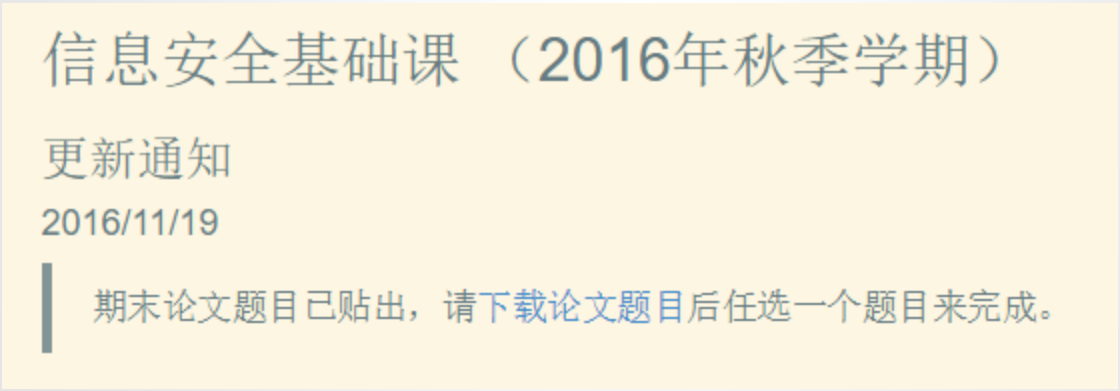
- ✓ http://qiqi789.github.io/teaching/info-security/

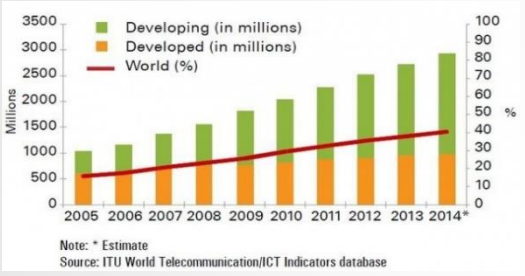


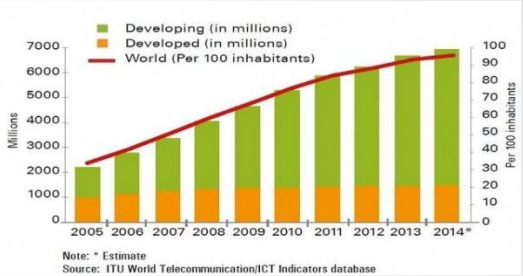
第10章 信息内容安全

- 信息内容安全概述
- 信息内容获取技术
- 信息内容识别与分析
- 信息内容控制和管理
- 信息内容安全应用
- 本章小结

信息内容安全概述

全球互联网发展状况





- 到2014年年底，全球互联网用户数量将达到30亿，占全球人口总数的约40%。在这些互联网用户中，2/3来自发展中国家
- 到2014年年底，全球手机用户数量将达到约70亿，普及率为96%，接近全球人口总数。其中超过50%约36亿来自亚太地区

数据来自国际电信联盟(以下简称“ITU”)《2014年信息与通信技术》

我国互联网发展状况

6.18亿 互联网用户

5.32亿 即时通信户

4.9亿 搜索用户

4.5亿 网络音乐用户

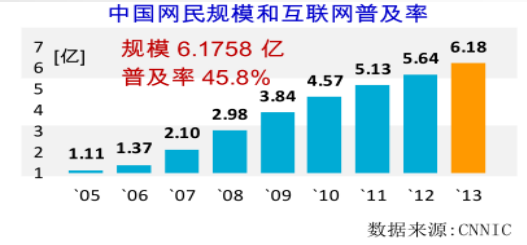
4.2亿 网络视频用户

3.02亿 电子商务用户

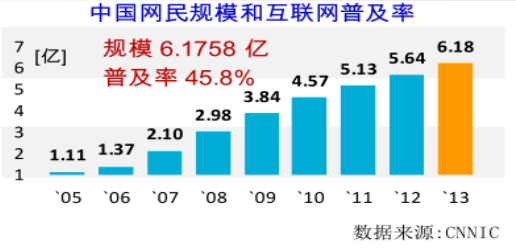
数据来源于CNNIC - 2014 年 5 月

我国互联网发展状况

网民人数：6.18亿

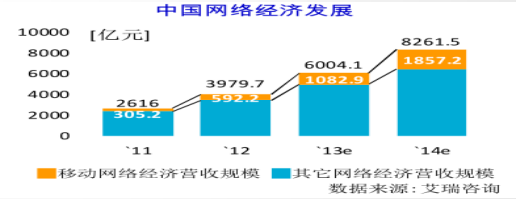


手机网民：5亿



网站数量：320万

网络经济规模： 8261亿



数据来源于CNNIC - 2014 年 5 月

互联网环境特点

- 便捷性
- 即时性
- 自由性
- 开放性
- 虚拟性
- 交互性
- …

■ 网络俨然已成为和现实世界并存的虚拟世界，给人们的生活方式、自由交往和沟通带来极大的便利。

■ 互联网上信息内容的非法传播和利用将会对社会稳定和国家安全具有较大的影响。

信息内容安全的重要性

- 2007年，胡锦涛总书记就强调要加强网络文化建设和管理。
- 2013年，习近平总书记在《中共中央关于全面深化改革若干重大问题的决定》：
  - 随着互联网媒体属性越来越强，网上媒体管理和产业管理远跟不上形势发展变化。
  - 面对传播快、影响大、覆盖广、社会动员能力强的微博客、微信等社交网络和即时通信工具用户的快速增长，如何加强网络法制建设和舆论引导，确保网络信息传播秩序和国家安全、社会稳定，已经成为摆在我们面前的现实突出问题。

信息内容安全已经成为国家信息安全保障建设的一个重要方面。

信息内容的概念

- 1995年，西方七国信息会议首次提出内容产业(Content Industry)的概念。
- 1997年，美国发布《北美产业分类系统》中，提出使用信息内容产业。
- 1996年，欧盟提出“INFO 2000计划”给出了信息内容产业内涵：
  - 产业范围：制造、开发、包装和销售信息产品及其服务的产业。
  - 主要表现形式：媒介的印刷品（书报杂志等）、电子出版物（联机数据库、音像服务、光盘服务和游戏软件等）和音像传播（影视、录像和广播等）。
  - 主要特点：数字化、多样性、易复制、易分发、交互性等。

信息内容产品的发展规模

- 2003年，全球内容产业年增长率达33%，正在成为全球经济最具活力、发展最快的新的增长点。
  - 美国内容产业年产值达7000亿美元，成为美国第一大出口产业，年增长率40%。
  - 欧盟内容产业年产值4300亿欧元，已超过电信和IT制造业。
  - 日本年产值为1600亿美元，超过钢铁业的两倍，相当于汽车业的一半。

信息内容安全概念

- 方滨兴院士对信息内容安全的定义：
  - 对信息真实内容的隐藏、发现、选择性阻断。
  - 解决的问题：发现隐藏信息的真实内容、阻断所指定的信息、挖掘所关心的信息。
  - 主要技术手段：信息识别与挖掘技术、过滤技术、隐藏技术等
- 李建华教授对信息内容安全的定义：
  - 研究如何计算从包含海量信息且迅速变化的网络中，对与特定安全主题相关信息进行自动获取、识别和分析的技术。
  - 根据所处的网络环境，又被称为网络内容安全(Network Content Security)。

