第 13 章

# 内容安全技术

**内容提要**

“内容”与“信息”有联系也有区别，一般可以认为内容是人们可感知的信息。虽然在概念上有“广义”与“狭义”之分，内容安全主要是指数字内容的复制、传播和流动得到人们预期的控制，而内容安全技术就是指实施这类控制的技术。当前，内容安全技术主要用于不良内容传播控制、数字版权侵权控制、敏感内容泄露等方面的控制与监测，在监管对象上已经从主要以文本内容为主过渡到以文本和多媒体监管并重的局面。本章重点介绍了面向文本内容的内容过滤、话题发现和跟踪技术，也简要介绍了多媒体内容安全技术。

**本章重点**

- 内容安全的概念；
- 文本过滤的基本方法；
- 话题发现和追踪的基本过程；
- 内容安全分级监管；
- 多媒体内容安全技术的基本内容。

## 13.1　内容安全的概念

在信息科技中，“信息”和“内容（Content）”的概念是等价的，它们均指与具体表达形式、编码无关的知识、事物、数据等含义，相同的信息或内容分别可以有多种表达形式或编码。信息和内容的概念也在一些特别的场合略有区别。一般认为，内容更具“轮廓性”和“主观性”，即在细节上有些不同的信息可以被认为是相同的内容，人们在主观上没有感觉到这些细节的不同对理解或识别内容有多大的影响。而信息具有自信息、熵、互信息等概念，可以用比特（Bit）、奈特（Nat）或哈特（Hart）等单位衡量它们数量的多少，因此一般认为信息更具“细节性”和“客观性”。在细节并不重要的场合下，内容往往更能反映信息的含义，也可以认为内容是人们可感知的信息或较高层次的信息，因此多个信息可以对应一个内容。

**例 13.1**　图像压缩编码中的信息与内容。

可以通过压缩编码减小一个数字图像的存储尺寸。当前常用的图像压缩编码方式是 JPEG 压缩，产生的图像文件为 JPG 文件。大量的图像压缩工具可以将其他格式的图像压缩为 JPG 文件，JPG 格式的图像也可以进一步压缩。设原图像编码文件为 A.TIF，它被压缩为 B.JPG，由于 JPEG 压缩是有损压缩，为了节省存储空间，压缩后的编码省去了一些高频信息，因此 A.TIF 和 B.JPG 表达的信息是不同的，但如果压缩程度不是太高，可以认为它们表达的内容是相同的。在现实中，人们会认为照片上的内容相同，只不过一个尺寸大些、一个尺寸小些。

随着数字技术、计算机网络和移动网络的发展，内容的复制和流动变得更加容易，这在一些情况下是人们需要的，但在另一些情况下，内容的肆意复制、传播和流动危害了一些组织和个人的利益，因此人们希望实施一定的控制和监管，获得可控性。显然，实施这类控制的依据是何种内容或信息在被复制、传播或流动，因此，内容或信息本身的含义直接与安全策略关联在一起，这也要求信息安全策略的执行需要预先识别内容或信息。内容安全就是指内容的复制、传播和流动得到人们预期的控制和监测。这里“内容”一词的定义主要基于以下 3 个方面：

（1）前述内容与信息的细微差别；

（2）当前国际上将数字视频、音频和电子出版物等称为数字内容；

（3）一些文献中的“内容”专指应用层或应用中的数据和消息。

当前，对内容安全的危害和需求主要体现在以下 3 个方面。

1）数字版权侵权及其控制

数字内容产业主要指影视和音乐的数字化制作和发行行业，包括 VCD、DVD、网络视频和 MP3 音乐的制作、发行企业等，涉及现代社会中的几乎每一个人，但是，数字视频和音频的盗版和非授权散布沉重打击了数字内容产业，也迟滞了网络技术在这一

