

## 百度六大知识类产品生产内容超10亿条

中国出版传媒商报 9月18日,百度知识垂类产品总负责人阮瑜在接受媒体采访时表示:过去十多年,百度知道、百度百科、百度文库等六大知识类产品累计生产内容已超10亿条。今年8月数据显示,百度APP日活已经超过2亿,其中大量用户每天都有知识相关的搜索需求。

8月12日知乎宣布完成F轮融资,总额4.34亿美元,由百度和快手联合投资。阮瑜表示,知乎后续会以小程序的方式进入到百度搜索,去满足专业、深度的内容需求,同时也会供给到信息流专业的内容里。阮瑜还强调,以百度百科、百度文库和百度知道为例,这三大产品已经和中国科协、

中国抗癌协会、人民日报数字传播、国家哲学社会科学文献中心等7000多家权威机构建立合作,近万名专家参与相关内容的编辑和审核。另外,百度百科已经有超过18万个词条得到科普中国项目权威专家的认证。百度还推出了百科医典项目,联合全国50多个权威机构,上百位院士、学会主委和行业顶级权威专家共建权威医学健康科普平台,计划今年内上线至少1000个优质疾病科普内容。

在百度AI和大数据等技术的支持下,百度还对优质知识内容进行了更多元化的呈现,如AR、全景、图表等富媒体形式,降低传播和认知门槛。今年,百度百科将联合专业内容生产者和机构,探索百科词条正文视频化。(晓雪)

# 工具类技术加速数字出版商业化

中国出版传媒报记者 陈莹

## 科大讯飞:语音合成助推有声阅读升级

如今,有声阅读已经成了现代化阅读场景下的强需求,有声书、音频更是出版业数字化转型的重要发展方向,语音识别、语音合成等相关技术的价值日益重要。科大讯飞依靠领先的语音技术可提供完整解决方案,目前,咪咕、QQ阅读、起点、追书、沃阅读、多看等上百个阅读资讯类平台都已接入了科大讯飞的语音合成服务。

传统人工配音方式不仅费时费力,还需要专业主播和剪辑人员,而现在通过语音合成技术,能够便捷实现文字到音频的转化,并且还支持背景音的自动匹配,极大地节省了人力成本和时间,提高效率。“合成配音的最大优势在于它可以7×24小时待命工作,不会有音质效果的浮动变化,并且可保证超高的合成准确率。”尤其是在针对长、难文本的情况下,语音合成能更有效地完成转化,提高有声书制作的效率。但科大讯飞“讯飞智声”产品经理刘焯秋并不讳言,相对于真人阅读,当前的语音合成技术在情感演绎上效果还不够好,需要继续优化算法效果。

目前,科大讯飞已服务中信出版集团等传统出版社、喜马拉雅等音频平台,得到等知识付费平台。刘焯秋介绍说,不同平台对语音服务呈现出迥异的诉求点,主要集中在如下四个方面。

第一类需要定制明星音库:为吸引更多的粉丝、声控等用户流量,客户需要用指定的明星发音人进行音频合成,也就是定制明星音库的需求,例如讯飞合作案例中用罗振宇的声音合成的《显微镜下的大明》。

第二类需要多场景发音人:不同的客户有不同的应用场景,例如小说、散文、科普等,而不同场景下所需发音人的韵律、音色、情感各有不同,客户对适合自己所需风格领域的发音人诉求较为强烈。

第三类需要个性化合成:个性化合成是指C端用户可用家人朋友或者自己的声音来合成音频,一方面满足尝鲜心理,另一方面声音也是一种陪伴,因此许多合

作客户为打造产品亮点,在这方面也会有比较强烈的诉求。

第四类需要高效高质量的合成效果:相比于真人播报,语音合成的优势之一就是能够有效节约时间、人力和物力成本,因此在保证质量的前提下,客户对语音合成成本的降低、效率的提高都会有比较高的追求。

为此,科大讯飞可以提供三种解决方案。

在线/离线语音合成产品:客户可在讯飞开放平台下载对应的技术开发文档,将语音合成能力集成到自己的工具或产品中,并根据需要购买或定制指定发音人,使用时只需通过相应的接口调用合成能力即可。