信息内容安全概念

- 信息内容安全是指信息内容的产生、发布和传播过程中对信息内容本身及其相应执行者行为进行安全防护、管理和控制。
- 信息内容安全的目标是要保证信息利用的安全，即在获取信息内容的基础上，分析信息内容是否合法，*确保合法内容的安全，阻止非法内容的传播和利用*。

非法内容概念

- 2000年颁布的《互联网信息服务管理办法》第十五条中有相关的规定：
  - 危害国家安全，泄露国家秘密，颠覆国家政权，破坏国家统一的。
  - 损害国家荣誉和利益的；煽动民族仇恨、民族歧视，破坏民族团结的；破坏国家宗教政策，宣扬邪教和封建迷信的。
  - 散布谣言，扰乱社会秩序，破坏社会稳定的。
  - 散布淫秽、色情、赌博、暴力、凶杀、恐怖或者教唆犯罪的。
  - 侮辱或者诽谤他人，侵害他人合法权益的。
  - 含有法律、行政法规禁止的其他内容的。

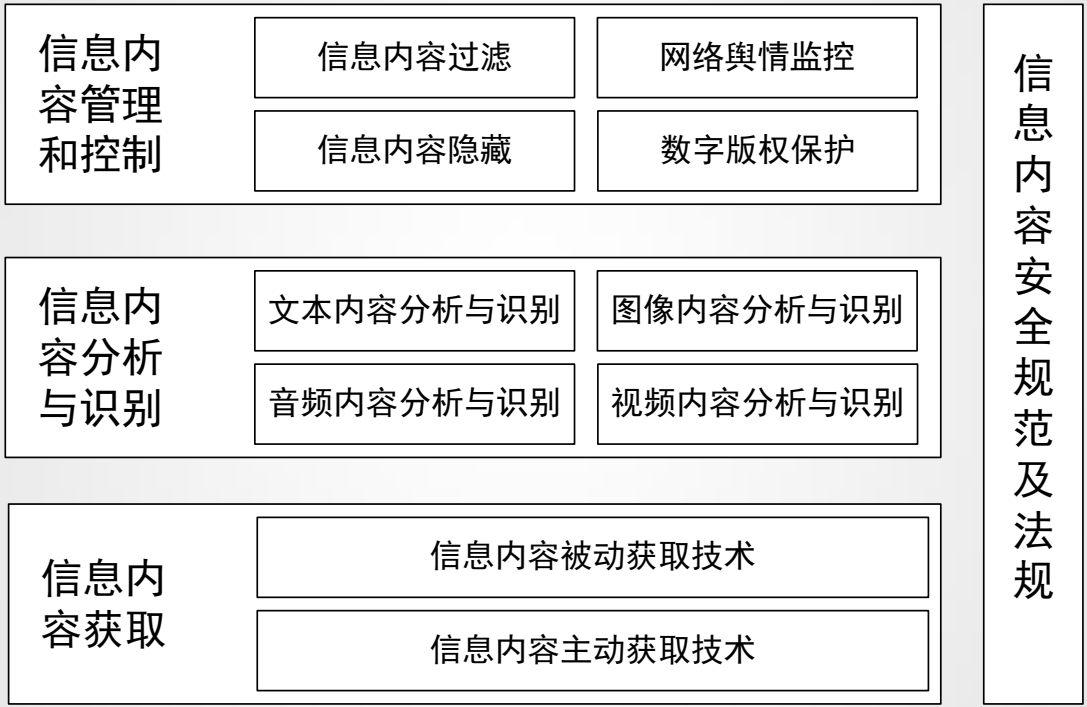
信息内容安全威胁

- 传统的信息安全威胁，如信息内容泄露、篡改、破坏、黑客攻击、计算机病毒等
- 互联网上各种不良信息内容泛滥
- 互联网上垃圾信息内容严重过载
- 互联网不良信息内容的传播和利用
- 互联网中信息内容侵权行为猖獗
- …

信息内容安全保护内涵

- 政治性方面：防止来自国内外反动势力的攻击、诬陷和西方的和平演变图谋；
- 健康性方面：剔除反动、暴力和黄色内容等；
- 保密性方面：防止国家和企业机密被窃取、泄露和流失；
- 隐私性方面：防止个人隐私被盗取、倒卖、滥用和扩散；
- 产权性方面：防止知识产权被剽窃、盗用等。
- 破坏性方面：防止病毒、垃圾邮件、网络蠕虫等恶意信息耗费网络资源。

信息内容安全体系架构


信息内容管理和控制
信息内容过滤
网络舆情监控
信息内容隐藏
数字版权保护
信息内容分析与识别
文本内容分析与识别
图像内容分析与识别
音频内容分析与识别
视频内容分析与识别
信息内容获取
信息内容被动获取技术
信息内容主动获取技术
信息内容安全规范及法规

信息内容安全和信息安全关系

- 信息安全层次模型（方滨兴院士）

信息内容获取技术

信息内容获取技术

- 主动获取技术：主动获取技术通过向网络注入数据包后的反馈来获取信息
  - 接入方式简单，能够获取更广泛的信息内容。
  - 会对网络造成额外的负担。
- 被动获取技术：被动获取技术是在网络出入口上通过镜像或旁路侦听方式获取网络信息
  - 接入需要网络管理者的协作。
  - 获取的内容仅限于进出本地网络的数据流，但不会对网络造成额外流量。

信息内容主动获取技术

- 在互联网发展初期，网站相对较少，信息查找比较容易。
- 伴随互联网爆炸性的发展，网络用户难以查找所需的资料信息。

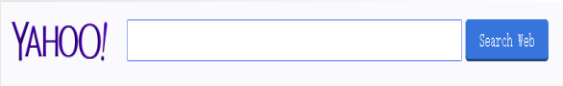
如何找到所需要的信息？

搜索引擎技术应运而生！

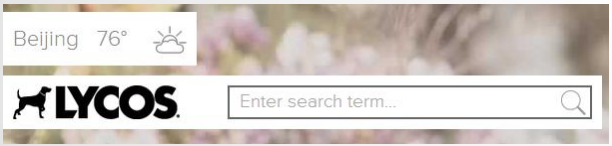
常用的搜索引擎


Bai du 百度
新闻 网页 贴吧 知道 音乐 图片 视频 地图 百科 文库 更多>>
百度一下


Google
Google.com.hk


YAHOO!
Search Web


搜狗搜索
新闻 网页 微信 问问 图片 视频 音乐 地图 百科 购物 更多>>
搜狗搜索
微信搜索 输入法 浏览器 网址导航 游戏中心


Beijing 76°
LYCOS
Enter search term...