行业中的应用。人们逐渐发现，对数字版权的侵权仅依靠法律手段是不够的，数字内容制作企业、内容制作者及管理部门迫切需要有遏制版权侵权的技术手段。

2）不良内容传播及其控制

不良内容的肆意传播是另外一个与内容相关的安全问题。在互联网上，任何拥有合法网络地址的团体或个人都可以发布内容，任何知道电子邮件接收地址的人均可以向该地址发送电子邮件，在各种动机的驱动下，造成了不良内容大量传播、垃圾邮件泛滥的情况。显然，政府、学校和邮件服务管理者希望阻止这些内容的传播或监控其发展。

3）敏感内容泄露及其控制

大多数工作环境在安全通信管理方面是松散的。例如，由于工作需要，政府、企业和科研单位允许工作人员对外收发电子邮件、上网并传输文件。这不免存在敏感信息泄露的问题，其中，敏感信息主要包括保密文件和与知识产权相关的资料等。为了制约这类现象，信息安全的管理者希望根据工作人员对外传输或接收的内容对网络通信进行控制。

4）内容伪造及其控制

随着数字多媒体技术的发展，出现了大量的数字媒体内容制作、加工和编辑工具。一方面，数字内容的制作者（尤其是影视行业）用这些工具提高了数字内容的质量；另一方面，这些工具也为数字内容造假提供了可能，使得逼真的假造内容屡次出现，不但对公众起到误导作用，也往往使得普通数字内容作为法律证据的效力遭到质疑。显然，人们需要能够核实数字内容的真伪，并且这种核实也能针对普通数字内容进行（即进行所谓的内容盲取证），而不依赖于这个内容曾经被数字签名。

内容安全技术就是获得以上控制和监管能力的技术，它可以分为被动与主动两类（见图 13.1）。被动内容安全技术不预先处理被监管的内容，它通过分析获得的内容本身判断内容的性质，并实施相应的控制策略。主动内容安全技术对被监管的内容先进行预处理，在内容中添加验证信息，在以后的监管中，它通过分析所获得内容中添加的验证信息来判断内容的性质，并实施相应的控制。后一种预处理主要包括对内容添加分级标志、数字签名、数字水印等可识别信息，它们方便了对内容性质的判定。一般认为，被动内容安全技术使用起来更方便，但主动内容安全技术的可靠性和准确性更高。

从国内外出版的文献看，内容安全技术也可以分为广义的内容和狭义的内容安全技术两类。广义内容安全技术指与内容及其应用特性相关的所有信息安全技术，包括数字版权保护、数字水印、多媒体加密、内容取证（包括前面提到的内容盲取证、第 9 章介绍的数字指纹与追踪码、第 6 章介绍的脆弱水印等）、内容过滤和监控、垃圾邮件防范、网络敏感内容搜索、舆情分析与控制、信息泄露防范等。狭义的内容安全技术主要包括广义内容安全技术中涉及内容搜索、过滤和监控的部分，如网络多媒体内容的非授权散布监控、内容过滤和监控、垃圾邮件防范、网络敏感内容搜索、舆情分析与监测等。

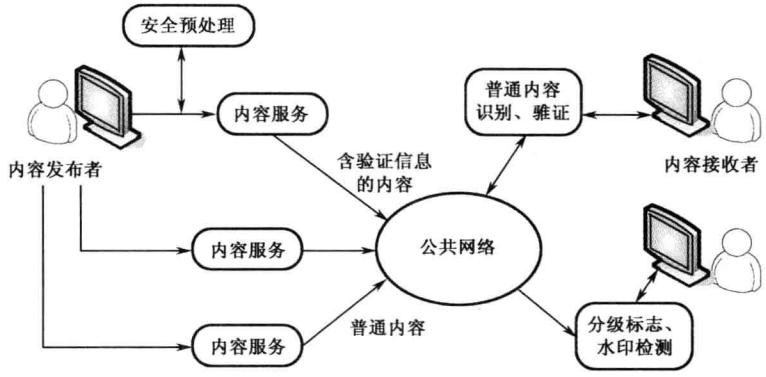


图 13.1　被动与主动内容安全技术实施环境

由于前面已经介绍了数字版权管理、数字水印、内容认证、数字指纹与追踪码等技术或概念，本章以介绍狭义内容安全技术为主，它的核心部分包括文本过滤、话题发现和跟踪、内容分级监管等，但由于当前这些技术正朝着同时监管多媒体内容的方向发展，本章也将简要介绍一些主要的多媒体内容安全技术。当前，面向网络环境的内容安全技术越来越受到重视，它普遍基于网络流量监测与网络内容搜索等技术获取信息，由于 8.2 节介绍了一些有关网关与流量监视的概念和技术，而网络搜索是常用的计算机技术（它一般指系统在后台下载和分析内容、在前台向用户提供查询），因此，本章并不展开介绍这些基础部分，而是重点介绍与内容分析、处理紧密相关的内容安全关键技术。

