

以知识图谱为工具优化高校科技成果统计效率

荣誉

长春师范大学，吉林长春，中国

【摘要】传统的高效科技成果统计一般采用表格形式将当年的成果数据按类别进行数量和性质等方面的区分，这种方式能够体现总体的发展水平，但在数据充分挖掘的当前成果应用背景要求下各种表格无法体现的趋势性的数据内在联系及细节变化。知识图谱是结构化的语义知识库，通过图结构表达实体、关系和属性，这种表达方式能够将数据不为人知或难以察觉的相似性及发展性特征表现出来，为科技成果的数据深度挖掘提供有效手段。本文结合 2024 年高效科技成果部分数据，采用知识图谱的手段对其进行处理，发现了一些传统表格统计方式无法体现的数据特征，为统计结果的深度有效应用提供了一条新视角路径。

【关键词】知识图谱；高校科技成果；统计分析；应用研究

【基金项目】吉林省教育厅科学技术研究项目（项目编号：JJKH20241019KJ，基于知识图谱的高校科技统计应用研究）

1.引言

在高校科技统计领域，当前的国内外研究重点主要体现在不同高校类别科研及统计的特殊性研究、科研投入与产出关系、科技成果统计应用方法等单维度等方面[1]。针对统计的方法研究及统计结果间的多层次内在联系研究方面研究内容较少。当前的单维度线性关系研究能够体现过去每年哪些领域有哪些科研人员产出了多少成果，这些能够在一定程度上指导当前的高校科研管理工作，但这种传统统计方法对数据的运用不足，无法深度体现数据结果之间的内在逻辑关系，比如成果与成果之间的横向比较、技术职务与成果水平的关系比较、不同年份科研项目规划在时代背景下的先进性比较等能够有效指导科研管理向科学化、精细化、可前瞻、有规划方向进步的科技统计方法与多层次结果结论呈现的研究非常少。此类方法虽能实现成果数量、类型等表层指标的快速汇总，但难以动态追踪学科交叉趋势、团队协作模式及成果转化路径等隐性关联[2-4]，导致统计结果碎片化、决策支持能力弱化[5-6]。知识图谱技术通过实体-关系建模与语义推理机制，可构建多维度科研数据关联网络，为解决上述问题提供了技术突破口。然而，现有研究多聚焦于宏观科研领域知识图谱构建[7-9]，针对高校场景的轻量化图谱设计、异构数据自动化对齐及统计效能量化评估仍缺乏系统性探索。在当前信息化时代，高校科学技术成果统计领域有新要求，成果统计不能仅仅体现冰冷的数字，要将数字之间的逻辑关系

和内在影响因素通过统计的方式呈现出来，有效指导科研管理部门的科学管理。很多专家提出了一些该并高校科技统计方法，如设计科学合理的统计指标、确定合理的统计时间和频率、建立稳定的统计队伍、完善科技统计的工作框架及采用先进的统计方法等方面[10]。整体上看高校科技统计正向着科学化、数字化的方向发展，但在智能化时代这些方法显然已经跟不上统计需求的发展。基于知识图谱的高校科技成果统计应用研究的研究目的是利用当前人工智能领域知识图谱的方法对高校科技成果进行统计与应用，直观体现当前高校科研项目设置是否合理、成果是否先进、总体科研能力是提高、科研成果是否对科技进步有贡献等深维度统计结果与结论。本项目的研究必要性在于当前成果统计方法局限于单一报表、先验性项目分类等传统数字罗列，没有深层次的横向比较与纵向比较，不能直观体现当前科研现状并有效指导高校科技项目研究论证发展及人才科研政策倾斜方向的问题。而知识图谱作为当前人类自动化向智能化迈进的关键方法在 DEEPSEEK、GPT、OPEN-AI 等人工 AI 领域已体现出传统方法所不具备的多层次、多维度、多交联对象特征，该方法运用于高校科技统计中能够展现出其特有的数据驱动高效率特征，从专家视角、用户视角、应用视角、数据视角多个维度上体现科研统计成果的内在逻辑关系，指导科技项目指南确定、科研人才培养、科技成果转化应用，将国内高校科研在管理上提升一定水平。

本文采用具有跨域关联、亮点预置特征的知识图谱处理方法对高校科研数据进行统计,能够实现对高校科技统计的数据驱动下的效率革命,呈现传统统计方法不能展现的统计数据与结果的内在联系,有效指导该领域的科研规划与管理,丰富高校科技成果管理的理论研究,探索知识图谱在科技统计领域的应用模式。