有声制作解决方案:客户只需将所需合成的文本发送过来,讯飞直接将合成好的音频发回给客户。

专业合成工具:目前讯飞提供的专业合成工具主要有两个,一个是更倾向于C端的产品——讯飞配音网站和APP,用户只需输入文本并选择发音人即可实时合成;另一个是面向B端的PC端合成工具,提供输入文本自动输出音频的功能,工具中会集成客户购买或定制的指定发音人。

“大V”加持正是如今有声市场的缩影,围绕作品、IP、作者、主播的全面竞争一触即发,科大讯飞能够为客户提供更高效、高质的语音合成服务及一体化的解决方案。

目前,科大讯飞能够提供更加多样的音库发音人,细化阅读场景,比如小说领域探索适合于玄幻小说、言情小说、悬疑推理小说的音库发音人,能够让有声阅读更加有趣生动。

针对IP、明星主播领域,科大讯飞可实现定制专属明星音库,吸引书粉、声控的用户群体。“借助于科大讯飞的个性化语音合成技术,给有声阅读市场提供更多个性化玩法,用户可以在平台上通过十句话来复刻自己的声音,用自己的声音读书讲故事,更符合现代人求新、个性化的需求。”刘焯秋表示。

网文翻译成本降低至1%、效率提升3600倍,术语校对准确率、查全率均超95%,每秒可校对超1000字;数据库搭建成本低至6万元……这些数据都是当前数字出版产业中工具类技术可以实现的效果。

当下,提供数字出版解决方案的平台和产品层出不穷,有些可覆盖从选题策划到营销推广的出版全流程,有些着眼于数据库、知识服务、在线教育等某一细分领域。而围绕有声书和智能编校的技术工具开发相对发展更为迅猛。

在人工智能、大数据、5G等技术赋能下,数字出版工具实现了成本降低、人力解放的目的,更加加速了出版业的数字化进程,帮助传统出版社、网络文学、音频平台等实现资源聚集和价值转化等商业目标。

《2018~2019中国数字出版产业年度报告》显示,2018年中国数字出版产业总规模8330.78亿元,其中,电子书收入达56亿元;艾媒咨询发布的《2018~2019中国有声书市场专题研究报告》显示,2018年中国有声书市场规模达46.3亿元,年均复合增长率为36.4%,2019年市场规模有望增至63.6亿元。面对这些可观的数字,出版业转型步伐随之加快,接入最具活力的技术平台,引入最具效能的工具产品,成为普遍的做法。目前,我们可以说,服务于数字出版各环节的工具类技术的市场潜力不容小觑。



三年间,全球出版市场中有声书品类的增长高居第一位。推文认为未来市场规模将扩展到30亿的海外用户,而且变现空间将变得更加广阔,可以通过付费阅读、广告、IP、线下纸质书出版等方式扩大营收。

有效盘活网文行业存量版权资源是推文科技的核心价值之一,在挖掘库存的同时,推文自身对原创内容保持开放态度。“推文不仅乐意与国内所有网文平台公司达成合作,同时也欢迎国内的独立作者、独立工作室即内容原创方进驻。推文与原创作者、原创机构同样采用版权授权、收入分成的合作方式,原创方将成为推文版权库中的一股重要力量。”童晔表示,甚至未来推文科技也不排除会吸收国外的原创作者或原创机构加入开放平台。

2017~2019年,“人工智能”写入政府工作报告,在这样的背景下,人工智能领域相关技术飞速发展,文字信息的处理效率也突飞猛进。随着人工智能、大数据等新技术融入出版产业链,“人工智能+出版”这个大方向也已经逐步渗透到多个细分领域中,“人工智能+校对”便是其中之一。

由浙江一意智能科技有限公司(以下简称“一意科技”)承接研发的专门校对稿件中使用不规范差错的工具——“术语通”,已成为全国科学技术名词审定委员会(以下简称名词委)科学传播项目宣传推广平台。