## 13.2　文本过滤

8.1 节介绍了防火墙技术，它可能在多个网络层次上实施过滤，一般基于地址或端口的过滤在基于应用数据的过滤之前执行。文本是最常出现的应用层数据形式之一，文本过滤不仅可用于防火墙，也适用于阻止垃圾邮件、防范信息泄露、搜索网络敏感内容和舆情控制等，这些应用也需要从截获或搜索到的数据中发现特定的文本内容或对文本进行分类，执行相应的安全策略。本节描述的文本过滤属于被动的内容安全技术，将在 13.4 节和 13.5 节分别介绍基于内容安全分级管理和数字水印的主动内容安全技术。读者应能理解主动内容安全技术和被动内容安全技术的不同。

最简单的文本过滤方法采用关键词查找，通过文字串匹配算法确定文本是否包含某些特定的词，进而确认文本类别。当前，研究人员提出了很多串匹配算法[130]，提高了匹配效率，但是，由于各个关键词的重要程度不同或它们之间的关联方式不同，发现它们的存在往往不能判断文本的特性。典型地，当系统发现一个文本包含一些不良词时，往往不能准确判断文章是从正面或从反面的角度使用这些字词，为了实施正确分类，系统可能需要知道不良词出现的频率、它们之间及它们与其他词之间的关联。

针对仅使用关键词匹配的不足，人们自然想到用更全面的特征判断文本内容的类

型。20 世纪 60～70 年代，Salton 等人[131, 132]提出了文本的向量空间模型，对文本过滤技术产生了深远的影响。这个模型将文本看作由不同的词条组成的高维向量($T_1,\cdots,T_N$)，根据不同的估计方法，词条$T_i$具有权重参量$W_i$，用于表示该词条对文本内容的重要程度，则($W_1,\cdots,W_N$)是 $N$ 维欧氏空间中的向量。在用于文本过滤时，一般$T_i$是经过选择的特征词条，维数 $N$ 也要按照计算能力进行控制，此时($W_1,\cdots,W_N$)也被称为特征向量，$W_i$的计算一般考虑了自然语言的特性。在以上文本表示技术基础上，典型的文本过滤方法包含如图 13.2 所示的步骤，其中，示例文本（有时也称训练文本）是用户用于生成待匹配特征的文本。在图 13.2 所示的系统中主要有 3 类技术。

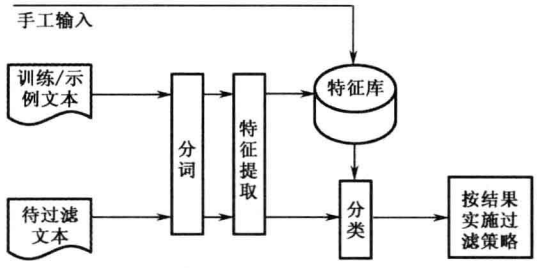


图 13.2　基于特征向量匹配的文本过滤

**1．分词**

分词是将文本语言分解为词。在英语、法语等西方语言中，空格是单词之间的分隔符号，因此计算机比较容易对西文文本分词。而中文由相互之间没有分隔符的字组成，但词仍然是表达含义的单位，一个中文词包括的字数不等，因此，中文分词的目的是要将文本文字分割成具有独立含义的词。必须特别指出，分词不但用于分解示例文本，也用于在实际过滤中分解待过滤的文本。

