

(2020春季 课程编号: 011184)



信息安全导论

第13章 内容安全技术

中国科学技术大学 曾凡平

billzeng@ustc.edu.cn



课程回顾：第12章 恶意代码检测与防范技术

12.1 恶意代码概述

- 定义、发展趋势、种类
- 攻击流程、攻击技术和生存技术

12.2 常见的恶意代码

- 计算机病毒
- 木马
- 蠕虫

12.3 恶意代码检测与分析技术

- 静态分析方法
- 动态分析方法
- 分类方法



第13章 内容安全技术

13.1 内容安全的概念

13.2 文本过滤

- 13.2.1 不良文本过滤主要方法
- 13.2.2 中文分词

13.3 话题发现和跟踪

13.4 内容安全分级监管

13.5 多媒体内容安全技术简介

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**。

原始图像和压缩图像：内容相同

- 例如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。

内容安全

- 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”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。

对内容安全的4个需求

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**仅仅依靠非技术手段是不够的**，数字内容制作企业、内容制作者及管理部门也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过。

被动与主动的内容安全技术



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

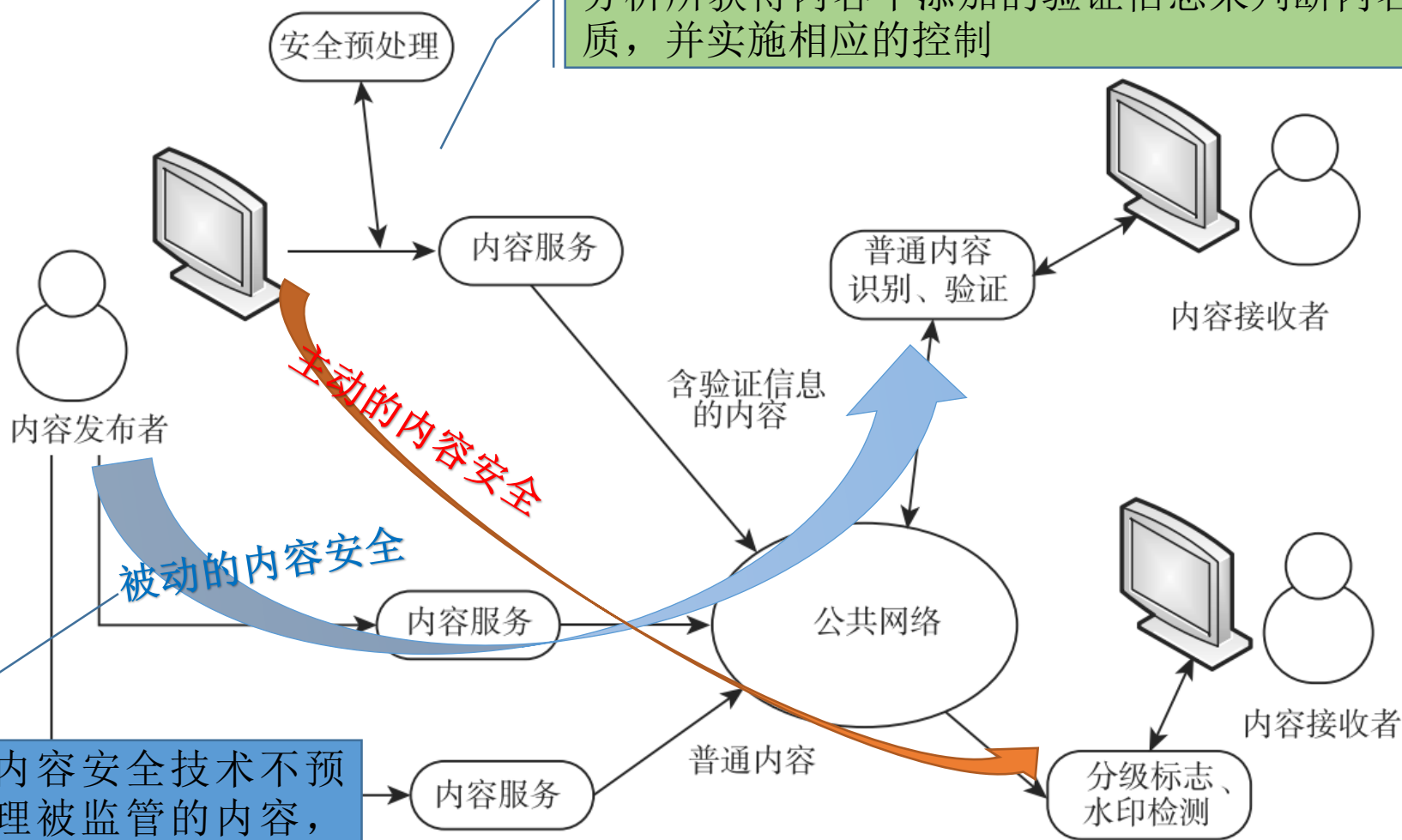


图13-1 被动与主动内容安全技术实施环境

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。

广义的内容和狭义的内容安全技术

- 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、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。