一意科技成立于2017年,公司研发团队聚焦自然语言处理、深度学习、数据分析等技术,并与国家相关语言文字规范相结合,经过3年积累,自主研发了

击整个海外小说市场。

童晔看重的是企鹅兰登、哈珀·柯林斯、阿歇特集团盘踞多年的海外大众出版市场,他们80%的营收依然来自纸质书,而他们尚未拿下的线上数字阅读市场就是推文科技的机会所在。

据了解,目前推文科技可将翻译内容一键分发至Kindle、Googlebooks、Applebooks等全球近50家海外主流数字出版平台,可以说技术和渠道是公司的核心竞争力,今年公司除了持续提升机器翻译的质量和继续完善AI生产系统流程之外,推文还考虑部署第三种语言及发展有声书。过去两

2019年,根据用户需求,名词委进一步开发了专门针对校对出版物中术语使用不规范差错的智能工具——术语通。术语通借助自然语言处理、大数据分析等技术,通过正确语料库建设,实现对术语使用不规范现象的智能判断。

作为首家语言智能交互平台作为专业校对云服务的开拓者,术语通率先将先进的语言智能核心技术通过互联网云服务对外提供。出版物质检技术为图书、期刊、报纸、网站等文字内容的质量检查和格式规范提供技术支持。平台还可以提供丰富的服务接入方式,WebAPI、Linux、Java SDK等接入方式,无系统平台及硬件限制,支持所有开放应用接入。

目前,人民邮电出版社、中原出版传媒集团、凤凰出版传媒集团、上海世纪出版集团、湖北长江报刊传媒集团、青岛出版集团等近30家出版机构提供了自动化转档、数据库自动化生产等服务。XBOOK平台已在科学出版社、法律出版社、中国纺织出版社、上海世纪出版集团、凤凰出版传媒集团、湖北长江报刊传媒集团等众多出版机构中应用。

目前,人民邮电出版社、中原出版传媒集团、凤凰出版传媒集团、中国科学出版集团等出版机构,中科院、社科院等机构主管的学术期刊,已将术语通纳入工作流程中,产品用户群体还在不断扩大,目前日活用户约2.5万。

## 推文科技:AI翻译使网文出海效率倍增

网文翻译成本降低至1%、效率提升3600倍,“网文出海内容开放平台”推文科技可以实现这样的超高效率正是依赖公司自主研发的人工智能翻译生产分发系统。这套全球首个针对网络文学的翻译生产分发系统,根据BLEU算法(一种机器翻译自动评价方法)的评测对比,在垂直领域的翻译质量已经超过谷歌。

不过,AI翻译只是推文科技网文出海业务系统的一环。该公司目标是打通网文出海全链条,帮助国内网络文学快速突破海外市场“天花板”。在推文科技网文出海内容开放平台funstory上,国内拥有海量网络文学内容的公司可以实现零成本翻译生产、分发变现的全链条数字出版。截至今年9月中旬,推文已与国内近30家网文平台签署版权合作协议;目前funstory.ai平台日更英文小说突破2600部,内容量位居行业第一。

推文科技创始人兼CEO童晔介绍说,推文科技不等同于百度翻译、谷歌翻译等开放性的“翻译工具”,推文与各版权方之间建立版权合作及收入分成关系,但并不提供免费翻译工具给版权方。

“以掌阅科技和推文科技的合作为例,在一个星期之内,推文科技完成了380本书的文本清洗、术语提取及翻译、人工标注、全本机翻、指代检查、封面英文化等工作,这几乎等于之前两三年时间内国内依靠人工翻译实现出海的网文的数量总和。”

推文科技目前打造的网文出海全链条包括版权库、AI翻译生产系统及分发变现渠道等各个环节,希望在未来1年实现超过10万部中国网文作品的海外投放,同时也试图用市场定价的1/10冲

## 术语通:“AI+校对”切入专业术语校对市场

一意科技大规模真实语料知识审核与智能校对云平台,并在此基础上开发了术语通。

近年来,国家新闻出版行政管理部门高度重视,采取了多项措施推广普及规范术语,比如:将术语规范使用纳入全国出版专业技术人员职业资格证书考试;将术语规范使用纳入干部颁标准;将术语规范使用纳入图书质量检查工作范畴。