目前，中文自动分词的基本方法是词典分词法，它将词典中给出的词作为文本词汇分割的依据。词典是系统预先构造的，但也可以通过机器学习的方法扩充[134]，其中包含了通常意义下认为有含义的所有词条。分词算法将文本中的字串依次与词典中的词比对，如果发现当前的字串与词条相符就把字串分割出来。词典的大小关系到分词算法的效果和效率。如果词典包括的词条比较多，分词效果就会比较好，但同时也会耗费更多的时间，因此设计人员需要在这两者之间找到一个平衡点。词典的数据结构也直接关系到分词算法的效率。最典型的分词词典有以下两类组织方法。

1）整词二分法

整词二分方法的词典结构分为首字哈希表、词索引表、词典正文三级。通过对首字哈希表的哈希定位，可以在词索引表中得到以该字为首字的词条在词典正文中的位置，进而可以在词典正文中通过整词二分法进行查找定位。这种算法的数据结构简单、占用空间小，构建及维护也较简单，但由于采用全词匹配的查询过程，效率较低。

2）Trie 索引树法

Trie 索引树是一种结合多重链表的树，基于它的词典由首字哈希表和 Trie 索引树节点两部分组成，它的构造可以用图 13.3 描述，其中，入口项数是指由首字后面加字可以组成的词数。Trie 索引树词典的优点是：在分词中，在系统对被分解语句的一次扫描过程中，无须预知待查询词的长度，沿树下行逐字匹配即可。例如，若当前文本出现“大案要案……”时，无须分别假设前 1、2、3、4 个字都可能是词并分别查找，仅需要从“大”沿树下行逐字匹配，可以分别发现“大”、“大案”、“大案要案”都是词。但是，Trie 索引树的构造和维护比较复杂，存储开销也较大。

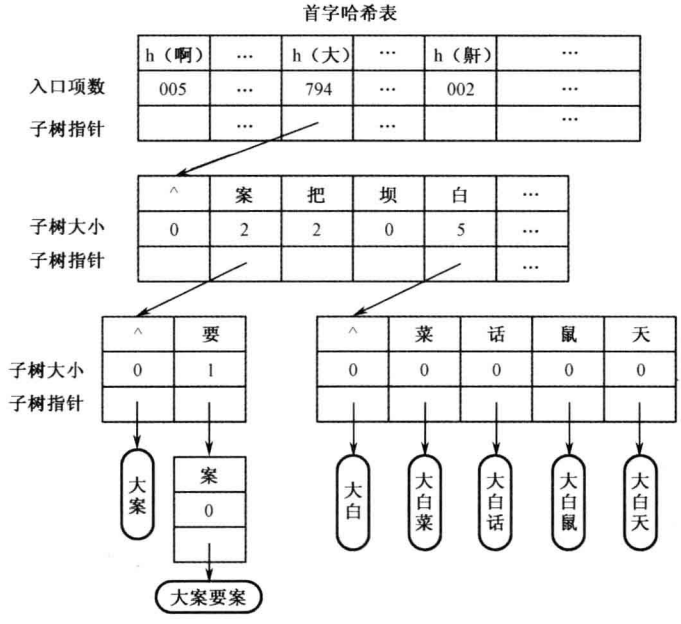


图 13.3　Trie 索引树的词典结构图

在前两类方法的基础上，研究人员提出了一些提高的中文分词方法。基于逐字二分法的词典查询是对前两种词典机制的综合。从数据结构上看，逐字二分法与前述的整词二分法的词典结构类似，但逐字二分法吸取了 Trie 索引树的查询优势，采用的是“逐字匹配”，这就在一定程度上提高了匹配效率。基于双字哈希机制的词典查询方法[135]根据汉语中双字词较多的特点，对基于 Trie 索引树的词典做出了改进，采用前两个字逐个哈希索引、剩余字串有序排列的结构，查询过程采用逐字匹配方法，这相当于使两字词以下的短语用 Trie 索引树索引，三字词以上的长词的剩余部分用线性表组织，避免了深度搜索。另外，研究人员还基于 PATRICIA 树和双数组 Trie 索引树提出了改进的词典结构。

