



# 科技人才政策细粒度知识图谱的 构建与应用研究



作者：曹树金，龙影，曹茹烨



汇报：曹树金，山东理工大学

日期：2022年7月12日

# 目录

## CONTENTS

01

引言

02

研究设计

03

图谱构建

04

应用分析

05

结论

### 1.1 研究背景

- 科技是第一生产力、人才是第一资源，科技人才的作用得到广泛认可和重视。
- 习近平总书记在二十大报告上强调要深入**实施人才强国战略**，坚持尊重劳动、尊重知识、尊重人才、尊重创造，完成人才战略布局，加快建设世界重要人才中心和创新高地。其中，实现具体落实目标，**离不开政策保驾护航**。

目前科技人才政策文本分析和应用所面临的问题包括**政策数量过多**导致科技人才政策服务提供与适用需求不对称、**政策文本内容丰富**但缺乏**细粒度揭示**导致难以快速定位到需求内容及所在位置、**政策间关系隐藏复杂**导致政策相关者难以快速找到政策来源和依据、政策间内容交叉导致难以快速对两个或者多个涉及相同主题的政策文本内容进行对比分析。

## 1.2 研究意义

- 信息化程度不断加深，政策相关者使用需求也不断精细化，如何快速有效、多维度深层次查询、分析政策文本内容，成为政策文本语义挖掘分析的关键问题。因此，科技人才政策文本细粒度分析相关研究引起了学术界的高度关注。
- 传统的解决方式大都是基于人工收集、整理并理解政策信息，因此容易造成人力成本高、效率低下、信息缺失、组织方式单一、信息融合利用难度高、对比分析难等问题。

本研究基于社会发展实际需求，聚焦于科技人才政策文本，根据用户需求构建高质量领域细粒度知识图谱，助力科技人才政策服务高效精准智能化发展，以求有效解决信息查询、分析成本高、效率低等问题。对政策文本的深层次语义价值展现与利用展开研究，以期为深入实施人才强国战略、创新驱动发展战略等提供一定支持。

### 1.3 科技人才政策相关研究现状

- 虽然国内对科技人才政策的研究较国外起步晚，但国内学者也取得了丰硕的成果。国内学者主要是从政策发展历程回顾、现状分析及未来演变趋势进行研究，其中涉及到特定类别科技人才的政策演变历史与基本动向、特定区域科技人才政策影响因素探究、地区科技人才政策实施效果评估比较、科技人才政策制定优化路径等，多数从政策外部特征进行深入分析，完成了政策研究的基础工作。
- 近年来科技人才政策的研究热点是时间纵向追踪分析和多维度计量分析，且研究成果相对较多。科技人才内容横向对比分析研究成果数量虽有增加，但主要都是借助质性研究方法对比国家之间、地区之间科技人才政策文件优势，包括：政策制定部门、政策文种、政策工具、政策目标等，以提出完善现有的科技人才政策体系建议。所以，目前对科技人才政策文本内容的知识要素分析不够全面，对其内容关联性、异同点的研究更为少见。

