

一场技术的博弈： 查重系统 vs 过关“妙招”

实习记者 翟冬冬

又是一年毕业季，“论文月”也随之到来。经过开题、中期检查之后，为了遏制论文造假，不少高校会借助文献检测系统来判断论文是否涉嫌抄袭。

为了帮助毕业生顺利“过关”，一些声称可降低重复率的“锦囊妙计”也在网上流传。日前，微信公众号“毕业有道”推出《了解知网查重原理，论文重复率3%以下也很容易》一文，其中介绍了变化措辞、翻译替代、图片变换等“降重”方法。

那么，这些“偏方”真的管用吗？

“偏方”会被系统迭代掉

随着计算机技术与互联网技术的发展，越来越多的文献信息被数字化。这些电子资料为工作、学习带来巨大便利的同时，也为抄袭、剽窃等行为提供了“方便”。

当前，论文相似性检测系统是反剽窃最有效的技术手段之一。国内的论文检测系统众多，其中中国知网(以下简称“知网”)的学术不端文献检测系统、北京万方数据股份有限公司(以下简称“万方”)的论文相似性检测系统和重庆维普资讯有限公司(以下简称“维普”)的论文检测系统是市场占有率较高的三大检测系统。

谈及这些系统的检测原理，中国科学技术信息研究所副研究员张英杰告诉科技日报记者，虽然各家的检测系统在细节设计上有所不同，但均基于相同的技术原理。它们都是将论文进行分解，而后对比资源库，并利用相关算法进行匹配度检测。

“匹配度检测也可称为文本片段比对，简单来说就是检测论文内容是不是与资源库中

的内容重复。”武汉大学信息管理学院副教授胡吉明说，论文上传系统后，系统一般会按照文章目录把文章“切”成几大段，之后再将其大段“切”成小段，将其与资源库中的文本内容进行比对。如果这一小段内容和数据库中的某个文本重复，那这部分内容就会被判定为涉嫌抄袭。

维普相关工作人员告诉科技日报记者，各家检测系统在设计上都尽可能仿照人的阅读方式，实现以机器代替人工，以求提升检测的精准度。

针对网上流传的“偏方”，多名业内人士向科技日报记者表示，系统设计的初衷是贴近人的阅读方式，以这个目标为设计出发点，随着系统升级，“偏方”自然会被迭代掉。据维普相关工作人员介绍，在算法上他们采用了多重防护机制，从而杜绝用“偏方”蒙混过关的情况出现；同时他们提供了“格式分析报告”，详细列出了送检文档中“图片”“空格”的数量，可供指导教师进行人工审查。

算法差异致反馈结果不同

当前，针对不同学历阶段，高校对毕业论文的重复率要求也不同。本科生毕业论文的重复率一般要求在30%或20%以下，硕士生论文则提至15%，博士研究生则规定要在10%甚至5%以下才算合格。现阶段论文检测一般由学校或学院组织，专业机构提供论文检测服务，相关检测数据或结果将会在系统前端呈现给学校。目前知网尚未开通个人检测服务，而万方、维普已开设个人检测窗口。

在送审前，不少应届生都会进行“自检”，依照学校对重复率的要求做进一步修改。在“自检”时不少学生会困惑，为何在两

家机构检测出来的相似度一个是15%，一个是20%？

维普相关工作人员向科技日报记者介绍，由于各家运用了不同的算法模型，基于不同的开发平台，因而造成相似度数值的差异。

各家的差别有多大呢？他打了个比方，如同苹果iOS系统和安卓系统一样，它们是按照不同思路设计出来的，各家的算法没法做同类对比。目前这方面没有相应的国家标准，各家根据自身的产品设计思路和原理模型提出了检测方法，于是就有了一定的差异。目前，市场上存在多种检测技术，如基于



字符串比较的方法和基于词频统计的方法等。”各家公司在技术上差距不大。”万方相关人员向科技日报记者介绍，不同算法可理解为不同的重复率判断方式。比如说，一篇文章中的某段话算不算抄袭，可能A算法认为80%的一致性是抄袭，B算法则规定70%的一致性是抄袭。“当然每种算法都有很复杂的计算模型，并没有例子中说的这么简单。”万方相关工作人员说。

抄袭判定标准有待更新

对于抄袭的定义，随着检测技术的发展也在改变。

南京大学信息管理学院教授苏新宁介绍，现在的抄袭可分为两种，一种是文字的抄袭，另一种是内容的抄袭。在论文检测技术发展的早期，一些作者会通过变换他人文章中词语的方式躲避检测，“这在早期的软件中不容易被查出来。”苏新宁说，但随着内容检测技术的发展，这种做法已经不灵了。

现在的检测技术并不是对文章进行逐字逐句的检测，而是用文章中出现的关键词和资源库中文章的词语作对比。如果这个词和资源库某篇文章的词大量相同，虽然顺序不同，但在检测系统中还是会被判定为相似。苏新宁举例说，如“我评价了这个问题”和“我对这个问题进行了评价”这两句话，如果后面的内容也都是这样颠倒语句，同样会被认定