360搜索+
新闻 网页 问答 视频 图片 音乐 地图 良医 更多
搜索一下

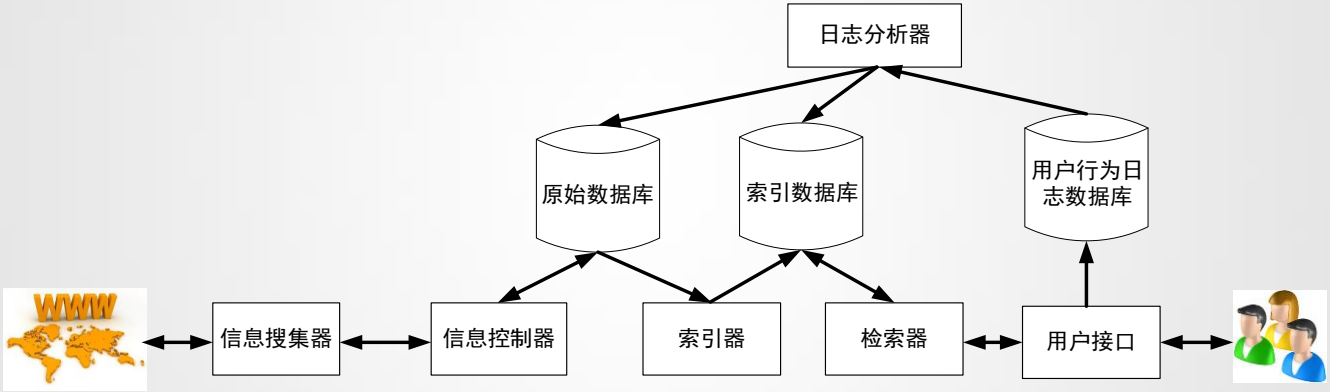
搜索引擎概念

- 搜索引擎(Search Engine)是一种在Web上应用的软件系统，它以一定的策略在Web上搜集和发现信息，在对信息进行处理和组织而建立数据库，为用户提供Web信息查询服务。
- 搜索引擎后台通过爬虫程序遍历Web，同时下载和存储分布在Web上的信息，并建立相应的索引记录；前端为用户提供网页界面，接受用户的查询请求，根据建立的索引按照一定的排列顺序为用户提供信息检索服务。

搜索引擎分类

- 根据工作原理，搜索引擎可分：
  - 全文搜索引擎(Full Text Search Engine)
  - 目录式搜索引擎(Directory Search Engine)
  - 元搜索引擎(Meta Search Engine)
- 根据搜索范围，搜索引擎可分为：
  - 综合搜索引擎
  - 垂直搜索引擎

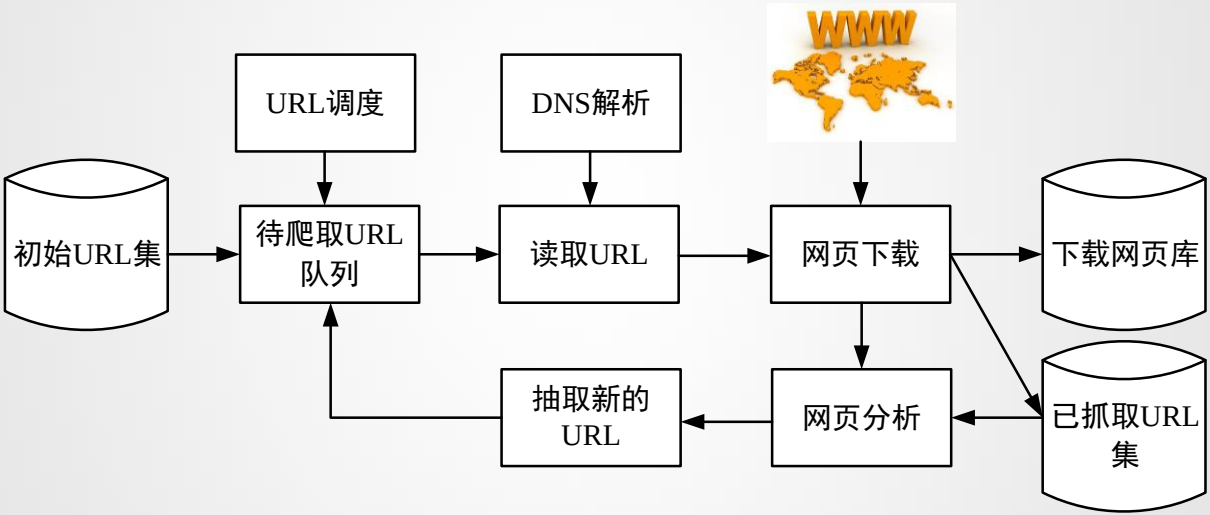
搜索引擎体系结构



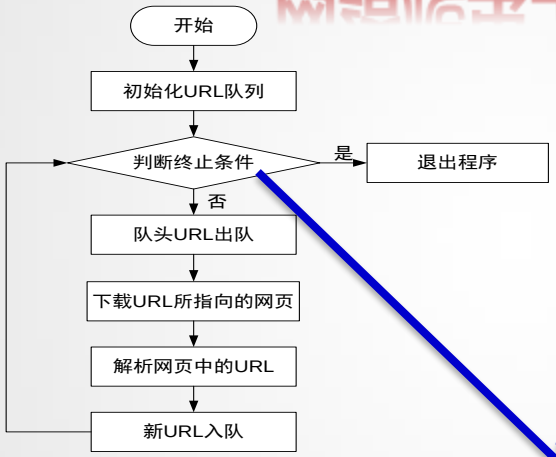
搜索引擎三段式工作流程

- 网页搜集
  - 批量搜集，增量搜集；搜集目标，搜集策略
- 预处理
  - 关键词提取、网页消重、链接分析和索引构建
- 检索服务
  - 查询方式和匹配、结果排序以及文档摘要生成

网络信息抓取技术原理


URL调度
DNS解析
WWW
初始URL集
待爬取URL队列
读取URL
网页下载
下载网页库
抽取新的URL
网页分析
已抓取URL集

网络爬虫工作流程




```
单机抓取算法
DS：
Url_QUEUE   uqueue;
History_LIST  hlist;
PROCEDURE:
Crawler(seed_url)// seed_url是起始URL队列集合
{
    in_queue (uqueue , seed_url);
    while （U=out_queue(uqueue) //从uqueue队列中移除url地址
    {
        wpage=http_get(u);   // 下载网页
        save wpage;     // 保存网页
        for each url in wpage  // 解析网页中的URL，看是否被访问过
        {
                if  url not in hlist
                then in_queue (uqueue , url); // 未被访问加入到uqueue队列
        }
    }
}
```

URL队列为空或满足某个爬行终止条件

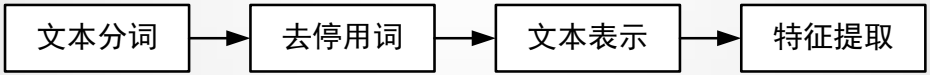
内容提纲

- 信息内容安全概述
- 信息内容获取技术
- 信息内容识别与分析
- 信息内容控制和管理
- 信息内容安全应用
- 本章小结

信息内容识别与分析

文本内容识别

- 分词技术
- 去停用词
- 文本表示
- 特征提取



分词技术

- 文本分词处理对象包括英文文本和中文文本两类，其中词是最小的、可独立运用的、有意义的语言单位。
  - 在英文文本分词中，单词被当作基本处理单元，单词与单词之间通过空格隔开，因此最为简单的方法使用空格与标点作为分隔符。
  - 中文文本中的分隔符(，：。、！？等)一般用来分割短语或句子，词与词之间没有明显的分隔符。

中文分词技术

- 将连续的字序列按照一定规范重组成有意义的分词序列的过程。
- 针对输入文字串进行分词、过滤处理，输出中文单词、英文单词和数字串等一系列分割好的字符串。

- 对文本进行有效的分词是实现人与计算机沟通的基础，也是文本内容处理的基础。
- 目前，文本分词技术已经广泛应用于信息检索、文本挖掘、机器翻译、语音识别等领域。

中文分词面临挑战

- 歧义识别问题
  - 交叉型歧义：汉字串ABC　AB，BC同时成词。如：结合/成，结/合成
  - 组合型歧义：汉字串AB，A，B同时成词。如：他/从/马/上/下/来；我/马上/就/来/了
  - 混合型歧义：同时包含交叉型和组合型歧义
- 歧义识别问题
  - 出现词典中没有的词。如人名、地名、机构名或一些新词等等。