在分词完成后，系统一般要对分词结果进行预处理，删除一些停用词，或者合并一些对定义文本性质不重要的词，如合并反复出现的人名、人称或同义词等。

### 2．特征提取

特征提取首先是指从示例文本中计算出能够表征文本特性的量。在向量空间模型中，对于词条 $T_i$，权重 $W_i$ 是其特征量，$(W_1,\cdots,W_N)$是整个文本的特征向量，它一般由前 $N$ 个最大权值组成，这 $N$ 个权值对应的词汇一般被称为特征词，可以认为它们对定义文本属性的贡献最大。当然，这 $N$ 个特征词或其一部分及它们的权值也可以由用户指定。

计算权值存在多种方法。最简单的权值计算方法被称为布尔向量表示法，它只考虑特征词在文本中是否出现，如果出现，向量中对应项为 1，如果不出现，向量中对应项为 0。这种方法易于实现，处理速度快，但在反映文章含义方面非常粗糙。好一些的方法是统计特征词条在文本中出现的频率（TF，Term Frequency）。它首先统计词条 $T_i$ 在文本中出现的次数 $N(T_i)$，然后通过不同方式归一化（如将 $N(T_i)$ 除以所有词条出现的总次数），得到 $n(T_i)$，最后将 $n(T_i)$ 作为 $W_i$。更复杂的权值计算方法包括基于信息增益（IG，Information Gain）、$\chi^2$ 统计量、互信息的方法，它们从不同角度度量了一个词条对定义文本含义的贡献程度。

对于被过滤的文本，也存在将其表示为特征向量的问题。在过滤系统对等待过滤文本进行分词并计算词条的权值后，根据特征数据库中的特征向量，可以得到由等待过滤文本在相应特征词上的权值所组成的特征向量 $(W_1',\cdots,W_N')$。

### 3．内容分类

内容分类是指过滤系统检查流经的文本、根据特征数据库判断文本属于哪一类文本的操作。在向量空间模型中，一般通过计算以上 $(W_1,\cdots,W_N)$ 与 $(W_1',\cdots,W_N')$ 的相关系数判断：当相关系数大于一个阈值时，可判断流经的文本属于 $(W_1,\cdots,W_N)$ 对应的那一类文本。

## 13.3　话题发现和跟踪

1996 年，美国国防高级研究计划委员会（DARPA）提出需要一种能自动确定新闻信息流中话题结构的技术。在随后相关的研究中，这类技术被称为话题识别与跟踪（TDT，Topic Detection and Tracking）技术[136]，它主要以网络新闻、广播和电视信息流为处理对象，将内容按照话题区分，监控对新话题的报道，并将涉及某个话题的报道组织起来，以某种需要的方式呈现给用户。总之，TDT 的主要研究目标是实现按话题查找、组织和利用来自多种新闻媒体的语言信息。随着 Internet 的普及，TDT 技术的应用意义越来越大。

话题（Topic）是话题识别与跟踪领域中的一个基本概念，它的含义与通常字面上的含义不同。在最初的研究阶段，话题与事件含义相同，一个话题指由某些原因、条件引起，发生在特定时间、地点，并可能伴随某些必然结果的一个事件。目前使用的话题概念的范围要相对宽一些，它包括一个核心事件或活动及所有与之直接相关的事件和活

动。如果一篇报道讨论了与某个话题的核心事件直接相关的事件或活动，那么就认为该报道与此话题相关。例如，搜寻飞机失事的幸存者、安葬死难者都被看作与某次飞机失事事件直接相关。因此，话题涉及某一类事件的报道。

TDT 的研究与开发主要集中在以下 5 个方面，从中不难看出，TDT 技术在运用中经历了从聚类（指在预先不知道类别的情况下将事件按照它们共同的特性归类）到分类（指在预先知道类别及其特征的情况下将事件归类）的过程。

**1．报道切分**

报道切分是指将从一个信息源获得的语言信息流分割为不同的新闻报道。一个新闻栏目通常包括很多条新闻报道，但是，这些新闻条目之间一般有一定的分割标识，或者在内容编排上有一些变化，这些都是分割的依据，而语言含义本身也是分割的基础。对于语音信号，新闻报道切分一般需要采用语音识别技术获得文字信息，因此，以下 4 项后继技术的输入一般仅为文本。