2.当前高效科技成果统计方法和结果运用存在的问题

高校科技统计是指用科学的统计方法对高校各项科技活动状况进行系统全面的调查,取得科学、系统、直观的数据,对统计数据的整理、汇总和分析,掌握高校科技活动的规模、结构和布局,发现其变化的规律,规范高校科研管理工作,为制定高校中长期发展规划和科研政策提供科学统计数据。高校科技统计为科研管理工作提供了重要的统计数据,科研管理本身就是科学决策,科学决策依靠科技统计数据,科技统计反过来服务于科学决策。高校科技统计的内容主要包含了科技人力、科技经费、科技机构、科技成果、国际科技交流、科技期刊、高校创新信息调查表等。这项工作是一项覆盖面广、综合性较强的工作,需要从科研、人事、财务、资产、基建、国际合作交流、学科建设和研究生管理等部门收集数据,并经过统计人员汇总、系统填报、系统校验、省教育厅主管部门审核等四个环节才能顺利完成。因此,如何做好高校科技统计工作,为教育科技行政管理部门提供真实、准确和科学的数据,必须通过高校每一位涉及科技创新与科技管理工作人员的共同努力才能达到。边缘化的倾向且普遍存在。在统计工作实施运行中,一般单位往往由于未设立专职的科技统计岗位,统计人员队伍体系不健全或统计人员专业能力与熟练程度达不到工作要求,一定程度上影响统计工作的正常开展。同时,管理部门在开展统计工作时,也存在临时“任务”突击进行统计的做法,即在短时期内发动并分别咨询各相关部门,填写提交数据或上门收集数据,按要求进行数据处理后经领导审核再报送上级主管部门。实际工作中,由于各部门各位统计人员对于统计指标的分类标准、统计的口径及计算方法的了解与掌握不完全一致,直接制约损害了统计结果的准确性和真实性。由于缺少综合开放式的管理,统计数据为高校各部门共享并服务的功能尚未有效的开发利用,高校教职工也没有得到及时有效的公共服务,不能有

效地利用科技统计工作成果来及时指导改进科技管理工作,也会在一定程度上对统计工作人员的工作缺乏应有的了解、理解与支持的现象。

目前科技统计工作的重心,还停留在科技情况普查、数据搜集、简单处理与汇总阶段,并没有通过深入的科学分析,形成研究分析报告,并应用于本单位的科研管理,为科技人员提供信息服务,进入能为高层领导作决策发挥积极参考作用的统计分析研究阶段。因此,力求迅速组织填报准确的数据,及时完成数据上报,仍是目前统计工作人员的主要任务,也是比较普遍的对统计工作性质的认识。单位与教师对统计工作的重视与支持程度,将直接影响科技统计工作的发展和科技统计功能作用的发挥。以吉林大学为例,早在2000年前就开发了动态信息管理系统。该系统提供对本校科研机构情况、科研项目经费及进度、获奖成果情况、科技论文和专利的查询与统计等功能。这种统计方法和手段是当前的主流方式,能够在一定程度上为本学效、本省的科技成果应用及科技规划提供参考,但在知识爆炸的当前社会,数据表面的意义已无法有效支撑大国竞争的科技发展需求。

3.知识图谱数据统计与挖掘技术

3.1 知识图谱的定义与构成

知识图谱(Knowledge Graph)是一种结构化的语义网络,用于描述现实世界中的实体及其相互关系。它以图的形式呈现,节点代表实体或概念,边表示实体间的关系或属性。知识图谱旨在将信息组织成易于理解和处理的结构,广泛应用于搜索引擎、推荐系统、自然语言处理等领域。知识图谱是结构化的语义知识库,通过图结构表达实体、关系和属性。其构成要素包括节点(实体)、边(关系)、属性(描述信息)。

3.2 知识图谱的应用领域

知识图谱技术主要应用于智能搜索、推荐系统、数据分析三大方向。其含义分别为通过语义理解提升搜索精度、基于知识关联提供个性化推荐、挖掘数据间的隐含关系。知识图谱在数据分析中的应用广泛,涵盖数据整合、关系挖掘、语义搜索、预测推荐、知识推理、可视化、自动化报告生成等方面。通过结构化知识表示和推理,知识图谱提升了数据分析的深度和效率。知识图谱这些特征对高校科研数据的整合处理挖掘分析及预测应用而言是非常合适的。

3.3 知识图谱在科技成果管理中的优势

知识图谱通过结构化的方式组织和管理科技成果信息，显著提升了管理效率和应用价值。以下是其在科技成果管理中的主要优势：
信息整合与共享：整合多源异构数据，打破信息孤岛，促进共享。
关系挖掘与分析：揭示成果间的关联，发现潜在合作机会。
语义搜索与智能推荐优势：支持复杂查询，提升搜索效率和准确性。
知识推理与发现优势：通过逻辑推理发现新知识，支持决策。
可视化与交互分析优势：图形化展示成果及其关系，帮助直观理解。
自动化报告生成优势：自动生成结构化报告，提升效率。
知识产权管理优势：有效管理