中文分词技术总结

分类	基于字符串匹配的分词方法	基于统计的分词方法	基于知识理解的分词方法
优点	实现简单，分词速度快	不需要基于切分词典，消除歧义	能识别未登录词；消除歧义
缺点	➢分词精度与词库相关 ➢不能发现交叉型歧义 ➢不能识别未登录词	➢ 经常抽出一些共现频度高、但并不是词的常用字组 ➢ 不能识别未登录词 ➢ 识别精度差，时空开销大	➢ 知识词库复杂； ➢ 分词精度与知识库相关。

去停用词

- 停用词(stop words) ：那些在文档中出现过于频繁(如超过80%以上的文档均出现该词)而对于检索没有区分意义的词，常见的停用词包括冠词、介词、连词。
  - 优点：停用词消除可以降低存储空间和计算时间
  - 缺点：有时消除的停用词对检索是有意义的，如：“的士”中的“的”，因此有些搜索引擎直接采用全文索引(full index)
- 主要的方法
  - 查表法：建立一个停用词表，通过查表的方式去掉停用词。
  - 基于DF（文档频率）的方法：统计每个词的DF，如果超过总文档数目的某个百分比(如80%)，则作为停用词去掉。

文本表示

- 抽象文本的内涵，用特征词来表示文本内容涉及的话题
  - 特征项（关键词），特征项权重（特征项代表文本内容的程度）
- 文本用特征向量表示
  - 特征向量的计算：TF-IDF

$\mathcal{X}$ 表示一个文档集合，$\mathcal{Y}$ 表示一个可能的话题集合。d 表示一个词典的单词数量。词典里单词的索引表示为 j 。

$TF(j, \mathbf{x})$ 表示词汇 j 在文档 $\mathbf{x}$ 里出现的次数，也叫做词频。

$DF(j, y)$ 表示词汇 j 出现在其他文档里并且话题不是 y 的出现频率，也叫做文档频率，用来度量词汇 j 是否在其他话题的文档里出现频繁。

给定一个文档 $\mathbf{x}$ ，和它的话题标签 y ，可以用一个特征向量来表示，这个向量的每个元素可以用以下的公式计算：

$$\Psi_j(\mathbf{x}, y) = TF(j, \mathbf{x}) \log(\frac{m}{DF(j, y)})$$

其中 m ，为训练样本中文档的总数。上面这个表达式计算的数量叫做 TF-IDF(term-frequency-inverse-document-frequency)。

直觉上，如果词汇 j 在文档 $\mathbf{x}$ 里出现的次数很多，但是在所有其他不是话题 y 的文档里没有出现的话，那么 $\Psi_j(\mathbf{x}, y)$ 应该很大。如果是这样的话，我们趋向于相信文档 $\mathbf{x}$ 是属于话题 y 。

不良图像内容识别方法

- 相比文本信息，数字图像具有信息量大、像素点之间的关联性强等特点。
- 数字图像：空间坐标和灰度（亮度）均用离散的数字（一般是整数）表示的图像。
  - 基本元素称为像素（Pixel）。
  - 可用矩阵或二维数组来描述。

不良图像内容识别方法

512x512 uint8

	1	2	3	4	5	6
1	161	161	161	160	161	155
2	161	161	161	160	161	155
3	161	161	161	160	161	155
4	161	161	161	160	161	155
5	161	161	161	160	161	155

```
val(:,:,3) =

val(:,:,2) =
```

257x256x3 uint8

```
val(:,:,1) =

  1 至 31 列

  255  253  255  255  255  255  255  255  255  255  255  255  255  255  255
  254  188  199  197  201  202  214  202  205  202  199  199  200  201  204
  255  205  205  206  217  217  228  220  215  212  210  208  209  210  213
  255  192  193  207  221  214  223  217  212  209  208  207  208  209  212
  255  208  209  222  232  219  226  225  226  223  221  220  220  221  224
  255  212  211  218  226  219  232  234  226  223  221  219  220  221  224
  255  211  211  217  228  228  243  236  229  226  225  224  225  226  228
```

```
237  236
136  136
137  137
127  128
132  131
126  126
126  127
129  129
127  128
128  129
128  129
```

不良图像内容识别方法

- 图像内容过滤技术
  - 根据图像的色彩、纹理、形状、轮廓以及它们之间的空间关系等外观特征和语义作为索引，通过人体敏感部位相关数据进行相似度匹配。
- 构建肤色检测、纹理分析、人脸模型、人体模型
  - 提取一些颜色、纹理、形状特征等低层次特征。
  - 利用统计方法和机器学习方法进一步转化为较高层次特征。

内容提纲

- 信息内容安全概述
- 信息内容获取技术
- 信息内容识别与分析
- 信息内容控制和管理
- 信息内容安全应用
- 本章小结

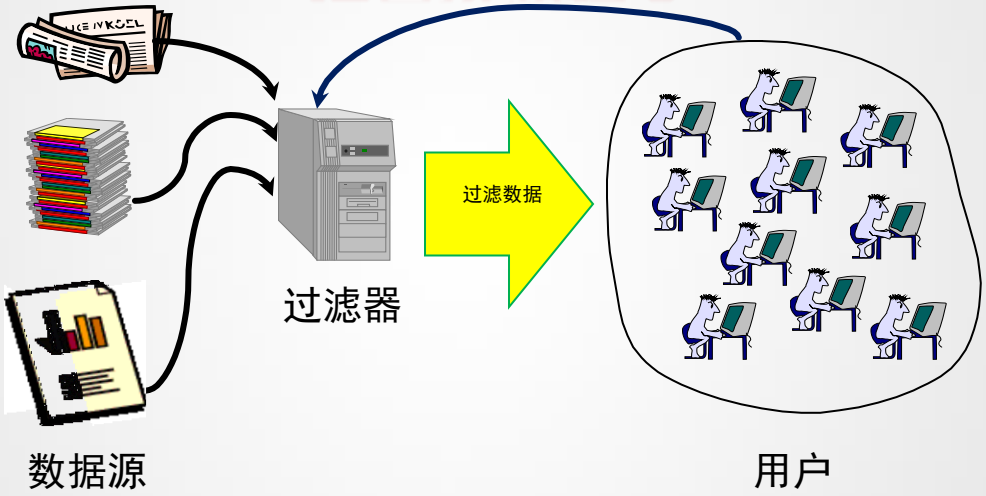
信息内容控制和管理

- 信息过滤技术
- 信息隐藏技术
- 数字水印与版权保护

信息过滤技术

- 信息过滤(Information Filtering，IF)
  - 信息的选择性传播
  - 根据用户的信息需求(User Profile)，在动态的信息流(如，Web，Email)中，搜索用户感兴趣的信息，屏蔽其他无用的和不良的信息。

信息过滤技术



IF vs. IR

	信息检索IR	信息过滤IF
需求表示	检索词(可组合)	兴趣模型
需求变化	动态	静态
信息源	静态	动态
目标	选择相关条目	过滤掉不相关的信息
了解用户	否	是
用户特点	大范围多用户短期使用	小范围少用户的长期使用