13.2 文本过滤

13.2.1 不良文本过滤主要方法

1. 基于关键字的过滤方法

- 基于关键字的过滤方法是不良文本过滤**早期常用的方法**。
- 首先由专业人员编制一个**不良文本关键字词库**，关键字词库中出现的字词都是经常出现在不良文本中的敏感词汇，能够很大程度地代表不良文本。
- 当有文本到来之后，对文本全文进行检索，通过比较简单的布尔逻辑运算进行匹配，当匹配超过一定**阈值**之后，系统就认为这篇文本是不良文本，给予过滤。

使用关键字匹配技术有以下三个难题

(1) 很难建立完整的不良文本关键字词库

- 网络信息如此之多，表达方式千差万别，文章主题繁冗复杂，任何人都不可能建立一个包括所有不良文本的关键字词库。不良文本关键字词库的不健全，就会导致不良文本过滤精确度的降低，又因为**关键字词有可能在正反两类文本中都频繁出现**，因此在采用不良文本关键字词库进行过滤时，经常会把关于某一个敏感主题的所有相关文本全部过滤掉。
- 目前，所有基于关键字的过滤都只是一个多层不良文本过滤系统的第一层过滤，用来区别对于一个主题的相关文本和不相关文本。然后对过滤得到的这两类文本的文本内容再进一步进行基于内容的动态过滤，最终把不良信息过滤掉。也可以通过不断的关键字过滤判断和人为的过滤方法相结合来进行不良文本过滤。



(2)不良文本关键词词库的滞后性

- 因为关键词词库的不完整性，需要不断地补充新的敏感词汇到不良文本关键词词库，这样带来的问题就是**敏感字词的滞后性**。系统总是在某一类的不良文本出现很多，但是没有被系统自动过滤之后，才会去抓取新的敏感词汇。过滤的滞后性也是一个必须考虑的问题。

(3)不良文本关键词词变形的难识别性

- 针对不良文本关键字过滤技术，很多不法分子采用拆分关键词的方法来逃避：使用特殊符号代替敏感字词；使用特殊符号间隔敏感字词；使用拼音替代敏感字词；故意使用错字或偏旁部首来代替敏感字词。这样也给采用关键词词库进行过滤带来了麻烦，而且由于没有能够准确的联系**语境**，只是单个的通过关键词词进行匹配过滤，导致了较高的误判率。



2. 基于分级标签过滤方法

- 分级标签过滤方法通过对不同的网页根据内容赋予不同的级别，以实现过滤。
- 根据网页内容的不同，分为普通级、一般限制级、严格限制级。青少年只能看到普通级别的网页，而成年人可以看到一般限制级别的网页，而包含反动等信息的严格限制级网页，则是要严格过滤掉的。
- 网络分级的顺利实施存在着以下几个问题。



- (1)网络分级目前还是一个**自愿采用的分级系统**，各个网站的管理团队如果采用了网络分级标签，那么他们就要为他们标记的内容承担责任。
- 而且有些网站，为了提高点击率，以获得经济利益，拒绝网络分级标签的使用，即使勉强采用，也会打擦边球，甚至故意标记错误。所以，网站运营者必须有很高的社会责任感，这个方法才可能有效。
- 另外，只是杜绝服务端并不能从根本上解决问题。在浏览器端，更大的问题在于“网络实名制”的实施，只有该规则的实施，才能避免不同浏览级别之间的用户混乱。比如，一个青少年注册账号谎称他是一个成年人，这样他就能看到一般限制级别的网页内容。



- (2)如果采用严格的技术来实现的话，**技术上存在着很大的问题。**
- 大型的网站特别是比较大的**BBS**，很多用户都有发帖和上传的权限，管理员也没有那么多的时间和精力去逐一地判断用户上传的是什么内容的东西。这种难度可以预见。
- (3)分级应**根据不同年龄有不同的约束内容。**
- 需要为不同的年龄段指定不同的浏览内容，这就需要有一个明确的标准，而且该标准就一定要让全社会的所有人都认可，不仅仅只是家长认可，同时也要吸收未成年人的意见。这样才能够做到标准的长久有效，为以后的网络分级打好基础。