“资源库也很重要。”张英杰表示，作为一个论文相似性检测系统，其资源库收录资源类型是否齐全、学科是否齐全、年限是否足够长、资源数量是否足够大等因素，都会对检测结果产生影响。从学科上来看，三家检测机构都已做到全学科收录；从收录文章的种类和数量来看，知网更具优势，万方在一些学科如医学领域有独家收录的文章，维普则在中文期刊数据库建设方面起步较早。

为是相似。也就是说，现在的检测系统除了对文章语法层面的词汇、句法结构进行分析，也会在一定程度上进行语义层面的检测。

说到相似和抄袭的不同，苏新宁坦言，现在的检测系统确实还存在一些局限。比如，对于一些公理性质的文字，系统也会判定为重复。如“一带一路”这类词如果在文章中多次出现，是否该被判定为抄袭，这值得商榷。

谈及对外文文献的比对检测，多位专家认为，我国目前在外文检测技术方面还比较薄弱。苏新宁介绍，一方面是因为外文文献数量大、种类多，相关资源库的建设工作尚处起步阶段。另一方面，中文和外语在语法和语义方面都存在较大差异，如何判定为抄袭也是个问题。但现在多数的期刊评审专家，一般都阅读过大量自己领域内的外文文献，对于送审论文是否抄袭了外文文献，会有自己的判断。

IT 辣评

点评人:本报记者 王小松

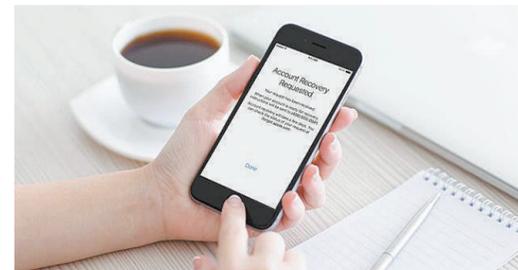
“0元打车”重现江湖 网约车市场硝烟再起



“前3单立减14元”“0佣金”“明日打车0元起”——一度平静的网约车市场最近热闹非常。继去年“试水”南京后，今年3月21日美团打车登陆上海，上线快车和出租车服务。据美团数据，在登陆后第三天，日订单量突破30万。3月27日，高德地图宣布推出顺风车业务，在成都、武汉两地率先上线，同时开启北京、上海、广州、深圳、杭州等城市的车主招募工作。

点评:一度平静的网约车市场又起波澜，许久不见的“补贴大战”也随之而来，无论是乘客还是司机都会觉得这是一件好事情。因为在滴滴之外，终于有了新选择。但长远看，烧钱是不可持续的，在让乘客得到实惠的同时也会让市场竞争再度陷入混乱。有了之前的经验和教训，相关管理机构也会更早介入。对网约车市场来说，便利、价格、服务是消费者最看重的3个因素，得用户者得天下，做好体验，用户会用脚投票。无论你是“老玩家”还是“新选手”，请牢记这一点。

苹果允许用户彻底删除ID 重隐私者得天下



北京时间3月30日，据彭博社报道，苹果公司表示将在未来几个月内更新其管理Apple ID的网页，允许用户下载苹果公司所持有的个人所有数据。该网站更新后，还将具备修改个人信息、暂停账号和彻底删除账号等功能。据了解，苹果公司制定新的隐私控制条款，旨在符合欧盟新的隐私保护法规《通用数据保护条例》，该法规将于2018年5月25日正式开始实施。

点评:脸书的数据泄露事件弄得满城风雨，其CEO扎克伯格面临前所未有的压力。拥有大量用户的苹果公司也不得不站出来表态。日前，苹果掌门人库克在接受美国媒体采访时表示：“脸书在个人隐私保护方面的水准太低，缺乏有效的保护……现在到了应该反思的时刻，应该想想下一步该做什么。”脸书数据泄露事件，或许也在某种程度上提醒了苹果，更严格的隐私保障策略也随之而来。在人工智能和大数据时代，重隐私者才能得天下。

微软分拆Windows部门 发展重点将从软件转向云计算



北京时间3月29日晚，微软宣布进行重大重组，操作系统Windows部门将与云平台Azure整合成为一个新的团队。Windows业务执行副总裁特里·梅尔森将离职，他曾在微软工作21年。这次调整后，微软重新成立两个新部门，目前负责产品的执行副总裁拉杰什·贾哈将负责用户体验与设备部门，而目前负责云计算和企业集团的执行副总裁斯科特·格斯里将负责云计算和人工智能部门。