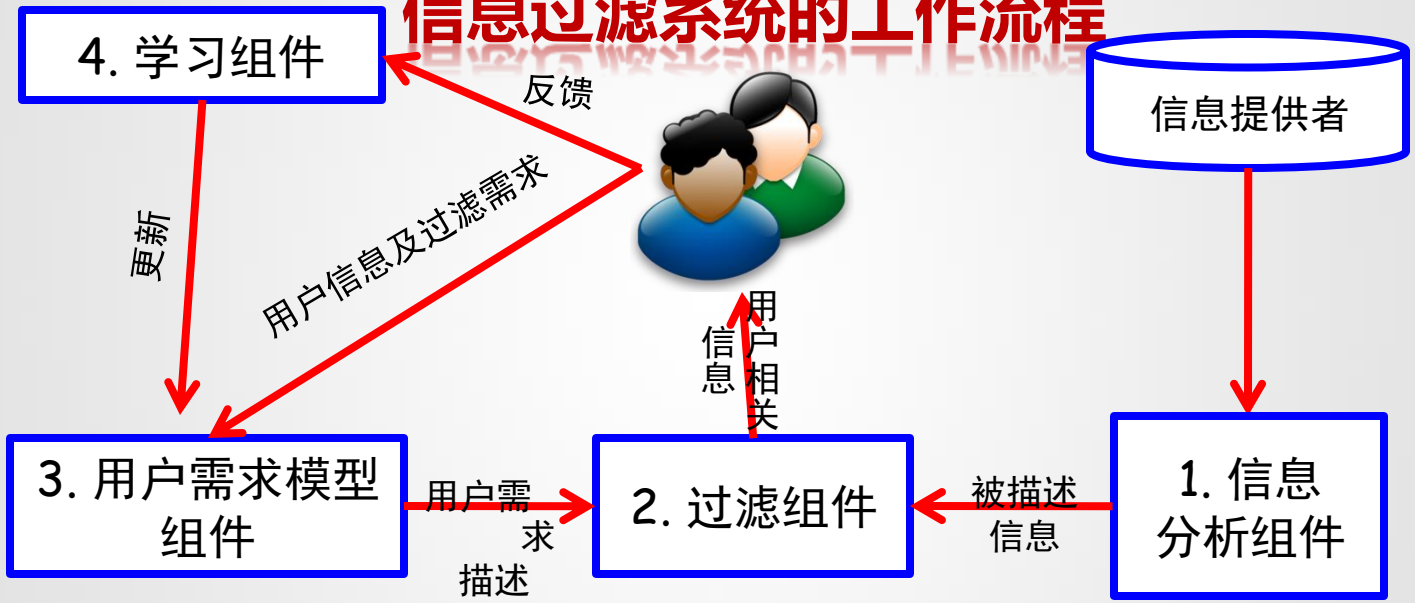
IF vs. IC and IF vs. IE

- IF vs. IC (Information Classification)
  - IF可以采用IC中的分类算法。
  - 某些场合下人们所称的IF实际就是一个IC问题，如不经过用户需求调整的垃圾邮件过滤。
  - IC中的分类范畴通常不会变化，而IF的用户需求会动态调整。
- IF vs. IE (Information Extraction)
  - IE是直接从自然语言文本中抽取事实信息，并以结构化的形式描述信息。比如抽取恐怖事件的时间、地点、人物等字段。
  - IE中不太关注相关性，而只关注相应的字段。
  - IF中要关注信息间的相关性。

信息过滤分类

- 按网络数据捕获方式：主动信息内容获取和被动信息内容获取
- 按过滤操作的位置分类：信息源过滤系统、信息过滤服务器过滤系统、用户端过滤系统
- 按过滤的方法分类：认知过滤系统、社会过滤系统、基于效用的过滤系统、基于智能代理的信息过滤
- 按获取用户知识的方式分类：显式知识获取过滤系统、隐式知识获取过滤系统、显隐混合知识获取过滤系统
- 按信息过滤的工具分类：专门的过滤软件系统、网络应用程序过滤系统、防火墙过滤系统、代理服务器过滤系统、旁路方式过滤系统

信息过滤系统的工作流程



信息过滤工作流程

- 信息分析组件：从信息提供者那里获取或收集信息(如文档、消息)，对信息进行分析并抽取其中特征信息，以适当的数据形式(如空间向量)来表示，表示结果将被输人到过滤组件中。
- 过滤组件：信息过滤系统的核心，主要用来计算信息源与用户需求模型的相关度。
- 用户需求：模型组件通过显式或隐式的搜集用户的信息生成用户需求模型，并将用户需求模型传递给过滤组件。
- 学习组件：学习组件通过发现用户兴趣变化，强化、弱化或取消现存有关用户的知识，来更新用户模型。常见的学习方法包括：观察学习、反馈学习和用户训练学习等。

信息过滤系统的关键技术

- 基于统计学理论的信息过滤系统
  - 用户需求模型和信息均可用向量空间模型表示，过滤组件采用统计算法计算用户需求模型与信息的相似性，最常见的可采用夹角余弦。
  - 学习组件要求用户决定过滤结果是否相关得到相应反馈，通过采用反馈学习方式来更新用户需求模型，主要更新用户的特征项及其权重。
- 基于知识的过滤系统
  - 在该系统中，主要基于知识论、本体论等中的相关知识，如语义网、神经网络、产品规则等，实现信息过滤。
  - 主要包括：基于规则的过滤系统、基于语义网络的过滤系统、基于神经网络的过滤系统和进化的基于遗传学算法的过滤系统等。

信息过滤系统的评估指标

- 对于集合大小为N的信息集合，实际与用户需求相关的集合大小为R。通过过滤组件进行过滤，若已经通过过滤的n条相关信息中，有r条与用户需求是相关的，即是符合用户需求模型的，则有 n-r条是与用户需求不相关的。

	相关	不相关	
已检索到	r	n-r	n
未检索到	R-r	N-n-R+r	N-n
	R	N-R	N

- 查准率(Precision) = r/n ； 查全率(Recall) = r/R
- 拒绝率(Fallout) = (n-r)/(N-R)
- 平均精度(Average Precision)是Precision-Recall曲线的面积。

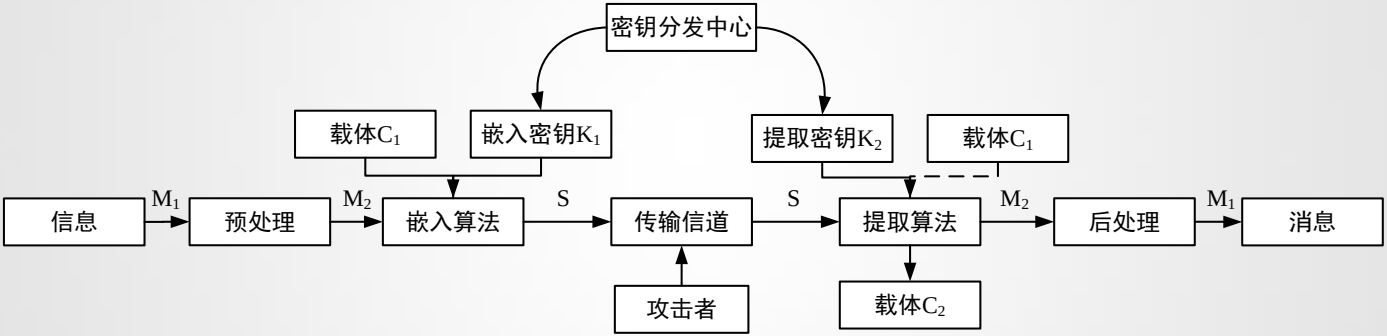
信息隐藏技术

- 信息隐藏(Information Hiding，IH)研究如何将某一个机密信息秘密隐藏于另一公开的媒介信息中。
- 公开的媒介信息
  - 数字媒体，如图像、音频、视频
  - 一般性文本信息
- IH技术的目的是隐藏秘密信息，减少第三方察觉的概率，使受攻击的风险降低，弥补加密技术的不足。