3. 基于地址库过滤方法

- 基于地址库的过滤方法可以分为以下三个类别：
IP地址过滤、URL过滤以及IP和URL相结合的过滤方法。
- **IP过滤是指通过封锁指定网站的IP地址**，实现对包含有大量不良文本的网站的过滤目的。该方法简单易行，但同时也存在着很大的不足：第一是颗粒度过粗的问题，一个大型的网站，由于内容繁多，每日都有大量要处理的信息，必然会导致管理困难。特别是对于大型的BBS和论坛来说，用户具有相当大的权限，拥有更大的自主言论的环境，网站中难免会有反面的信息存在，用IP过滤的话会直接屏蔽掉整个网站。



IP过滤的不足之处

- 而且还有可能一个IP地址对应着好几个网站，通过IP过滤方法进行过滤，则会一刀切，合法的、非法的网站都直接给过滤掉；
- 第二是由于代理的普遍使用，即使已经对某些网站进行了屏蔽，但是现在好多网站都提供了很多的国外可用代理。作为普通用户来说，只要很简单的设置，通过这些代理，可以很容易地访问到已经被屏蔽的网站。

URL过滤

- 基于IP过滤存在的不足，提出了URL过滤方法。URL过滤方法直接定位不良文本在互联网上的具体位置，直接对该网页进行屏蔽，可以实现精确定位，准确过滤。
- 但是URL过滤也有不足，那就是在海量网页中会存在很多的不良网页，这时网页地址一个一个地输入，而且当我们每次访问一个网页的时候，都需要和所有的不良网页的URL进行比较，会造成很大的浪费。
- 除此之外，如果一个IP对应的所有网页都是不良文本的话，这个时候，反而是采用IP过滤方法比较好，采用URL过滤方法的话，不仅输入时麻烦琐碎，而且当访问新的网页地址需要和所有不良文本的URL地址进行比较的时候，会加大运算量。

IP过滤和URL过滤的结合

- 通过分析IP过滤和URL过滤的优点和不足，发现这两种过滤方法具有很大的互补性：IP过滤颗粒度较大，URL过滤颗粒度较小。两种方法结合使用后，对于一个网站下大部分的网页都是不良信息的情况，则采用IP过滤；如果是一个网站下只有极少一部分是不良文本这种情况，则采用URL过滤。不过，需要看到的是，即使采用IP和URL相结合的过滤方法，除了上面提到的过滤方法中使用代理访问的问题需要解决之外，还有很大的缺陷：该方法只能针对已经知道的不良文本的网络地址进行过滤，对于新增加的不良网站和网页，则没有任何的作用，这些新产生的不良信息会很容易地通过。而在这个互联网数据大爆炸的时代，网络上每天都会产生大量的新的不良信息，这就给不良网站和网页地址的收集工作带来了很大的麻烦，因此维护IP和URL黑名单就会很复杂、很困难。

4. 基于内容的动态过滤法

- 对于同一个话题，正反两方言论在关键词词使用上存在很大的重复性，因此在使用关键字过滤方法进行不良文本过滤过程中，会有**很高的误判率**。这时，就需要进一步根据文本的具体内容来判断指定文本的具体类别属性。
- 基于内容的过滤方法主要有：**K近邻法**，**贝叶斯方法**，**神经网络算法**，**潜在语义索引法**等。
- 该类方法首先将文本的表示形式具体化，然后再通过指定的过滤算法，对于指定的文本划分类别属性，最后将对不良文本进行过滤。



- 和不良文本关键词词库过滤方法相比，基于内容的动态过滤法能够比较深入地分析文本内容的具体倾向性，然后再进行类别属性划分，从而实现对动态文本流的过滤，并有比较高的精度。
- 但是，基于内容的过滤方法并不是最完美的过滤方案，目前文本的表示方法主要采用向量空间模型方法。该方法将文本的内容表示成一个词的集合，用单个的词来表示一个文本，词与词之间是孤立的，没有考虑词之间的上下文联系。
- 在计算两篇文本的相似度时，只是通过遍历计算两篇文本之间存在的特征选择算法选择出来的词的相似度，最后累加得到两篇文本总的相似度，却忽略了文本的语义属性。



- 比如，“中国队战胜了美国队”和“美国队战胜了中国队”，意思完全相反，但是如果使用不考虑文本语义属性的过滤方法，这两句话就会有100%的相似度，算法会认定这两句话表达的是一个意思，这和实际情况完全不符合。
- 后来，也有学者对该类方法提出了改进，提出了基于格语法的不良文本过滤等多种基于语义的过滤方法。这类方法主要考虑了文本中词之间的前后联系，词在文本中所处的位置以及在一个语句中词所处的位置等信息。该方法加强了文本的语义表达，理论上可以达到很高的过滤精度和较低的误判率，可同时该类算法过于复杂，计算量过于庞大，执行速度慢，目前来说根本不可能满足实时过滤的要求。如何提高该类方法的执行速度是目前研究不良文本过滤的主要方向之一。