点评:微软这家老牌IT企业的一举一动仍牵动着我们的神经，毕竟在这个地球上绝大多数电脑运行的都是由这家公司开发的操作系统。然而随着大数据、云计算和人工智能的发展，变局也随之而来。微软的业务重心发生转向也是自然而然的事情。与微软早期几乎将全部精力都集中在Windows上不同，微软首席执行官纳德拉上任以来，一直在将业务重心向云和智能方向转移。过去4年里，微软已成为云计算市场的重要玩家。因此，不必惊讶，新时代早已开启。

(本版图片除标注外来源于网络)

eSIM 时代，运营商的末日还是新生

第二看台

本报记者 刘艳

自近日中国联通宣布在国内首发 Apple Watch Series 3 的 eSIM 和 LTE 功能，并选择在上海、天津等 6 座城市率先启动“eSIM 一号双终端”业务后，坊间便传出“运营商的末日要来了”的声音。

eSIM，即嵌入式 SIM (embedded SIM) 的缩写。顾名思义，与传统可插拔的 SIM 卡不同，eSIM 是将 SIM 卡直接嵌入到设备中，最终让 SIM 卡槽消失，让设备拥有更轻薄的机身，并可实现与终端直接交互。它通过 APP 或者云端，在全球范围内将终端智能设备连接到所

选择的当地网络，让设备始终处于最佳网速的状态。

人们之所以有了这样的想法，是因为 eSIM 通过空中写卡的方式实现远程配置，理论上能够自由切换运营商网络。同时，eSIM 将不仅限于手机，一切可携带、便携式电子设备均可通过它实现上网、语音通话等功能，物联网因此又增加了新备注。

但这并不意味着在选择运营商上可以任性，你需要获得运营商同意，更要国家监管层面的放行。

面向更大的物联网市场

将传统 SIM 卡直接嵌入到终端设备芯片上，

让用户无需插入物理 SIM 卡，这是 eSIM 提供的看得见的便利。而 eSIM 的真正价值则体现在，它为物联网这个广阔的市场预设了连接条件，使移动通信的范围从手机向更多设备扩展。

eSIM 的概念并不是最近才有的，技术难度也远未到高不可攀。在 eSIM 的推广上，苹果与三星一改在专利战中的针锋相对，二者始终统一战线推动该项技术的落地。

但如果把 eSIM 仅仅看作是几个手机厂商的“兴风作浪”，这样的看法未免有些片面了。当频繁更换的 SIM 卡无法适应未来的物联网环境时，eSIM 卡的优势尤其是它的“一号多终端”的特性，将使具备 eSIM 技术优势的企业在物联网市场上大有作为。

虽然物联网的大规模实现还尚需时日，但未来无论是汽车、家电、智能设备，还是公共设施的互联都将高度依赖网络却是不争的事实。为此，国内的运营商已经在利用 eSIM 卡积极部署自己的物联网平台。

通信行业著名科技自媒体人王云辉说：“从人与家庭的连接到万物连接，eSIM 卡将为运营商带来新的增长点，但他们的商业模式也将面临一次前所未有的颠覆。”

王云辉强调，在以物体作为通信连接点的物联网江湖中，运营商面对的不再是单个消费者，而是整个垂直领域，还要面对业务模式、技术特征等方面的调整，这就要求电信运营商不仅需要在功能方面创新，更要在运营方面作减法，通过降低成本等方式获得更多客户、更高收益。

运营商间的竞争或将加剧

SIM 卡与手机联手走进了空前繁荣的智能手机时代，并使电信运营商虽然只有向用户销售 SIM 卡一种商业模式，却创造了二十多年的辉煌。虽然 eSIM 卡的出现不会直接颠覆运营商现有的商业模式，但可能会加剧运营商之间的竞争。

此前，中国联通率先与苹果公司开展 eSIM 终端产品合作，建成了全球首个运营商拥有自主知识产权的 eSIM 管理平台，研发出中国市场唯一的 eSIM 电子证书服务体系，完成了 IT 支撑系统的全网升级改造，推出了基于 eSIM 的消费物联网解决方案，也促成了这一次 eSIM 业务的登场。

这样的铺陈并不代表 eSIM 顺利落地。事实上，全球电信运营商在 eSIM 开展初期普遍选择了“不合作”的态度，因为 eSIM 就像电信运营商驻守“领地”的一道“壁垒”。

如通信行业专家马继华所说：“以 SIM 为基础，运营商完成了整个通信鉴权和客户管理，也掌控了整个产业链的核心。如今，几乎所有的业务都绑定在号码之上。”

好在，失去控制权对电信运营商来说只是一种“远虑”，毕竟所有的 eSIM 技术、标准以及号码管理，至今都是以运营商的技术路线为首是瞻，终端企业更必须与电信运营商一家一家谈合作协议，用户转网更换运营商还有号码和规则等门槛。

只是，运营商的竞争模式必须进行相应调整，才能让得到的远远超过失去的。



视觉中国

扫一扫
欢迎关注
畅游 IT 时空
微信公众号