专利、商标等知识产权信息。科研绩效评估优势：提供全面的科研绩效评估工具。

4.知识图谱统计技术用于高校成果数据分析案例

某高校科研管理部门希望利用知识图谱技术整合和分析近年来的科研成果数据，包括科研人员构成、项目、论文、专利、获奖情况等，以提升管理效率并为科研决策提供支持。本案例以 2024 年吉林省高校人员构成情况为基础，采用知识图谱的办法对其整合处理，发现内在规律。下表 1 为 2024 年吉林省高校人员构成情况，对该数据进行基于知识图谱的数据分析。

表 1.科技人力资源情况表（2024 年吉林省高校人员构成）																
		合 计	其中： 女性	教师技术职务系列人员						其他技术职务系列人员						辅助 人员
				小计	教授	副教 授	讲师	助教	其他	小计	正高 级	副高 级	中级	初级	其他	
	编号	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15
合计	01	29608	16028	19749	4369	6902	6327	1976	175	9774	616	2332	3945	2656	225	85
自然科学	02	5815	3064	4917	1285	1783	1516	296	37	893	48	278	383	171	13	5
工程与技术	03	11744	5532	9399	1766	3280	3185	1102	66	2321	156	644	998	485	38	24
医药科学	04	9805	6263	3847	980	1371	1056	396	44	5914	376	1236	2290	1850	162	44
农业科学	05	1367	648	1110	287	354	396	71	2	256	19	72	113	49	3	1
其他	06	877	521	476	51	114	174	111	26	390	17	102	161	101	9	11
博士研究生	07	10762	4955	9378	3028	3548	2728	38	36	1384	234	504	514	121	11	0
硕士研究生	08	11463	6855	6619	694	1922	2513	1415	75	4844	182	907	2109	1608	38	0
大学本科	09	6925	3932	3696	634	1419	1074	505	64	3229	181	826	1244	802	176	0
大学专科	10	329	210	48	9	13	11	15	0	222	4	32	69	117	0	59
中专及以下	11	129	76	8	4	0	1	3	0	95	15	63	9	8	0	26
29 岁以下	12	1577	1038	988	2	7	214	691	74	583	0	5	66	417	95	6
30-34 岁	13	3810	2307	2406	47	248	1380	657	74	1395	0	18	546	778	53	9
35-39 岁	14	5919	3457	3326	156	1049	1749	356	16	2586	11	246	1251	1040	38	7
40-44 岁	15	6944	3740	4848	676	2187	1813	164	8	2090	54	637	1087	290	22	6
45-49 岁	16	4029	2045	3008	775	1432	728	71	2	1012	106	405	431	65	5	9
50-54 岁	17	3964	2053	2650	1076	1216	333	25	0	1283	169	602	455	50	7	31
55-59 岁	18	2835	1236	2053	1194	736	110	12	1	765	224	411	109	16	5	17
60 岁及以上	19	530	152	470	443	27	0	0	0	60	52	8	0	0	0	0

1) 对上表进行数据预处理与实体识别
人员类别：教师技术职务系列人员（教授、副教授等）、其他技术职务系列人员（正高级、副高级等）、辅助人员。
学科分类：自然科学、工程与技术、医药科学等。
学历层次：博士、硕士、本科等。
年龄分段：29 岁以下、30-34 岁等。
性别属性：女性占比。
属性标注：实体： 工程与技术_教授
学科： 工程与技术 职称： 教授 人数： 1766 女性占比： 5532/11744 ≈ 47.1%

2) 关系建模
学科→职称分布（如工程与技术包含教授、副教授等）。
学历→年龄分布（如博士研究生主要集中在 35-39 岁）。
交叉关联：
学科与学历的关联（如医药科学中博士占比）。
职称与年龄的关联（如教授多集中在 50-54 岁和 55-59 岁）。
衍生指标：
高职称占比（教授+正高级占总人数的比

例)。

青年人才储备(29岁以下人员占比)。

3) 关键分析维度

学科规模:工程与技术(11,744人)和医药科学(9,805人)是主力学科。农业科学(1,367人)与其他学科(877人)规模较小。

职称分布:高职称集中度:自然科学教授占比($1285/4917 \approx 26.1\%$)高于工程与技术($1766/9399 \approx 18.8\%$)。