- 即使基于内容的不良文本过滤在精度上已经有了很大的提高，但是该方法由于计算量比较大，造成了运算速度慢，实时性过滤比较差等不可回避的问题，其中，基于语义分析的不良文本过滤方法计算量更大。
- 而且，由于网络上文本表达方式的多样性以及不法分子为了躲避惩罚对词进行了“加工”，也给文本内容的预处理工作带来了更高的技术要求，从而进一步限制了过滤精度的提高。
- 此外，计算机自然语言处理技术本身的复杂性，也将会使得实现实时处理且达到较高的过滤正确率和较低的误判率还有很长的路要走。

13.2.2 中文分词

- 基于中文的文本分类和基于英文的文本分类，最大的差别就来自于在基于中文的文本分类过程中，首先需要先进行分词。
- 基于中文的文本分类技术起步比较晚，大多都是照搬国外的文本分类技术，然后在其基础上再进行改进，使其适应中文文本分类技术。但是，中英文在写法规则上有很大的不同，英文在书写过程中，单词之间都是由空格来隔开的，词与词之间是自动间隔的，不同的只是同一个词的词性不同，根本不存在错误划分词的可能性，可以百分百准确的分词。但是中文的书写方法却是写完一个完整的句子或者表达了一个完整的意思之后，才会用标点符号将其隔开。

中文分词

- 从根本上说，中文书写是以句子为间隔，而英文是以一个一个的词为间隔的。而在英文文本分类技术中，大多数词采用了基于向量空间模型的文本表示方法，而向量空间模型则是将文本中的每一个词作为模型中的一个维，因此英文文本分类技术都是以文本中自动间隔开的词为基本单位进行文本分类。中文的分类方法又大都是照搬西方的技术，这就要求在中文文本分类过程中，首先需要将文本中连续的字符串按照一定的算法将其划分为独立的、有实际意义的词，即进行分词操作。
- 国内在分词研究方面取得了引入瞩目的成绩：
 - <https://baike.baidu.com/item/中文分词>

13.3 话题发现和跟踪

- 话题发现和跟踪的关键技术是“聚类”，相关的4个概念如下：
- 定义13.1 **报道**：包含两个或多个独立陈述某个事件的字句的新闻片段。
- 定义13.2 **事件**：由某些原因、条件引起，发生在特定时间、地点，涉及某些对象（人或物），并可能伴随某些必要结果的一个特例。
- 定义13.3 **活动**：一个互相关联的事件集，这些事件发生在特定的时间、地点，有共同的焦点或目的。
- 定义13.4 **话题**：一个核心事件或活动以及与之直接相关的事件或活动。可以简单地将话题理解为若干关于同一事件的报道。

话题自动发现的流程

- 话题发现的目的就是把互联网上大量关于同一事件的报道聚合为话题，换句话说，话题内包含关于同一事件的大量报道。
- 话题发现是为网络舆情监控服务的，它的任务是把大量讨论同一事件同一话题的新闻报道聚合在同一个簇里面使之有序呈现，减少呈现给用户的结果列表中的重复和冗余。这样做的另一个好处是，可以以话题为粒度考察其受关注的程度即热度，从而更好地评估话题的潜在影响力，为舆情的监管和预警提供依据。