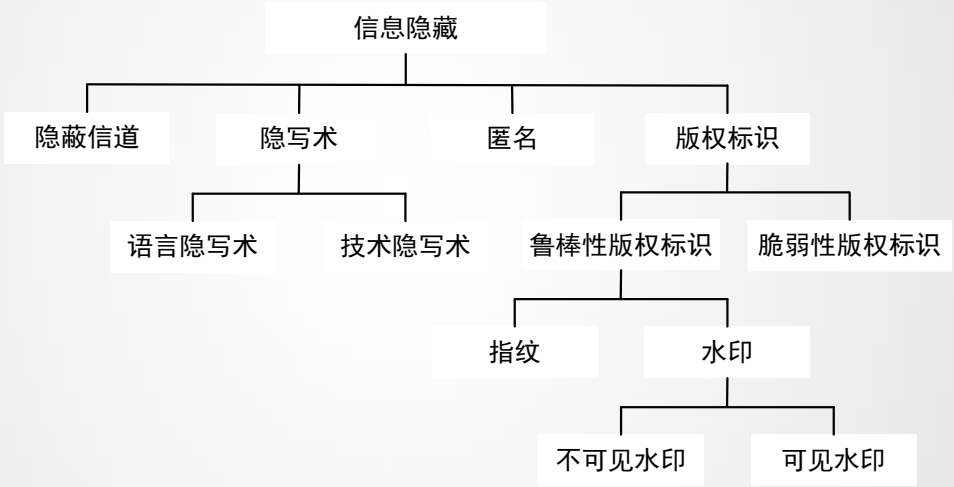
信息隐藏 VS. 密码学

- 两者之间有些差别
  - 信息传输方式不同
  - 信息保护的形式和时间不同
- 密码学和信息隐藏是信息安全领域两大重要的分支，两者并不矛盾。
- 在有些情况下，信息隐藏技术会用到加密技术，通过先加密机密信息，然后把类似乱码的机密信息用嵌入算法隐藏到公开媒介中，可实现具有更好的安全性。

信息隐藏技术模型


密钥分发中心
载体C_1
嵌入密钥K_1
提取密钥K_2
载体C_1
信息
M_1
预处理
M_2
嵌入算法
S
传输信道
S
提取算法
M_2
后处理
M_1
消息
攻击者
载体C_2

典型IH的分类


信息隐藏
隐蔽信道
隐写术
匿名
版权标识
语言隐写术
技术隐写术
鲁棒性版权标识
脆弱性版权标识
指纹
水印
不可见水印
可见水印

其它IH的分类

- 根据嵌入和提取所用参数的不同
  - 提取时不会用到原始载体图像*C*称为盲检测；反之称为非盲检测。
  - *C*1=*C*2：无损(可逆)IH模型；反之有损IH模型。
  - K1=K2：对称算法；非对称算法。
- 根据载体类型分类
  - 文本为载体技术
  - 图像为载体技术
  - 音频为载体技术
  - 视频为载体技术
- 根据嵌入域分类
  - 时空域：直接修改载体信息的冗余部分
  - 变换域：变换到另外一个空间（频域）

IH的特性

- 鲁棒性(Robustness)：载体不因某种攻击或改动而导致隐藏信息丢失的能力。
- 不可检测性(Undetectability)：要求嵌入隐秘信息的载体与原始载体之间具有一致性。
- 嵌入容量(Capacity)：在单位时间内或在一个载体内最多嵌入信息的比特数。尽可能提高该值。
- 透明性(Invisibility)：经过一系列隐藏处理，目标数据在质量上没有明显的降低，但隐藏的数据却无法人为的看见或听见。
- 安全性(Security)：嵌入算法具有较强的抗攻击能力，即它能够承受一定程度的攻击，但隐秘信息不会被破坏。
- 自恢复性(Self-repairability)：在嵌入隐秘信息的载体遭受破坏的情况下，能够从留下的片段数据中恢复出隐秘信息，且恢复过程中不需要原始载体的能力。
- 对称性(Symmetry)：嵌入过程和提取过程具有对称性，以减少存取难度。

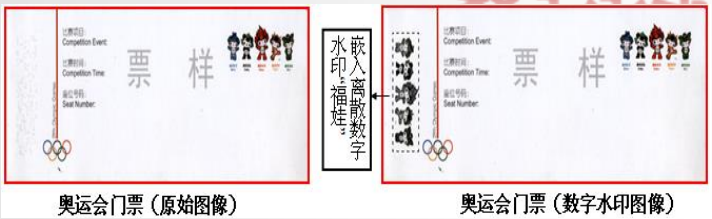
IH的典型算法

- 空间域(Spatial Domain)算法
  - 最低有效位LSB(Least Significant Bit)算法
  - 基于亮度统计特性的Patchwork算法
  - 调色板算法
  - …
- 变换域(Transform Domain)算法，又叫频域算法
  - 基于离散余弦变换(DCT)算法。
  - 基于小波变换(DWT)算法。
  - 基于离散傅立叶变换(DFT)算法。
  - …

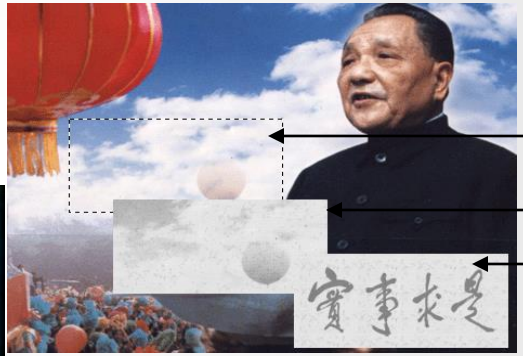
数字水印与版权保护

- 在信息隐藏技术中，隐写术和数字水印是两个主要的分支
  - 隐写术主要实现隐秘通信；
  - 数字水印(Digital Watermarking)主要用来实现版权保护、真伪鉴别、认证和完整性检测等
- 数字水印技术(Digital Watermarking)
  - 把标识版权的数字信息嵌入到被保护的数字产品中
  - 通过相应的技术手段使这段特定的数字信息不被人感知
  - 只有通过具有相应权限的拥有者通过专用的检测器或阅读器才能判断其是否存在。