医药科学特殊性:教师系列人员仅占38.7%(3,847/9,805),但其他技术系列人员占比60.3%(5,914/9,805),可能涉及临床或交叉岗位。

高学历年龄分布:博士研究生(10,762人)集中在35-39岁(3,326人)和40-44岁(4,848人),反映科研主力年龄段。硕士研究生(11,463人)中,30-34岁(2,406人)和35-39岁(3,326人)占比较高,可能为青年骨干。

学历断层:中专及以下学历(129人)中,辅助人员占59人,显示低学历人员多从事辅助岗位。

整体性别均衡:女性占比54.1%(16,028/29,608),但学科差异显著,医药科学女性占比最高($6,263/9,805 \approx 63.9\%$)。工程与技术女性占比最低($5,532/11,744 \approx 47.1\%$)。

职称晋升路径:教授职称在50-54岁(1,076人)和55-59岁(1,194人)达到峰值,显示职称晋升周期较长。副教授集中在40-44岁(2,187人),可能为晋升教授前的关键阶段。

青年人才瓶颈:29岁以下人员中,助教(691人)和初级职称(417人)占主导,高级职称极少(教授仅2人)。

4) 知识图谱应用建议

人才结构优化:学科均衡-加强农业科学与其他学科的高职称人才引进。青年培养-针对30-34岁硕士群体设计晋升通道,提高博士研究生中青年教授比例。

性别政策建议:在工程与技术领域推动女性科研人员支持计划,优化性别比例。

退休预警与梯队建设:55-59岁教授占比高(1,194人),需提前规划退休后的人才补充。

40-44岁副教授群体(2,187人)是未来教授储备重点。

跨学科协作潜力:医药科学中“其他技术

职务系列人员”占比高,可探索与教师系列的跨学科合作模式。

重点领域人才强化:扩大工程与技术领域教授规模,优化医药科学交叉岗位配置。青年人才激励:为30-44岁群体设计快速晋升通道,增加科研资源倾斜。性别平衡计划:在工程领域设立女性科研基金,提升女性职业发展支持。退休衔接机制:针对55岁以上高职称人员,建立“导师制”传承经验。

5. 结论

通过知识图谱技术,高校可以优化科技成果统计效率,具体方法包括数据整合与清洗、实体识别与关系抽取、知识表示与存储、知识推理与扩展、语义搜索与智能查询、可视化与交互分析、自动化报告生成和科研绩效评估。案例分析表明采用知识图谱的方法能够深入挖掘现有表格化高校科研数据的内在联系,提出行之有效的建议。这些方法显著提升了数据管理效率和分析能力,广泛应用于科研项目管理、技术转移、合作网络分析、成果展示和知识产权管理等领域。该领域目前存在的不足包括目前数据采集范围和质量仍有待提高,部分高校数据存在缺失或不准确的情况,知识图谱的构建和更新需要进一步优化,以适应科技成果的动态变化和复杂关系,未来可将知识图谱技术应用于高校科技成果转化的全过程,探索与其他新兴技术(如人工智能、大数据)的融合应用,进一步提升科技成果管理的智能化水平。

参考文献

- [1] 韩月.基于知识图谱的高等教育科研立项管理研究[D].2012,西北农林科技大学.
- [2] 邓三鸿,苏新宁,徐瑶.基于中文数据库的知识图谱绘制方法及应用:以创新研究论文的分析为例.现代图书情报技术,2011,27(1),55-62.
- [3] 刘春辉,陈宏,王慧.知识图谱在科研选题中的应用价值探讨.现代医药卫生,2017,33(22),3534-3536.
- [4] 胡伟.学会科学知识图谱,搞定文献 so easy!河南科技,2017,(13),40-42.
- [5] 陈秋华,戴永务,陈钦,苏时鹏,余建辉.基于 bibliometrix 知识图谱的可视化分析——以福建农林大学为例.福建农林大学学报(哲学社会科学版),2018,21(1),85-92.
- [6] 李忠云,李岩,李太平.如何评价大学高层

- 次人才——基于 CiteSpace 可视化分析.高等农业教育, 2017, (4), 15-20.
- [7] 刘叶玲, 孙继湖, 陈宝国.基于知识图谱的效率可视化研究.系统工程理论与实践, 2018.38 (4), 1026-1037.
- [8] 陈悦, 陈超美, 刘则渊, 胡志刚, 王贤文.知识图谱的作用.科学通报, 2015, 60 (10), 910-918.
- [9] 侯剑华, 李旭彦.基础科学研究前沿格局的知识图谱解析[M].北京: 科学出版社, 2019.
- [10] 王雅娇, 刘育兵, 等.基于 CNKI 数据的知识图谱分析[J].企业科技与发展, 2018, (9): 148-151.