**2．新事件识别**

新事件识别的目标是识别出以前没有报道过的新闻话题。当前，新事件识别技术采用了类似于文本过滤的方法，它一般也用特征提取算法得到事件报道的特征向量，这些特征向量组成了事件特征库。对于一个新的报道，识别系统计算它的特征向量并比较特征库中的向量，确定报道的事件是否已经存在。在不存在的情况下，系统将这篇报道描述的事件作为一个新事件，并对事件特征库进行扩充。

**3．报道关系识别**

报道关系识别是对两篇报道做出分析，判断它们描述的新事件是否在讨论同一个话题。报道关系识别技术也与前面介绍的文本过滤技术有类似之处，当前普遍采用特征向量比较的方法，相互比较的特征向量来自被分析的两篇报道，对于特征向量相似的，系统认为两篇报道在讨论同一话题。通过这种方法可以将报道同一个话题的事件聚集在一起。

**4．话题识别**

话题识别的目的是将新闻报道归入不同的话题类（或称为话题族）。实际上，以上 3 种技术都是为最终的话题识别做准备的，是话题识别的前期步骤，最后通过报道关系识别，识别系统已经将报道同一个话题的大量新的事件聚集在一起，接下来的工作是进一步将它们整理归类并描述它们。从模式识别的角度看，话题识别可以看作对事件的聚类，因此研究人员运用了大量的聚类技术，包括增量 K-Means 聚类、Agglomerative 聚类、单遍历聚类、层次聚类算法、DBSCAN 密度聚类等技术。

**5．话题跟踪**

与话题识别可以被看作是聚类过程不同，话题跟踪可以被看作是分类过程，它是指识别出某个新闻报道是否属于某个已知话题的技术。通常，跟踪系统已经通过前期

的话题识别获得了各个话题的基本特性，通过比较新闻报道的特征，判断出新闻报道所归属的话题。通过对不同网络地址范围实施搜索，话题跟踪系统可以判断舆情的传播情况。

## 13.4　内容安全分级监管

内容安全分级监管是一种主动内容安全技术，它指在内容发布前，在内容中嵌入分级标识，随后的各种监管措施基于分级标识进行。在基于分级标签的内容分级管理框架（见图 13.4）中，分级标准处于核心地位，它约定了内容分级、生成并嵌入标签及在监控和过滤中识别标签的方法。为了管理互联网上日益泛滥的不良信息，保护儿童的身心健康，W3C 组织（www.w3c.com）推动了“Internet 内容选择平台（PICS，Platform for Internet Content Selection）”规范的制定，提出了基于分级标签的内容分级管理技术，得到了一些网络内容服务、客户端和浏览器程序开发商的支持。