数字水印与版权保护



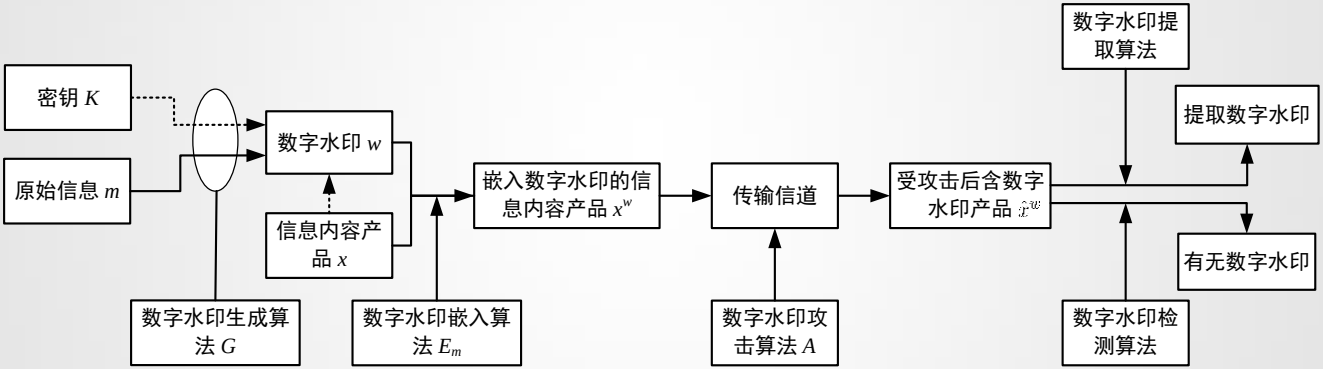
奥运会门票（原始图像）　　奥运会门票（数字水印图像）



数字水印基本特征

- 鲁棒性：含有数字水印的信息内容产品经过几何变换、压缩、加噪、滤波等攻击后，水印信息仍然可以正确的检测并提取出来。
- 不可感知性：主要是针对不可见水印而言，指从人类视觉上和采用统计方法也无法检测或提取数字水印信息。
- 安全性：即使攻击者知道数字水印算法的情况下，也无法实现未经授权的数字水印嵌入、检测/提取和未经授权的数字水印删除等操作。
- 可证明性：在含有数字水印的信息内容产品在遭受到盗版、侵权或泄露等行为时候，数字水印技术可以为用户提供安全、可靠且毫无争议的版权证明。
- 嵌入容量：一般而言，对于数字水印系统而言，其嵌入容量要求相对较小，而隐写术则通常要求较大的嵌入容量。

数字水印系统框架

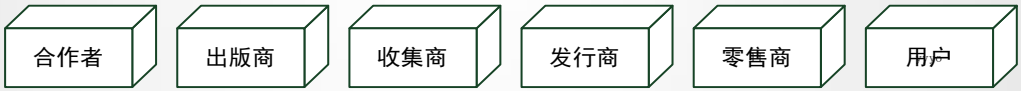

密钥 K
原始信息 m
数字水印 w
信息内容产品 x
数字水印生成算法 G
数字水印嵌入算法 E_m
嵌入数字水印的信息内容产品 x^w
传输信道
数字水印攻击算法 A
受攻击后含数字水印产品 $\hat{x}^w$
数字水印提取算法
提取数字水印
有无数字水印
数字水印检测算法

数字水印分类

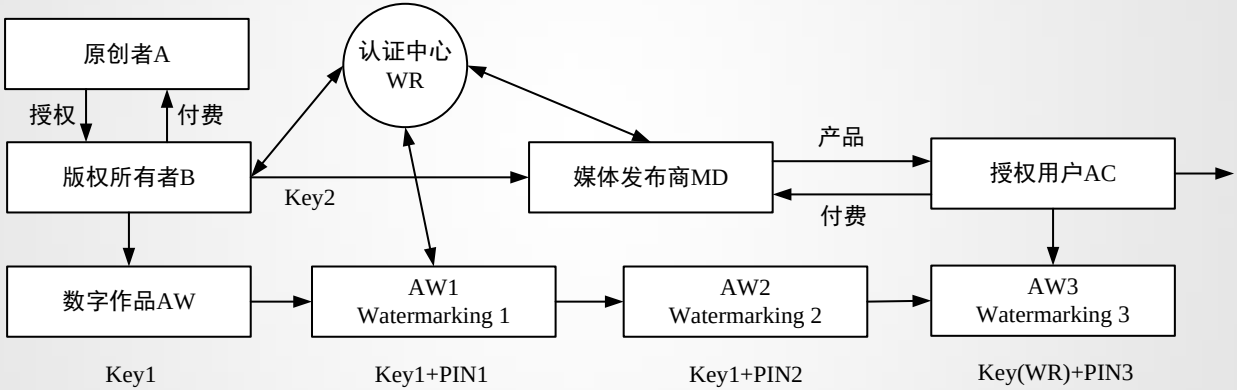
- 按数字水印所附载信息内容类型分类
- 按数字水印的外观分类
- 按数字水印的内容分类
- 按数字水印的特性分类
- 按数字水印的检测/提取过程分类
- 按数字水印隐藏的位置分类
- 按数字水印算法的可逆性分类
- 按数字水印算法的用途分类（版权保护水印，票据防伪水印等）

数字水印在数字版权保护中的应用

- 数字版权保护技术(Digital Rights Management, DRM)
  - 对各类数字内容知识的知识产权进行保护的一系列软硬件技术，用以保证数字内容在整个生命周期内的合法使用，平衡数字内容价值链中各个角色的利益和需求，促进整个数字化市场的发展和信息的合法传播。



基于数字水印的数字版权保护系统



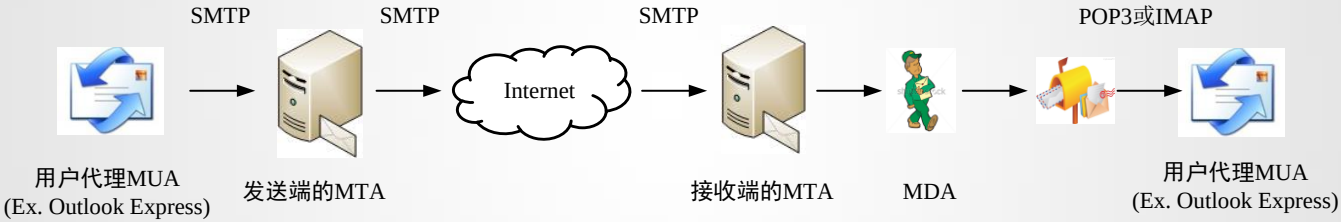
信息内容安全应用

- 垃圾电子邮件过滤系统
- 网络舆情监控与管理系统

垃圾电子邮件的概念

- 《中国互联网协会反垃圾邮件规范》
  - 收件人事先没有提出要求或者同意接收的广告、电子刊物、各种形式的宣传品等宣传性的电子邮件。
  - 收件人无法拒绝的电子邮件。
  - 隐藏收件人身份、地址、标题等信息的电子邮件。
  - 含有虚假的信息源、发件人、路由等信息的电子邮件。
  - 含有病毒、恶意代码、色情、反动等不良信息或有害信息的邮件。
- 基本特征
  - 不请自来
  - 带有商业或政治目的
  - 虚假，误导性

电子邮件系统传输过程


SMTP
SMTP
SMTP
POP3或IMAP
Internet
用户代理MUA
(Ex. Outlook Express)
发送端的MTA
接收端的MTA
MDA
用户代理MUA
(Ex. Outlook Express)

