**电子科技大学2018-2019学年第 2 学期期 末 考试A卷**

考试科目：知识产权管理 考试形式：论文 考试日期： 2019 年 5 月 7日

成绩构成比例：平时 40 %， 期中 0 %， 实验 %， 期末 60 %

本试卷由 一 部分构成，共 1 页。考试时长： 分钟 注：学生课下自行查阅资料完成

|  |
| --- |
| 得 分 |
|  |

**论文题目：**

1. 新类型作品的著作权保护（例如：影视作品/娱乐节目/软件等）之我见

2．某一特定行业（例如：智能手机行业，生物制药行业等）的专利纠纷及其启示

3. 知识产权案例分析（例如：试论“微信”商标之争等）

4. 某一特定产业的知识产权竞争态势分析及中国的对策研究

5. 互联网+时代专利/著作权/商标/反不正当竞争保护的前沿问题研究

**论文撰写要求：**

1. **写作要求：**从以上四个论文方向中任选一题，具体题目或副标题可以自拟，运用《知识产权管理》课程相关知识进行分析。
2. **字数要求：**3000-4000字。

**3.写作周期：**2019年4月19日-5月7日。

**4.论文提交：**2019年5月7日晚9-10节，提交到沙河校区二教201。

**5.撰写格式和内容要求：**

（1）文档大标题采用小二号黑体；二级和三级标题采用小四号楷体；正文采用五号宋体，单倍行距。

（2）充分发挥同学们的创造能力，文字简练，逻辑清晰，有理有据。

（3）拒绝抄袭，避免雷同。一经发现，概作零分处理。

（4）提交时只需提交本试卷页面和撰写正稿，本试卷页面作为封面一并装订。

（5）文档封面和正文每页都需签名。

网络爬虫与其法律风险

1. 网络爬虫的定义

网络爬虫是一种按照一定的规则，自动抓取互联网信息的程序或脚本。用非专业的术语来描述这个过程，大概就是利用程序将人可以通过互联网浏览到的信息自动的按照一定规则下载并储存。生活中常见的例如百度、谷歌等搜索引擎便是利用网络爬虫程序爬取各个网站的内容并通过算法进行展示，而各种股票交易数据库等专业数据库也大多是通过爬取专业网站的公开内容建立的。

网络爬虫的运作过程通常可以分为如下三个阶段：1. 信息的采集，在这个阶段爬虫模拟用户进行用户登录等、浏览信息等操作，并将所设定的信息进行爬取采集。2. 数据的处理，通常来说爬取到的数据进行脱敏或提取等操作。3. 数据的存储，在这个阶段爬虫通常会将上一个阶段获得的数据存储到本地数据库中，以待进一步的使用。

正如在上文所介绍的那样，网络爬虫其实是一个在现实生活中已经被广泛利用的技术，而大部分网站也有被搜索引擎嗅探到并展示的需要，为了统一规范，Robots 协议成为了一个解决方案。根据中国互联网协会《互联网搜索引擎服务自律公约》第七条的定义，机器人协议（robots协议）是指互联网站所有者使用robots.txt文件，向网络机器人（Web robots）给出网站指令的协议。具体而言，robots协议是网站所有者通过位于置于网站根目录下的文本文件robots.txt，提示网络机器人哪些网页不应被抓取，哪些网页可以抓取。

1. 网络爬虫存在的法律风险及常用的规避方法

爬虫存在的法律风险主要可以归结为以下三个方面：

1. 不遵守 robots 协议的法律风险

《互联网搜索引擎服务自律公约》第七条规定，“遵循国际通行的行业惯例与商业规则，遵守机器人协议（robots协议）”。第八条则规定“互联网站所有者设置机器人协议应遵循公平、开放和促进信息自由流动的原则，限制搜索引擎抓取应有行业公认合理的正当理由，不利用机器人协议进行不正当竞争行为，积极营造鼓励创新、公平公正的良性竞争环境。”

具体的案例有“奇虎科技诉百度不正当竞争案”，法院经审理认为，Robots协议被认定为搜索引擎行业内公认的、应当被遵守的商业道德，被告奇虎公司在推出搜索引擎的伊始阶段没有遵守百度网站的Robots协议，其行为明显不当，应承担相应的不利后果，最终判决认定，被告北京奇虎公司的行为违反了《反不正当竞争法》，判决奇虎公司赔偿百度公司经济损失及合理支出共计人民币70万元，并驳回百度公司的其他诉讼请求。

通常来说，最好的规避此风险的方法是遵守网站的robot协议。

1. 对突破网站设置的技术防护措施带来的法律风险

正如双十一大量用户对淘宝的访问请求会对其服务器带来巨大的压力一般，爬虫因为程序设置的效果，过于频繁与快速的访问请求会对网站的正常维护带来压力，网站也通常会利用各种技术防范爬虫过于泛滥。

在《反不正当竞争法》第十二条第二款中有规定，经营者不得利用技术手段，通过影响用户选择或者其他方式，实施列妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行的行为。而如果所突破的是某些特定领域，《刑法》第二百八十五条规定，违反规定侵入国家事务、国防建设、尖端科学技术领域的计算机信息系统的，不论情节严重与否，构成非法侵入计算机信息系统罪。

程序员在编写网站爬虫时一般会对网站采用的各种技术进行了解，并规避可能存在的风险。

1. 爬虫爬取的信息可能带来的法律风险

爬虫的最终目的是为了爬取数据，而相比较传统领域，互联网领域的著作权、隐私权、专利权等显得尤其复杂，难以一概而论。就爬虫运作机制而已，其本质是模仿用户的行为进行爬取，对于用户可以阅读并保存的信息一般是不存在侵权问题的。而对于用户不能访问到的，或者是无法通过正常行为保存而通过技术手段保存的，就很有可能存在侵权的风险。

