

拓尔思大数据： 知识图谱赋能公共治理

股票代码 300229
广州拓尔思大数据有限公司
拓尔思知识图谱研究院

□本版撰文 信息时报记者 李丹
实习生 张文科

疫情期间,一张基于大数据整合的疫情热力图,可清楚显示各地疫情实时动态、传播情况等,在疫情防控中发挥着重要的作用。今年2月,广州拓尔思大数据有限公司(以下简称拓尔思大数据)开发的“拓尔思流行病学溯源追踪系统”,依托各地的一体化在线政务服务平台,整合疾控、公安、通讯、医疗机构等多方面的数据,发挥着快速甄别密切接触者、分析疫情传播路径、评估疫情扩散风险等作用,这是大数据助力公共卫生管理的一个缩影。

从IC卡到大数据 为城市构建治安感知网

据了解,在天河科技园科韵园区成长起来的广州拓尔思大数据有限公司,前身为广州科韵大数据技术股份公司,成立于2001年,由20余名国企下海员工联合创办。

“我们是天河土生土长的企业,公司成立之初主要是做流动人口管理。早在20年前,广州的治安问题比较突出,尤其是外来人口多,城中村人员复杂,犯罪案件频发。为有效管理租住在城中村的外来人员,我们开发了一套流动人口管理系统软件——通过录入外来人员照片、指纹、身份证数据等信息,由公安机关发放IC卡暂住证。”拓尔思集团知识图谱研究院首席科学家王亚强介绍,该系统实现了流动人口信息的上网,与公安系统的网上追逃数据库联系在一起,如果某人是在逃人员,只要来登记,很快就会被发现。

此后,该公司开始在“互联网+泛安全”领域深耕,并于2018年底被拓尔思集团并购,更名为拓尔思大数据有限公司,如今该企业已从最初的20余人小团队扩展到现在的200多人的规模,并跻身于国内大数据行业的前列。据悉,截至今年3月,拓尔思大数据共拥有三项发明专利及六十多项软件著作权,自主研发了水晶球分析师平台、水晶分布式数据库、安拓知识图谱平台、海贝分布式搜索引擎等多种大数据服务产品,现已为国家多个部、委,二十多个省、一百多个城市的信息化建设提供服务,涉及公共安全、金融等多个领域。

构建知识图谱 走向人工智能的阶梯

近年来,大数据、人工智能概念火热,但如何搭建起二者之间的关系,实现人工智能在更多场景的应用呢?作为“水晶球知识图谱”的总设计师,王亚强给出了解决方案:“先把大数据构建成为



拓尔思大数据研发团队。信息时报记者 李丹 摄

自己领域的知识图谱,形成机器学习、模型训练的知识资源,正是通往人工智能的阶梯”。

“知识图谱是模拟人类的思维,围绕实体、实体的属性、实体的关系搭建起来的知识架构。”王亚强现场演示了谷歌知识图谱的应用,比如在谷歌上输入某名人关键词,页面右边随即出现了关于此人的简介、人物关系、新闻报道等信息,形成了一份直观的信息档案。知识图谱实现了对数据的汇总和集纳,使得用户不再需要根据文字片段和网站名称来逐个判断哪条链接更有价值,大大节省了用户搜索的时间。王亚强介绍,与谷歌类似于百科全书的通用型知识图谱不同,拓尔思开发的安拓知识图谱平台属于领域型的知识图谱,融合了大数据、可视化技术,对海量多源异构数据进行融合、关联、标注和知识化处理,实现知识构建管理、知识语义检索、智能文本提取等功能,用于经侦、政务管理、应急管理等多个专业领域。

在互联网时代随着网购、移动支付等的应用普及,经济诈骗案时有发生。为此,拓尔思大数据基于知识图谱技术开发了金融风险控制系统,可提前预防和减少经济诈骗的发生。据介绍,按该系统通过整合企业的成立日期、法人代表、股权结构、人员关系等大数据,找出哪些企业涉及非法集资、反洗钱、传销、个税诈骗等经济犯罪,随后根据企业与企业、人员与人员之间的关系,整合企业的信用度,预测企业的经济犯罪风险。“如今经济犯罪非常高发,而且范围很广。公安经侦部门可借助该系统及时介入,今年起我们已与广州的公安部门开展了此类合作,提供相关服务。”王亚强介绍道。

探索“知识图谱+” 提供更多智慧解决方案

据介绍,除了医疗、治安、金融领域,拓尔思大数据积极探索大数据、知识图谱技术在更多场景的应用。如“知识图谱+新闻传播”,在新闻传播的舆情分析方面,知识图谱的引入可实现对全球媒体、国际智库、国际社交媒体等海量国际传播数据的数据整合、数据关联与价值发现,进而协助有关部门开展国际舆情研判、国际传播能力分析、效果评估等工作,从而全面提升对外宣传工作的精准性、针对性和有效性。

在政务智能服务方面探索“知识图谱+政务服务”。知识图谱的技术基础使得机器学习等人工智能算法得以运用。以智能服务机器人为例,首先利用知识构建工具对海量非结构化知识源文档进行导入提取,一键生成标准化的问答知识图谱,实现知识梳理智能化。随后在机器学习等人工智能算法的运用下,机器人不断地学习公众的问法,不断学习,从而不断补充未知问题、完善政务办理过程中的智能引导服务。

目前,有100%的副省级城市、89%的地级以上城市、49%的县级城市都已经开展了智慧城市建设。王亚强认为:“不少智慧城市都将数据集中管理,但不同的数据之间的关系尚未建立。”在他看来,“知识图谱+智慧城市”能够把数据之间的关系和数据一起存储,和各种算法模型一起配合,形成真正的“城市智脑”。王亚强预测,知识图谱技术也许就是智慧城市的下一个制高点。