作为国务院授权、代表国家开展术语审定事业的国家队,名词委自1985年成立以来,已累计开展了100多个学科的术语审定工作,公布了130余种、60万条权威术语。以此为基础,名词委积极拥抱新技术、探索新模式,主导开发了一系列相关产品,比如术语在线,已累计面向用户提供了超2000万次术语查询服务。

上海精灵天下数字技术有限公司(以下简称“精灵天下”)是上海张江数字出版文化创意孵化器的人驻企业。作为知识服务技术及运营公司,精灵天下可以向出版传媒机构提供面向全媒体移动时代的、AI和大数据驱动的新一代阅读视听、数字教育、智能化学术和知识服务生态技术平台系统。目前,该公司主推产品是火鸟文档助手和XBOOK平台,两个产品相辅相成,前者将不规范的原材料加工为规范的标准材料,从而为后者平台进行内容生产运营提供充足的标准化素材。

在精灵天下董事长覃云川看来,市场上有三类出版社对转档工具的需求比较大:一是存在海量存量资源的出版社,二是希望大幅提高出版数字化程度的出版社,三是融合出版进入运营期,对纸电内容同步出版需求较高的出版社。

火鸟文档助手恰好能够满足出版机构从排版文件到规范化xml文件的自动化加工需求,解决人工转档过程周期长、成本高、难以融入自动化融合出版流程的问题。

火鸟文档助手效率,单服务器转档速度平均为2秒钟/页,10分钟/本(按每本书平均300页计),每天(24小时)可以处理144本书。由于全自动进行,因此可以通过部署多个服务器进行扩容,以倍增的方式进一步提高效率。例如,2台服务器288本、3台服务器432本,以此类推。

目前,火鸟文档助手采用两种计费方式,一种是在线转档,按使用次数收费;二是为出版机构做本地化部署(可以集群),按处理内容页数收费,转档成本约为当前行业平均价格的一半,转档效率比当前行业平均水平提高数倍以上。最重要的是,它可以嵌入出版机构的自动化融媒体出版过程。

一方面,火鸟文档助手可以为互联网内容平台提供不规范内容(主要是大量的pdf)的自动化加工处理能力,把它们变成适合于自动化发布的规范化文章。另一方面,精灵天下致力于通过火鸟平台,在出版社的海量优质内容与互联网内容平台优质渠道之间,建立一种自动化分发的能力。

“这种应用模式拓宽了出版社的内容分发渠道,提升了出版社的数字化盈利能力,同时也为互联网内容平台提供了海量、优质的内容来源,是一个合作共赢的应用模式。”覃云川进一步介绍说,公司研发的SaaS工具XBOOK平台,更是提升出版效率和数字化能力的利器。“以前,开发一个数据库至少要60万元,现在只要6万元。过去2年间,公司把制作数据库的成本降低至1/10、自动化程度提高到90%、交付周期缩短了6倍。”

据了解,XBOOK平台可为出版机构提供包括融媒体、大数据、知识图谱、人工智能、电子商务、区块链八大功能的全媒体生产运营平台,主要满足出版机构多样化数字内容生产、运营、多种方式提供智能化知识服务等需求。

不久前上线的商务印书馆全媒体生产运营平台,就是精灵天下与之前历时近3年研发出来的,开创和应用了多项先进出版科技。拥有著作权近10项、联合专利6项,尤其是在融媒体、大数据、人工智能、知识图谱的融合应用方面实现诸多创新。

目前,火鸟文档助手已为商务印书馆、科学出版社、凤凰出版传媒集团、上海世纪出版集团、湖北长江报刊传媒集团、青岛出版集团等近30家出版机构提供了自动化转档、数据库自动化生产等服务。XBOOK平台已在科学出版社、法律出版社、中国纺织出版社、上海世纪出版集团、凤凰出版传媒集团、湖北长江报刊传媒集团等众多出版机构中应用。

精灵天下:转档工具+SaaS工具实现数字出版完整解决方案