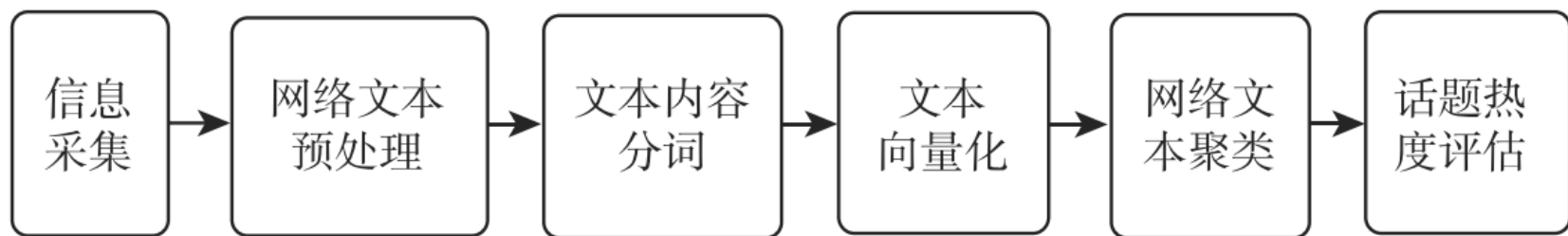
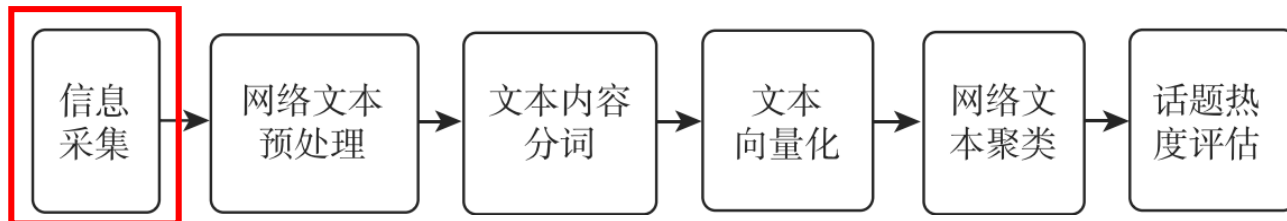
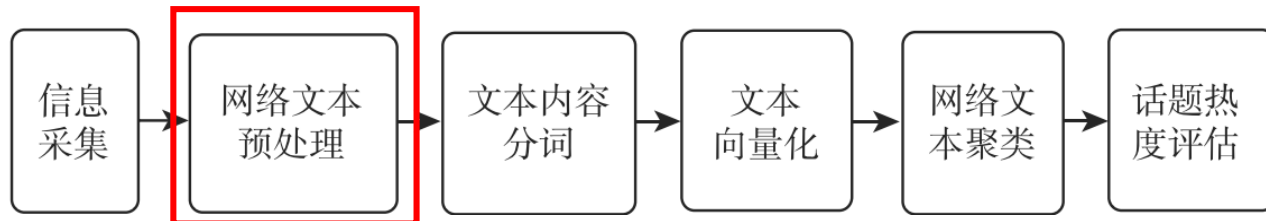


图13-2 话题自动发现的流程



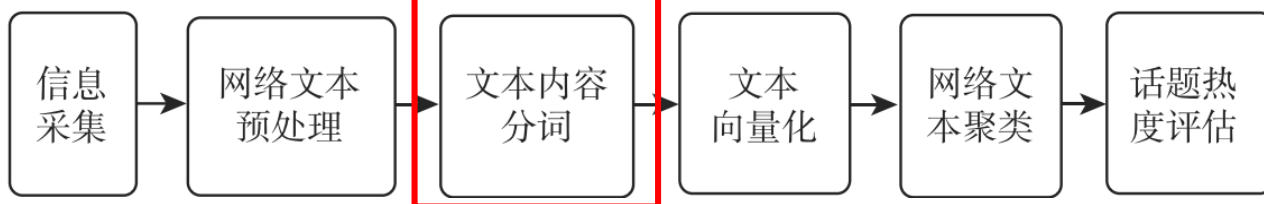
- **信息采集阶段利用网络爬虫工具，从指定的Internet网站把Web网页等互联网信息资源抓取到计算机本地进行存储。**网络爬虫为人们批量获取互联网信息提供了很大的方便，人们可以通过简单设置目标网站、爬行间隔和存储位置等选项，就可以源源不断地获取目标网站上的不断更新。而爬虫工具内部都设置好了爬行策略、线程分配和爬行深度，除非必要，一般很少需要用户更改这些比较复杂的设置项。
- 常见的爬虫工具有spider、crawler等，会定期地按照预设地址查看相应网页，网页发生变化时就获取该网页，根据网页中的链接不断访问。网络爬虫访问页面的过程是一个遍历的过程，可按照广度优先或深度优先的策略访问完所有的链接。互联网信息的采集可以通过门户网站、论坛、聚合新闻等方式，不同方式中信息的范围和更新速度不一样。



- **网络文本处理也称预处理**，即把互联网网页源码信息进行处理，包含剔除无关字符清洗源码、提取正文和必要的附带信息（如来源、作者、发表时间等）。
- 话题发现的操作对象是网页的标题和正文以及其他一些相关的信息，网页上的导航信息、版权信息和广告信息不是应该关注的对象，在处理中要过滤掉。同时，为了得出话题后热度计算的方便，会在处理中提取出与热度计算有关的一些网页元数据，如网页的点击量。
- 基于HTML的标记语言特性，可以利用基于标签的层次结构分析来剔除无关信息，利用正则表达式抽取目标内容。