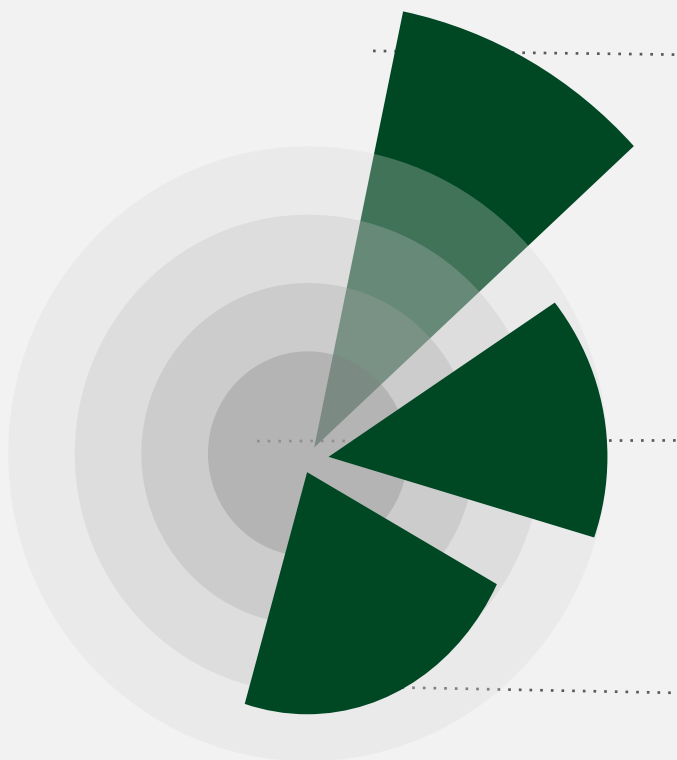
### 1.3 科技人才政策文本分析方法技术现状

- 目前科技人才政策文本研究成果主要集中在量化分析，所采用的政策文本分析方法技术主要集中在三大类：**政策工具类**、**文献计量类**和**文本挖掘类**。
  - (1) 政策工具类：该类方法通常是借用 Rothwell 等提出的政策工具法进行定量研究，如通过建立政策目标和政策工具网络，识别不同时期的核心政策节点，确定政策主题演化过程。
  - (2) 文献计量类：采用统计学的方法对政策文件进行政策主题、关键现状等描述性统计分析。
  - (3) 文本挖掘类：这类方法是采用文本挖掘方法，借助于机器学习、数据挖掘等 IT 技能，对政策文件的内容进行情感分析、语义分析。
- 上述方法技术对于政策的研究多聚焦于外部环境和实施成效，**较少从政策工具视角深入关注政策文本内容本身**，且以往针对政策文本的研究分析更多地关注单篇政策文本外部属性的量化分析，**忽略了政策单个文本自身及多个文本之间内部具体内容知识之间的关系**，**缺乏对科技人才政策文本的深层语义分析和历史纵向演变对比分析**。

### 1.3 科技人才政策知识图谱研究现状

- 面向科技政策领域构建知识图谱能够将知识赋能科学技术、科技人才等以知识密集型领域为代表的领域。已有学者构建以科技项目、学术论文、专利、科技动态等为主要数据源，以科技成果、作者、机构、科技项目、主题关键词等为主要实体，以支持面向科技领域的学术影响力评价、比较优化、关联挖掘为主要目的领域知识图谱。
- 在构建科技政策领域知识图谱的过程中，预先定义的实体与关系模型能够在较大程度上影响领域知识图谱质量及应用效率，如在领域相关的数据分析、关联挖掘、问题发现等场景中的应用。部分学者已对科技政策领域知识图谱的实体与关系模型展开研究，但由于科技人才政策文本在行文结构、语言特点及精准度、专业性、时效性等方面与其它类型的科技文本存在不同的要求，需要解决的问题也不同，所以在构建科技人才政策知识图谱时其本体模型、知识抽取、实体消齐、知识推断等也需要采取不同的方法技术。

## 2.1 研究内容



### 01.科技人才政策细粒度知识图谱本体模型构建

梳理科技人才政策文本特点和知识图谱相关理论和技术基础后，结合多维政策分析框架，提炼细粒度知识图谱的实体关系类型，从语义层面构建科技人才政策细粒度知识图谱本体概念模型。