而另一些情况下，即使是公开的数据，若这些数据为网站的核心竞争力所在，而爬虫爬取数据完成后又利用这些数据进行商业运作，这种损害竞争优势的行为即构成不正当竞争，有名的“新浪微博诉‘脉脉’不正当竞争案”即此类代表。

涉及到用户个人信息的爬取，即使是公开的个人信息通常网站与用户也存在隐私协议，而爬虫是否经过用户同意认可也无从得知。如果是不公开的个人信息的爬取和利用机会一定触犯法律。

针对信息，一般来说如果需要利用爬取的信息进行进一步研究，通常会在爬取完毕后进行数据的脱敏操作。相类似的，近几年较热门的机器学习方向通常需要大量的数据集，而数据集的一个重要来源即网络爬虫，论文的发布一般会对使用的数据集进行隐私处理，以防止侵权行为的发生。

1. 个人案例分析

下面将会分析三个本人所正在开发的项目中爬虫的运作以及针对可能存在的法律风险进行的额外工作。

第一个项目核心为文本编辑器的智能推荐，即用户在使用我们提供的网页编辑器进行一些特殊文章（如合同、借条、文书等专业性较强的文本）的撰写时，输入句子的一部分，便可以通过后端的算法分析数据集并提供可能的建议。这一个项目采用人工智能机器学习的算法，需要一定数量的文本作为数据集进行分析。在制作爬虫爬取数据集这个阶段，我们对可能存在的法律风险进行了调研，并采取了一定的手段进行规避。

首先，对于网络文章的爬取，我们确认了爬取文章的网站不存在额外的协议禁止转载或再利用，规避了可能存在的版权风险。同时，我们确认了我们与爬取文本的网站直接不存在直接竞争，尽量不构成不正当竞争。在制作爬虫的阶段，我们严格遵守网站提供的 robots.txt 协议，同时因为项目时间相对宽裕，我们降低了并发数，即尽可能降低了爬虫对网站服务的影响。

此项目的另一个数据来源为用户在使用编辑器后的文章，我们计划利用用户的文本形成循环润化机器学习算法。于是我们在用户注册使用的协议中指出了我们将再利用他们的文章，并且提供选项可以主动禁用。对于爬取到的用户文章也开发了脱敏算法进行脱敏操作避免触犯隐私权。

另一个项目是一个新闻知识图谱的研究。主要是爬取新闻网站的新闻，在提供转发的同时对新闻内容进行拆分，并进行文本理解分析，最终以知识图谱的形式进行展示。经过调研我们发现，根据《[著作权法](https://www.baidu.com/s?wd=%E8%91%97%E4%BD%9C%E6%9D%83%E6%B3%95&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)》第二十二条规定，为报道[时事新闻](https://www.baidu.com/s?wd=%E6%97%B6%E4%BA%8B%E6%96%B0%E9%97%BB&tn=SE_PcZhidaonwhc_ngpagmjz&rsv_dl=gh_pc_zhidao)，在报纸、期刊、广播电台、电视台等媒体中不可避免地再现或者引用已经发表的作品; 可以不经著作权人许可，不向其支付报酬，但应当指明作者姓名、作品名称，并且不得侵犯著作权人依照本法享有的其他权利。对此我们在爬取新闻的时候特地进行了作者姓名、作品名称的爬取，对捕捉到“不允许转载”相关字眼的新闻不做保存及分析，以减少新闻爬虫可能带来的法律风险。至于针对爬虫本身的优化如上一个项目一般做了基本的处理，没有特殊的优化。

在两个团队项目之外，个人兴趣驱动下我本人编写了一个喜爱的图片网站的爬虫，对关注的画师的过往作品进行了爬取。爬虫本身性能不高故不会对服务器造成严重影响，而作品本身只用于个人欣赏，也没有向外传播，因此也不触犯专利权、著作权等。

1. 总结

综上所述，爬虫作为一门技术，其本身的特殊性导致在使用的过程中不免的会存在一定的法律风险，从不正当竞争，到专利权、著作权，一不小心就有可能踩坑。在利用爬虫的初期可能只是小事很难被注意到，但一旦有所疏忽，对产品未来就有可能有着毁灭性的打击。正确的方法即是在编写爬虫之初便认识到可能存在的风险，了解技术存在的底线，并采取措施对风险进行规避，保证项目或产品的顺利进行。

当然也存在着好消息，在信息聚合领域存在着“避风港原则”，即指在发生著作权侵权案件时，当ISP（网络服务提供商）只提供空间服务，并不制作网页内容，如果ISP被告知侵权，则有删除的义务，否则就被视为侵权。如果侵权内容既不在ISP的服务器上存储，又没有被告知哪些内容应该删除，则ISP不承担[侵权责任](https://baike.baidu.com/item/%E4%BE%B5%E6%9D%83%E8%B4%A3%E4%BB%BB/8921969)。而最新的互联网版权管理条例对信息展示增加了版权限制，未获得版权的内容，禁止对公众发布。对于一般的个人来说，爬虫及抓取的风险相当之低，而对抓取的数据的展示则要尤为重视。

在较为详细的了解爬虫及存在的法律风险之后再进行爬虫的开发，是防止在不知情的情况下触犯法律最好的方法。