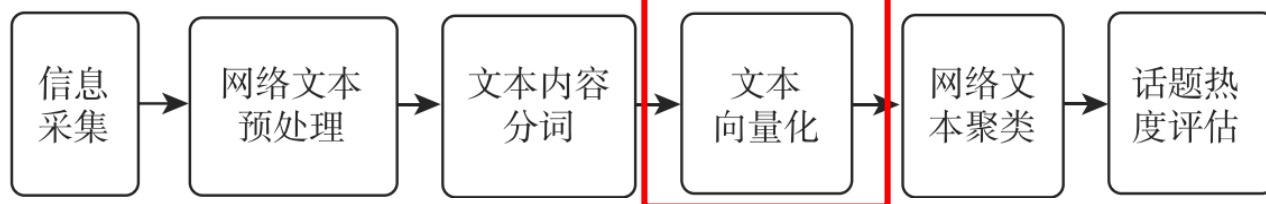
- HTML语言中标签标记了各部分内容的性质和地位，本书据此来构建结构树，通过分析总结出无关信息在结构树中的路径，采取剪枝的方法实现清洗网页源码。正则表达式主要用于基于文本的搜索和编辑，可以实现数据验证、文本替换，还可以根据模式匹配从字符串中提取子字符串，可以利用正则表达式的这种提取功能从网页源码中抽取以条件定位的有用信息内容。
- 预处理得出的网页元数据将被存储到数据库表中，抽取的正文内容被以文档的形式存储到本地硬盘，并与数据库表中的纪录保持一一映射。网页的元数据包含网页的URL、来源网站、存储位置、网页标题、时间、点击量、跟评数等。
- 网页URL作为网页的主码标识，和存储位置一起用于网页文档与网页数据库表元纪录的关联，和来源网站一起用于检测该报道是否在来源网站的首页上，网页标题用于直观代表网页，时间、点击量和跟评数用于计算报道话题的关注度。



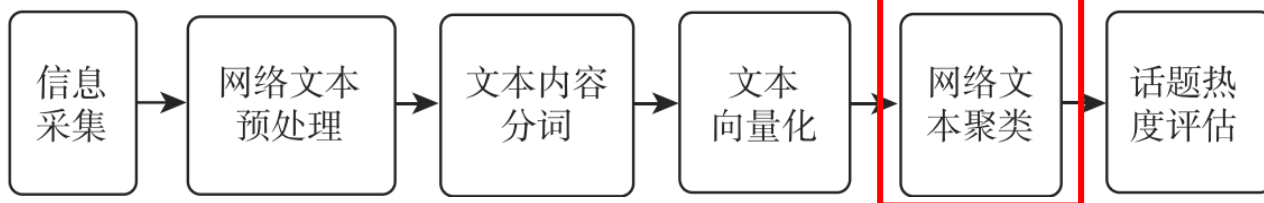
- **文本内容分词**是在汉语文本处理，且选择词语作为文档特征表达的特定情况下必要的一个步骤。
- 英语因为单词之间空格的存在，使得程序可以很容易地识别单词与单词之间的界限。
- 而汉语文本处理时，如果选择以字或句子作为特征则不需对文本内容进行分词处理，比如文本摘要的研究，一般以句子权重来组合出摘要。
- 在话题发现研究中，需要以词语来构建文本的多维向量空间，进而将文本转化为一个空间向量以便进行接下来的相似评估，所以需要文本进行分词。



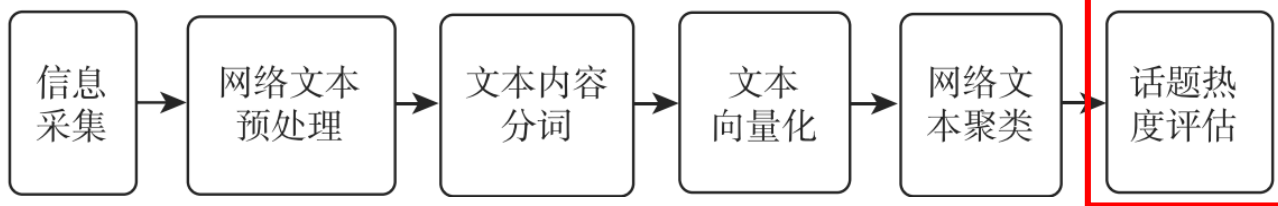
- 目前国内分词方法很多，有词典匹配法、设立标志法、人工智能法等，但绝大多数分词方法还是建立在词典的基础上，辅之以词法、语法和语义规则进行字符串匹配。
- 文本分词在词典的指导下，将文本内容分解为一个个汉语词语并添加空格隔开，以便于计算机识别进行后面的相应操作。
- 能够标识报道所代表事件的主要是具有实际意义的名词、动词，也就是说，对话题发现能起到关键作用的往往是这些词，而副词、连词、介词和叹词等词的作用微乎其微，并不能显著地区分不同的事件。对涉及同一事件报道的空间向量构建的研究也证实了，各报道非零的维度交集主要是这些实词。
- 在分词的过程中，通过词性标注对词语加以选择，忽略掉副词、连词、介词和叹词，以减少计算和存储，同时也可以提高话题发现的精度。