### 02.科技人才政策细粒度知识图谱的知识抽取

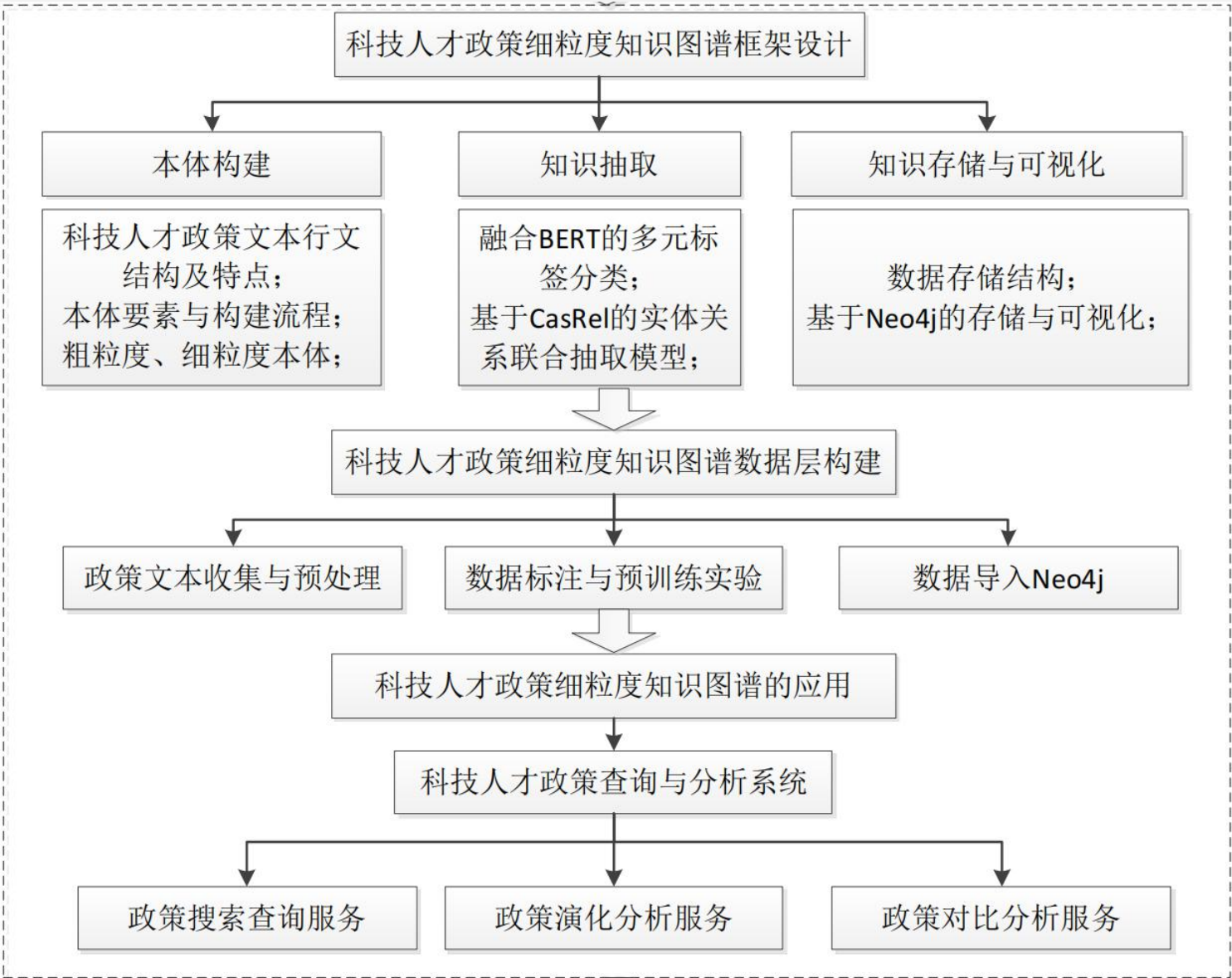
- (1) 政策文本多元标签分类；
- (2) 政策文本细粒度知识抽取；
- (3) 政策文本细粒度知识存储与可视化。

### 03.科技人才政策细粒度知识图谱的应用

结合政策相关主体的实际需求，从政策搜索查询、演化分析、对比分析服务三方面展开科技人才政策细粒度知识图谱的应用场景分析。



2.2 研究技术路线图



### 2.3 实践难点

#### 难点一

关于政策文本分析框架制定与量化评估研究成果较多，关于科技人才政策语义分析研究可参考文献资料较少。尤其是在科技人才政策细粒度知识图谱本体概念模型构建中，各个维度下的细分实体类别概念较难界定。