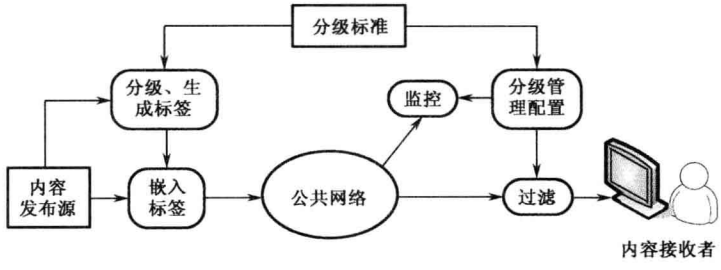


图 13.4　基于分级标签的内容分级管理框架

内容安全分级监管主要包括内容分级、生成并嵌入标签及根据识别的标签实施监管等几个环节。任何接受监管的内容必须要按照统一的要求被分级，一般一个级别包含内容类别标识和等级标识，如“暴力 2 级”。标签不仅记录以上内容类别和等级信息，一般还包括分级标准颁布组织、时间戳等标签信息。PICS 规范没有给出标签防伪技术，但实用系统不难进行这方面的扩展。当前，PKI 和数字签名技术被应用到了标签生成中，这样，标签的真实性和被保护内容的完整性均可以得到保证。标签的嵌入与保护的具体文档格式相关，一般采取以下嵌入方法：① 对于常用的 HTML 格式，可利用 HTML 格式的 META 标记，将标签嵌入在 HTML 文件头中；② RFC-822 约定了 Internet 中一些文本消息的格式，它们涉及电子邮件、HTTP、FTP、GOPHER、USENET 等应用协议，可以利用这类消息头存储标签，另外，PICS 定义了 HTTP 协议扩展，允许 Web 服务器处理获得分级标签的请求；③ 由用户发出请求，再由可信的第三方——“标签局（Label Bureau）”针对特定的 URL 向用户提供标签。

**例 13.2**　Web 敏感内容服务商通过 PICS 标签的自律措施。

在 IE 4.0 浏览器的 Internet 选项中有一个内容设置功能，它可以防止浏览一些受限

制的网站。之所以浏览器能自动识别某些网站是否受限制，是因为在网站网页的 META 标记中已经设置好了该网站的 PICS 级别，而该级别是由美国原 RSAC（Recreational Software Advisory Council）评定的，该组织这样做的目的是保护儿童不受不良信息的危害。1999 年，RSAC 并入 ICRA（Internet Content Rating Association），后者是美国家庭在线安全协会（Family Online Safety Institute）的下属机构。一个包含 PICS 标签的 META 标记如下：

< META http-equiv="PICS-Label" content="(…1996.04.16T08:15-0500，r(n 0 s 0 v 0 l 0))">

## 13.5　多媒体内容安全技术简介

近年来，互联网上以图像、视频和音频为代表的多媒体内容正以惊人的速度增长，出现了以视频新闻、播客、视频下载、网络电视、视频广告、流媒体、P2P、歌曲下载等为传播方式的网络多媒体内容产业，用户与日俱增。目前，基于互联网的图像与视频、音频节目内容已经成为网络文化的重要组成部分，对人们文化消费和意识形态的影响越来越大。与此同时，由于缺乏技术监管手段，淫秽色情、暴力血腥、恶搞、变态、反动有害、盗版的多媒体内容正在通过互联网快速传播，造成了十分恶劣的影响，使多媒体内容安全受到更广泛的重视。

多媒体内容安全技术在目的和实施框架上与面向文本的内容安全技术类似，它主要通过监管多媒体内容的散布情况来制约不良或盗版内容的传播。但是，由于多媒体内容以信号编码的形式存在，也是数字电影和音乐的发售形式，因此，多媒体内容安全技术包括了大量的多媒体编解码、信号处理和模式识别等技术，也更多地与版权保护联系在了一起。本节仅简单介绍多媒体内容安全技术。

### 1．被动多媒体内容安全技术

被动多媒体内容安全技术通过检测或搜索未经过相应安全预处理的网络多媒体内容，确定不良、盗版内容的传播、散布情况，或者识别伪造的内容，并执行可能的处置。当前，被动多媒体内容安全技术主要包括网络多媒体识别、内容伪造取证等技术，前者预先知道网络多媒体的内容或相关特征，需要发现其散布情况，后者不知道网络多媒体的内容，需要通过专门的验证方法判断内容的真实性，甚至定位篡改位置。

已经出现了多种网络媒体的识别方法。当前一些简单的监管系统主要采用网页分析与网页信息抽取的方法判断多媒体的违规散布，即通过关键字搜索检测媒体散布的线索，但这样做的可靠性不高，违法者容易通过修改媒体内容的名称等方法避开监管，因此还需要分析媒体本身。研发人员已经在多媒体内容识别方面采用了各种特征提取和分类手段确定视频或图像的类型，已经在色情内容识别等方面获得了一些有效的方法，但是，由于音视频内容一般尺寸较大，若要对普通内容进行识别，直接做常规匹配的难度较大。内容杂凑（Content Hash）是一种新出现的多媒体内容发现技术[137]，它也称为感