- **文本向量化**是汇总分词后文本中的词语，将这些词语作为空间向量的维度构建文本表示的多维向量，然后将各词的文档词频统计值和逆文档词频统计值运用TFIDF(token frequency and inverse document frequency)公式转换为一个权重值，用以表示文本在这个词语代表的维度上的值，进而将文本表示为一组关键词及其词频为权重的空间向量。
- 聚类的过程就是两个向量之间进行相似程度计算的过程，正是文本向量化这一步骤使得文本型的网页转换成了数值型的向量，使得聚类计算成为可能。词语的文档频率指的是词语在该文档中的出现次数，用于衡量该词语对文本的重要性。逆文档频率指的是包含这个词语的文档占全部文档的比重的一个倒数，用于衡量该词语在所有文档间的重要性。



- **网络文本聚类**采取一定的组织策略调度文本向量参与相似度计算，并建立话题的向量表示方法。
- 聚类的任务是，确定两个文本是否可以被认定描述了同一话题，或者文本是否可以被认定属于和它进行比较的话题。
- 向量之间的相似程度计算是通过夹角余弦公式来实现的，夹角余弦公式基于欧几里得距离计算空间向量之间的距离。当这个距离小于某个预设阈值时，认为向量是相似的，向量所代表的文档也被认定以较大概率属于同一话题，否则不属于同一话题。
- 不同的聚类算法采取不同的文本向量组织调度策略，以不同的方式进行相似或相异的比较，将大量无序的文本划归为若干类簇，使得不同类簇描述不同事件，得出不同的话题。



- **话题热度评估**综合考虑话题中所有报道的点击数、回帖数、报道频率和时间频率等参数，来评估该话题受到关注的程度，并结合刊载源的级别、传播力、影响力来衡量话题的重要程度，综合两者共同表征报道和话题的热度。
- 热度反映了话题代表的舆情所具有的紧要程度，可以帮助用户筛选最紧要的舆情事件优先进行处理，将有限的人力、精力、物力用到最关键和最需要的地方，从而体现舆情监控系统的智能性和易用性。



- 在话题发现的所有步骤里面，信息采集和网页文本处理为网络文本聚类提供了数据来源，这是基础；
- 文本内容分词和文本的向量化是必要的转化，以使后面的步骤可以顺利进行；
- 网络文本聚类是话题发现的核心，聚类算法的选择决定了话题发现的具体策略，算法的效率和精度也关系着话题发现的有效性，一个能够在互联网信息处理背景下发挥优秀聚类作用的聚类算法，将为网络舆情监控的话题发现功能提供良好的实用性保证；
- 准确地说，话题热度评估不属于话题发现的组成部分，但是从网络舆情监控的角度出发，它显然有利于话题发现的定量分析与把握。

13.4 内容安全分级监管

- **信息内容分级标准：**是整个信息内容主动监管系统的基础，包括内容分级的词汇表、分级标记和分级操作方法。信息内容分级标准化工作的主要内容应包括：内容分级标准制定和维护过程的规范与管理、通用信息内容分级标准及分级操作方法的制定、对不同应用领域内容分级标准的细化指南等方面。
- **分级信息的发布：**信息内容分级标准在信息发布环节的应用主要体现在信息的分级标记上。将信息分级处理与发布流程控制相结合能更好地保证分级的准确性。对不同类型的信息，采用不同级别的发布管理与控制，在分级的准确性和效率方面求得平衡。由于信息标记体现了发布方对信息内容分级处理的责任，因此对标记信息自身需要采取保护措施，防止标记信息被篡改。



内容安全分级监管

- **分级信息的使用：**信息内容分级标记的应用可以在信息服务提供和信息服务使用两个层面上实现。在信息服务提供层面上，主要是针对不同的信息受众群体的需要，对信源提供的信息进行分级过滤和组合。
- 从目前网络应用的发展趋势来看，信息服务将逐步取代单纯的信息提供成为网络信息资源的主流，而分级信息服务则是与分级信息发布过程相对的一个过程。
- 在信息服务使用层面上，主要是根据单个信息受众自身的需要，同时结合有关管理手段的需要对所接收的信息进行分级过滤处理。所接收的信息可能来自分级信息服务机构，或直接来自分级信源。一般而言，分级信息服务能在较大的范围内产生影响，但分级管理的粒度受限，二分级信息终端过滤能根据用户的需求实现精细而灵活的调整。