垃圾邮件的特征分析

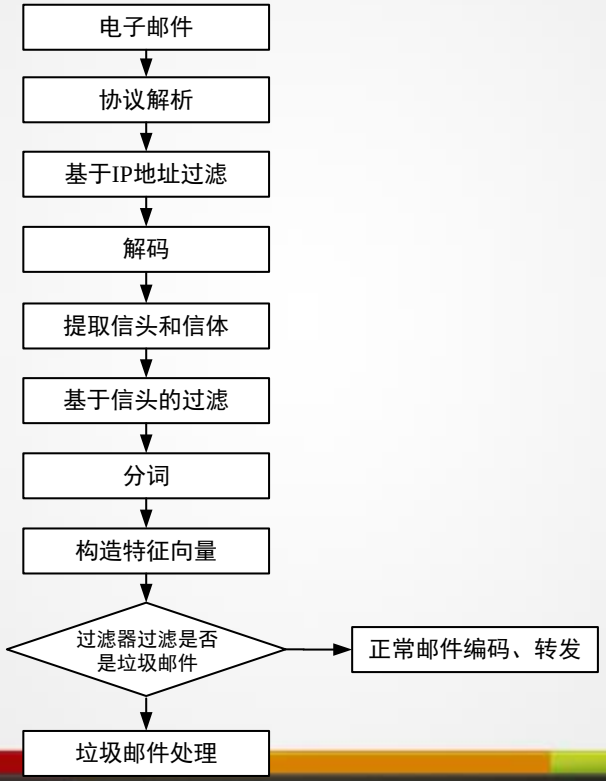
- 垃圾邮件的检测特征来源主要包括：通信特征、信头特征、信体特征

通信特征	IP地址是否可信	1
	IP链接数量、频率是否异常	1
信体特征	信体的大小问题，过大(包含内嵌资源或是大邮件轰炸)或批量空信	1
	附件的大小问题，附件过大	2
	附件的类型问题，为声音，图片，可执行文件，或包含恶意宏	1
	信体，附件包含关键词	2
	信体，附件语义分析包含垃圾信息	3

垃圾邮件的特征分析

信头特征	X-mailer没有或是特殊字段	2
	Mail From字段不相同或反向解析与真实的IP不符或包含关键词	2
	Received：时间有误，传送时间长，其中标识的IP地址有误，有3个以上Received或包含关键词	1
	Reply-to：与From字段不相同或包含关键词	1
	Message-id伪造，whois查询的结果该域名不存在	1
	Data：时间在当前时间之前	1
	Subject：包含关键词	1
	Cc：抄送人字段包含关键词	2

垃圾邮件过滤系统流程


电子邮件
协议解析
基于IP地址过滤
解码
提取信头和信体
基于信头的过滤
分词
构造特征向量
过滤器过滤是否是垃圾邮件
正常邮件编码、转发
垃圾邮件处理

典型的垃圾邮件过滤技术

- 基于黑白名单的过滤技术
- 基于关键字的过滤技术
- 基于统计的内容过滤技术
- 基于规则的内容过滤技术
- 基于邮件行为识别的过滤技术
- 图片垃圾邮件的过滤技术
- ...

网络舆情监控与管理系统

- 网络舆情具体以下方面的特点
  - 自由性
  - 交互性
  - 多元性
  - 偏差性
  - 突发性

网络舆情监控系统架构


用户
部门领导
界面显示
多方式上报
舆情简报
敏感信息监测
舆情预警
舆情检索
舆情跟踪
趋势预测
热点发现
网络舆情服务
舆情知识库
舆情信息挖掘
热点话题识别
倾向性分析
…
趋势分析
舆情信息预处理
自动排重
网页去噪
自动分词
…
语义分析
信息检索库
网络舆情分析处理
网络爬虫技术
通用网络爬虫
垂直型爬虫
批量型爬虫
论坛
新闻
博客
微博
贴吧
其它
网络舆情信息采集

网络舆情监控系统架构

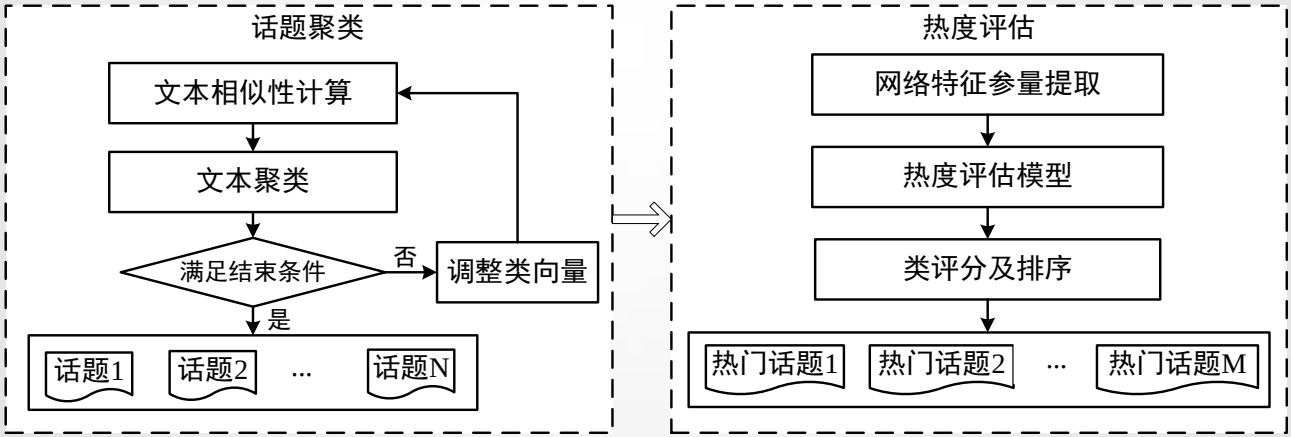
- 网络舆情信息采集
  - 采集各种论坛、新闻留言板、博客、微博、贴吧等信息源的各类信息，主要以文本为主，同时也包括图像、音频和视频等多媒体信息；
  - 能够实现满足用户需求的定向网络舆情信息地抓取；
  - 支持具有多线程、分布式采集功能的高速采集技术；
  - 支持具有身份验证的网络的采集，需要提供合法的用户账号；
  - 内置自动转码功能，可以将Big5或Unicode编码统一转换为GBK进行后续处理。

网络舆情监控系统架构

- 网络舆情分析处理
  - 自动排重：DSC(Digital syntactic clustering)算法、改进的DSC-SS算法(DSC-supershingle)、I-Match算法、基于关键词匹配的向量空间模型检测算法等
  - 网页去噪：主要用来识别并排除与网页主题无关的噪音信息，如广告信息、版权信息等，从而实现网页净化。
  - 自动分词
  - 语义分析：运用各种机器学习方法、挖掘与学习文本、图像等深层次概念

网络舆情监控系统架构

- 舆情信息挖掘模块
  - 话题识别与跟踪



网络舆情监控系统架构

- 舆情信息挖掘模块

➢ 倾向性分析：

- 主观文本情感分析
  - ✓ 词语情感倾向性分析
  - ✓ 句子情感倾向性分析
  - ✓ 篇章情感倾向性分析
  - ✓ 海量数据倾向性预测
- 文本主观性分析

网络舆情监控系统架构

- 网络舆情服务
  - 舆情跟踪
  - 趋势预测
  - 热点发现
  - 敏感信息监测
  - 舆情预警
  - 舆情检索
  - 舆情信息显示
  - …

本章小结

- 掌握信息内容安全相关概念及与信息安全关系
- 掌握常用的信息内容获取技术
- 掌握文本信息内容识别与分析技术
- 熟悉多媒体信息内容识别与分析技术
- 熟悉信息内容控制与管理技术
- 了解信息内容安全的应用