知杂凑（Perceptual Hash）或指纹化（Fingerprinting），这类技术首先提取待发现内容的基本特征数据，前者一般尺寸较大，而得到的特征数据具有小尺寸和低碰撞性的特点，在这方面类似于密码技术中的杂凑值，但它对不同的编码格式不敏感，因此，网络搜索系统可以基于内容杂凑去识别搜索到的多媒体，避免了采用大数据作为匹配依据的复杂情况。内容杂凑是数字多媒体的稳定特征，在不显著改变内容的情况下，内容侵权者难以实质性地更改其基本信息，因此难以避开监管。在我国，与内容杂凑类似的技术也称为零水印。当前，相关研究普遍试图发现与内容更相关且性质更加稳定的统计特征，基于这些特征计算内容杂凑并形成高效的查询和过滤。2006 年开始，美国一些企业已采用基于内容杂凑的搜索技术监管其生产的数字内容散布情况，一些国外网站陆续采用了基于内容杂凑的技术限制受版权保护的数字内容上传。

另一类典型的被动多媒体内容安全技术是数字内容盲取证[138]。由于普通多媒体内容本身也存在一些制约关系，如一幅图像中的太阳光照角度是相同的，并且在物体透视效果上满足一定的规律，因此，可以通过分析这些约束条件的满足情况发现篡改痕迹，并识别内容伪造及其区域。

**2. 主动多媒体内容安全技术**

主动多媒体内容安全技术主要包括基于分级标签和数字水印嵌入这两类技术，而正如第 6 章已经介绍的，数字水印又可分为鲁棒水印和脆弱水印，它们可分别面向版权保护和内容伪造识别。

本章前面已经介绍了基于分级的网页内容安全技术，它显然是一类典型的主动内容安全技术，它的基本原理也可以用于多媒体内容安全技术。但是，通过前面的描述也不难看出，分级标签的嵌入受到文件格式的制约，另外，违法者可以架设自己的网站发布非授权的内容，这些网站不会支持使用分级标签。而鲁棒水印技术弥补了以上不足，鲁棒水印与合法发布的多媒体内容紧密地结合，违法者难以在不显著破坏多媒体感知质量的情况下消除水印，因此水印成了“黏合力强”的标签。虽然鲁棒水印可以作为分级标签，但当前更多地用它表示版权所有者的信息或内容购买者的信息，在后一种情况下，水印通常也称为数字指纹（Digital Fingerprint）。在版权保护的应用中，版权管理部门或司法机构可以通过验证水印维护版权所有者的利益，也可以通过验证数字指纹检举非授权传播者，这种技术在一定程度上能够制约内容使用者肆意违规散布内容。

脆弱水印是一种主动内容取证技术，其原理可以参阅第 6 章。相比于内容盲取证技术，脆弱水印验证的正确率较高，也能确定篡改位置，但盲取证适用于未经安全预处理的内容。

## 13.6　小结与后记

本章介绍了数字内容安全技术的基本概念、主要技术方法和应用场景。数字内容安

全技术涉及的范围较广，但其核心是针对数字内容的制作和散布进行相应的监管和控制。内容安全技术分为被动内容安全技术与主动内容安全技术两类，前者不事先预处理被监管的内容，它通过分析获得的内容本身判断内容的性质，实施相应的控制，后者对被监管的内容进行安全预处理，以后通过分析所获得内容中添加的预处理信息来判断内容的性质，并实施相应的控制。文本过滤、多媒体内容发现与分析是主要的被动内容安全技术，内容安全分级监管与数字水印等构成了主要的主动内容安全技术，这些技术已经被实际应用于监管不良内容和盗版制品的非法散布。

当前，数字内容安全需求仍在增长，内容安全技术也在迅速发展。随着多媒体内容安全技术的重要性日益提高，预计将会有更多的新技术、新方法被运用到这一领域中来。

## 论述与思考

1. 请讨论“信息”与“内容”的关系。
2. 被动内容安全技术和主动内容安全技术分别指什么？各有什么优劣？
3. 在向量空间模型下如何构造一个文本过滤系统？
4. 话题发现和追踪技术的过程是什么？聚类和分类分别在哪个环节用到？
5. 请分别简述内容杂凑和数字水印在多媒体内容安全技术中的作用。