉感知 44 4.3 音频信号分析和编码 47 4.3.1 音频信号的特征分析 47 4.3.2 音频信号的数字编码 48 4.3.3 数字音频信号的解析 48 4.4 音频信息特征抽取 49 4.4.1 基于帧的音频特征 50 4.4.2 基于片段的音频特征 51 4.5 本章小结 52 4.6 习题 53 第5章 图像信息特征抽取 54 5.1 数字图像的表示方法 54 5.2 图像颜色特征提取 56 5.2.1 颜色直方图特征 56 5.2.2 颜色聚合矢量特征 59 5.2.3 颜色矩特征 60 5.2.4 其他颜色特征 61 5.3 图像纹理特征提取 61 5.3.1 灰度共生矩阵 61 5.3.2 Gabor小波特征 62 5.3.3 Tamura特征 63 5.3.4 纹理特征 64 5.4 其他图像特征 64 5.4.1 边缘特征 64 5.4.2 轮廓特征 65 5.5 本章小结 66 5.6 习题 66 第6章 信息处理模型和方法 67 6.1 文本模式匹配算法 67 6.1.1 经典单模式匹配算法 67 6.1.2 经典多模式DFSA匹配算法 71 6.2 分类算法 73 6.2.1 线性分类器 74 6.2.2 最近邻分类法 75 6.2.3 支持向量机 76 6.2.4 传统Bayes分类方法 78 6.2.5 向量空间模型法 79 6.3 本章小结 80 6.4 习题 81 第7章 信息过滤 82 7.1 信息过滤概述 82 7.1.1 信息过滤研究的历史 83 7.1.2 信息过滤的分类体系 84 7.1.3 信息过滤的应用 86 7.1.4 信息过滤的评价 86 7.2 内容安全的信息过滤 87 7.2.1 信息过滤与其他信息处理的异同 87 7.2.2 用户过滤和安全过滤 88 7.2.3 现有信息过滤系统及技术 90 7.3 基于匹配的文本过滤 92 7.3.1 特征字串匹配查全率估算 93 7.3.2 准确率估算试验 94 7.4 基于邻近类别分类的过滤 95 7.5 本章小结 96 7.6 习题 97 第8章 数字水印 98 8.1 数字水印概述 98 8.1.1 数字水印的历史 98 8.1.2 数字水印的现状 99 8.1.3 数字水印分类 101 8.1.4 数字水印基本要求 102 8.1.5 数字水印的应用领域 104 8.1.6 数字水印的发展趋势 106 8.2 数字水印理论与模型 108 8.2.1 系统数学模型 108 8.2.2 数字水印的一般定义 108 8.2.3 数字水印的基本特性 109 8.2.4 数字水印与密码学的区别 110 8.3 数字音频水印技术 113 8.3.1 数字音频水印算法 113 8.3.2 数字音频水印攻击 114 8.3.3 数字音频水印算法评价准则 116 8.4 数字图像水印技术 116 8.4.1 数字图像水印算法 116 8.4.2 数字图像水印攻击 119 8.4.3 数字图像水印评价准则 121 8.5 数字视频水印技术 123 8.5.1 数字视频水印算法 123 8.5.2 数字视频水印攻击 126 8.5.3 数字视频水印技术的特殊要求 127 8.6 一种基于DCT视频水印的改进算法 128 8.6.1 算法模型介绍 128 8.6.2 算法基本思想 129 8.6.3 嵌入算法步骤 129 8.6.4 提取算法步骤 130 8.6.5 仿真试验分析 131 8.7 本章小结 136 8.8 习题 136 第9章 网络舆情监测与预警系统 137 9.1 舆情系统的背景和应用范围 137 9.1.1 现状 137 9.1.2 舆情系统的发展趋势 139 9.1.3 舆情系统的应用 142 9.2 舆情系统的功能分解 143 9.2.1 技术发展背景 143 9.2.2 高仿真网络信息深度提取 148 9.2.3 高性能信息自动提取机器人技术 149 9.2.4 基于语义的海量文本特征快速提取与分类 150 9.2.5 多媒体群件理解技术 151 9.2.6 非结构信息自组织聚合表达 152 9.2.7 非结构信息数据挖掘技术 153 9.3 互联网论坛信息分析 154 9.3.1 面向互联网论坛的定点网站深入挖掘机制 155 9.3.2 异构数据归一化存储与目标站点热点查询 156 9.3.3 监控目标热点自动发现功能 156 9.4 本章小结 157 9.5 习题 157 参考文献 158

章节摘录

插图：多词级别中一种思路是应用名词短语作为特征项，这种方法也称为Syntactic Phrase Indexing。
另外一种策略则是不考虑词性，只从统计角度根据词之间较高的同现频率（Co-Occur Frequency）来选取特征项。
采用名词短语或者同现高频词作为特征项，需要考虑特征空间的稀疏性问题，词与词可能的组合结果很多，下面仅以两个词的组合为例进行介绍，根据统计，一个网络信息检索原型系统包含的两词特征项就达10亿项，而且许多词之间的搭配是没有语义的，绝大多数组合在实际文本中出现频率很低，这些都是影响多词级别索引实用性的因素。
3.2.3语义与语甩级别语义特征如果我们能获得更高语义层次的处理能力，例如实现语义级别（Semantic Level）或语用级别（Pragmatic Level）的理解，则可以提供更强的文本表示能力，进而得到更理想的文本分类效果。
然而在目前阶段，由于还无法通过自然语言理解技术实现对开放文本理想的语义或语用理解，因此相应的索引技术并没有前面的几种方法应用广泛，往往应用在受限领域。
在自然语言理解等研究领域取得突破以后，语义级别甚至更高层次的文本索引方法将会有更好的实用性。
3.2.4 汉语的语义特征抽取1.汉语分词汉语是一种孤立语，不同于印欧语系的很多具有曲折变化的语言，汉语的词汇只有一种形式而没有诸如复数等变化。
另一方面，汉语不存在显式（类似空格）的词边界标志，因此需要研究中文（汉语和中文对应的概念不完全一致，在不引起混淆的情况下，文本未进行明确区分而依照常用习惯选择使用）文本自动切分为词序列的中文分词技术。
中文分词方法最早采用了最大匹配法，即与词表中最长的词优先匹配的方法。
根据扫描语句的方向，可以分为正向最大匹配（Maximum Match，MM）、反向最大匹配（Reverse Maximum Match，RMM），以及双向最大匹配（MM））等多种形式。

编辑推荐

《信息内容安全管理及应用》：信息采集的原理及方法文本、图像、视音频的特征提取技术面向内容安全的分类原理有方法内容安全的典型应用案例

第一图书网, tushu007.com

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:http://www.tushu007.com