#### 难点二

政策文本存在多源异构性，包含结构化与非结构化信息，数据筛选与预处理难度较大。并且政策文本知识抽取技术多样，难以选择适配度较高的算法模型。

#### 难点三

科技人才政策相关主体需求越来越多样化、精细化，对科技人才政策细粒度知识图谱的应用方面提出了较高要求。

## 2.4 解决方法

融合自上而下、自下而上的本体构建方式，对现有政策本体进行改良，邀请领域专家评估鉴定。

调研各政策制定部门、政策执行者、政策受益者等相关主体的使用场景与具体需求。

### 01. 解决方法

充分调研分析科技人才政策文本语言与结构特点、已有本体研究成果、政策相关主体实际需求。

### 02. 解决方法

### 03. 解决方法

尝试多种政策文本解析技术：融合Bert的多元标签分类模型和实体关系联合抽取模型CasRel标注框架。

### 04. 解决方法

### 3.1 验证分析

#### 01、政策多元标签分类

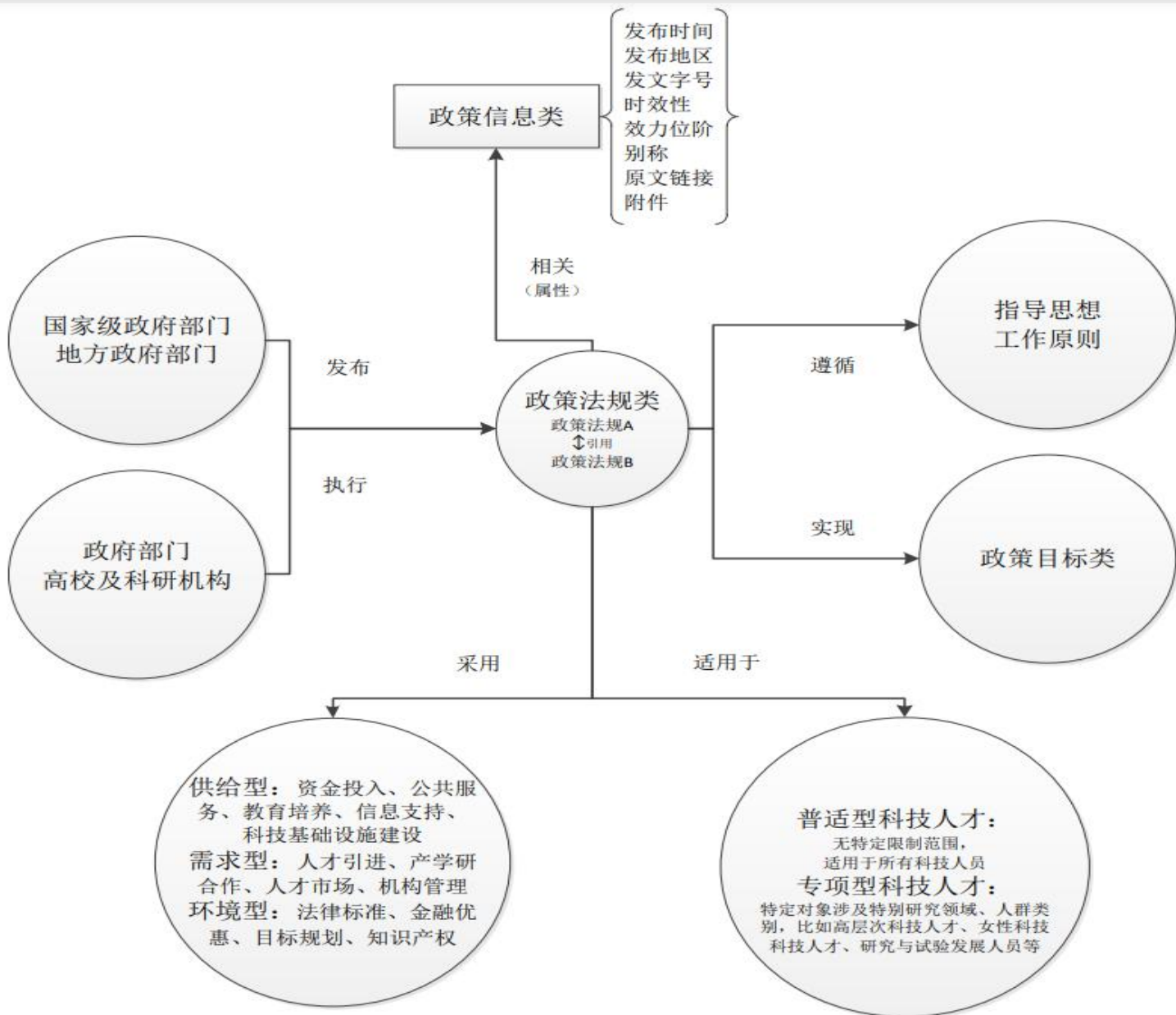
- 政策目标分类：人才引进、人才培养与发展、人才激励、人才管理与服务和科研管理。
- 模型选择：基于多元标签分类基础模型Bert-base-chinese构建科技人才政策文本的目标分类模型
- 数据标注与模型训练：人工标注4000份政策文本，采用随机抽样的方式将其以8:1:1的比例进行模型训练。本次分类模型评估指标F1值为0.80，准确率为0.88，效果较好。