13.5 多媒体内容安全技术简介

- 由于数字媒体易于无损地拷贝、分发等特性，人们也可以借助数字技术和互联网，免费并且没有任何质量损失地批量复制和发行受知识产权保护的数字媒体产品和内容。未经授权的访问、复制、发行具有知识产权的数字产品使得数字媒体业遭受了巨大的损失。
- 就我国而言，大量的数字作品盗版和非授权使用严重地影响了我国影视、软件、游戏、音乐等行业的进步，阻碍了我国文化产业健康正常的发展，对我国的文化安全构成了严峻的挑战。
- 在开放的网络环境下，数字媒体产业迫切需要的有效的技术手段来保护知识产权和保障数字内容的创作者、出版商、发行商的商业利益。

多媒体内容安全有迫切的现实需求

- 缺乏对数字媒体的权益管理和内容保护将导致大量的负面现象，不但会滋生大量的对数字媒体产品的盗版行为，对数字媒体业造成巨大的冲击，而且也会对我们的经济、社会和文化造成多重的伤害：
- 它会延缓宽带以及数字媒体的发展脚步；削减对新设备、新应用的需求并导致IT市场的停滞；同样，从社会的层面来看，由于数字媒体的停步不前，社会的教育文化程度就会下降；商业系统的运营成本会增加而利润却会减少；个人的生活环境和生活品质自然也会降低，最终将阻碍我国社会和文化的繁荣和进步。
- 如何对数字媒体产品进行有效的权利管理和保护，维护数字多媒体的内容安全，保护数字媒体避免未经授权的访问、复制和发行，成为十分迫切的重大研究课题。

数字多媒体内容安全面临的问题

- **(1)如何鉴别一个数字媒体作品的创建者。**传统作品一般采用签名的方式，但对于数字媒体作品来说，一般的签名很容易擦除或伪造，对签名的确认也缺少一般作品所能依据的笔迹或图章等证据。
- **(2)如何确定数字媒体作品创建者的版权声明。**数字媒体作品的创建者有时会对自己的作品声明保留权利，或附加一些版权信息。对这些要求如何确认也是一个问题。而且版权声明往往携带相对大量的信息，很多情况下不能用与签名一致的方法对作品嵌入这些信息。



数字多媒体内容安全面临的问题

- **(3)如何公证一个数字作品的签名与版权声明。**一个数字作品上可以附加多个签名和版权声明，创建者或他人也可能否认。对签名真伪的鉴别以及对版权声明的确认不能仅仅由创建者来执行，而且必须通过第三方进行验证。这又涉及一系列技术、管理、法制的问题。
- **(4)如何控制用户访问数字媒体作品的权限。**用户购买了数字媒体作品获得使用权后，如果缺乏监督和控制用户访问及使用数字媒体作品权限的手段，就无法有效地阻止其非法复制、分发或使用该数字媒体作品。因此，在开放的网络环境下，数字媒体产业迫切需要有有效的技术手段来保护数字多媒体的内容安全和保障数字内容的创作者、出版商、发行商及消费者的利益。

数字权益管理(DRM)技术

- 在这种背景下，**数字权益管理(DRM)**应运而生，提供从数字内容的创作者，到发行者，到消费者的整个价值链的权益保护，并且结合了新的商业模式为数字媒体业增加了新的机会。
- 现在，**DRM**技术已经从第一代发展到了第二代。第一代**DRM**技术侧重于数字内容的加密、防止未授权的使用，即保证只把内容传递给付费用户，没有实现全面的数字权益管理。第二代**DRM**则扩展到对基于有形或无形资源的各种权益进行描述、标定、交易、保护、监督和跟踪以及对权益所有者进行管理，即第二代**DRM**管理所有相关的权益，而不是局限于数字内容的访问控制。



第13章 作业

- 作业

- 1. 对内容安全的需求主要体现在哪几个方面？
- 2. 不良文本过滤主要有哪些方法？各自的工作原理是什么？
- 3. 话题发现主要包括哪几个步骤？
- 4. 多媒体内容安全面临的主要问题有哪些？

- 实践（自己研究，不考核）

- 从百度学术网站查阅2019年以来的DRM文献：
 - 网站：<https://xueshu.baidu.com/>
 - 关键词：数字权益管理(DRM)