#### 02、实体关系联合抽取

- 模型选择：基于CasRel标注框架构建科技人才政策文本的实体关系联合抽取模型。
- 数据标注与模型训练：人工标注700份政策文本，采用随机抽样的方式将其按照7:2:1的比例划分为训练集、验证集、测试集进行模型训练。本次 CasRel 联合抽取模型评估指标F1值为0.77，召回率为0.70，准确率为0.85，效果较好。

### 3.2.科技人才政策细粒度知识本体模型

每一个概念类的细分实体类别都遵循所属概念类的关系和属性，粗粒度本体模型中的属性集可以直接映射到细粒度本体模型概念中来，所以本研究提出的科技人才政策细粒度知识本体模型的细分实体关系如图所示，共 **32** 个细分实体，**15** 种关系。





3.3 本体模型

北大法宝  
PKULAW.COM

【法宝引证码】CLI.12.2163461

福建省公务员局、福建省人力资源开发办公室关于引进高层次人才专业技术职务任职资格确认有关问题的通知

制定机关：福建省公务员局 福建省人力资源开发办公室

发文字号：闽人发〔2011〕72号

公布日期：2011.05.04

施行日期：2011.05.04

时效性：现行有效

效力位阶：地方规范性文件

法规类别：职位职称

福建省公务员局、福建省人力资源开发办公室关于引进高层次人才专业技术职务任职资格确认有关问题的通知

闽人发〔2011〕72号

各设区市公务员局（人事局），平潭综合实验区管委会办公室，省直有关单位人事部门：

为贯彻落实中共中央组织部、人力资源和社会保障部《印发〈关于支持留学人员回国创业的意见〉的通知》（人社部发〔2011〕23号）、《中共福建省委办公厅省人民政府办公厅关于印发〈福建省引进高层次创新创业人才暂行办法〉等三个文件的通知》（闽委办〔2010〕2号）精神，加大高层次人才引进力度，提升职称工作服务发展的水平，现就“985”工程高校、中国科学院系统和海外引进的已聘专业技术职务的高层次人才专业技术职务任职资格确认工作通知如下：

一、从国家“985”工程高校、中国科学院系统引进的聘任满1年以上的教学科研人员，与我省高校按照原单位聘任的专业技术职务，签订3年以上期限聘用合同的，由我局（办）确认其专业技术职务任职资格。

3.4 粗粒度知识获取

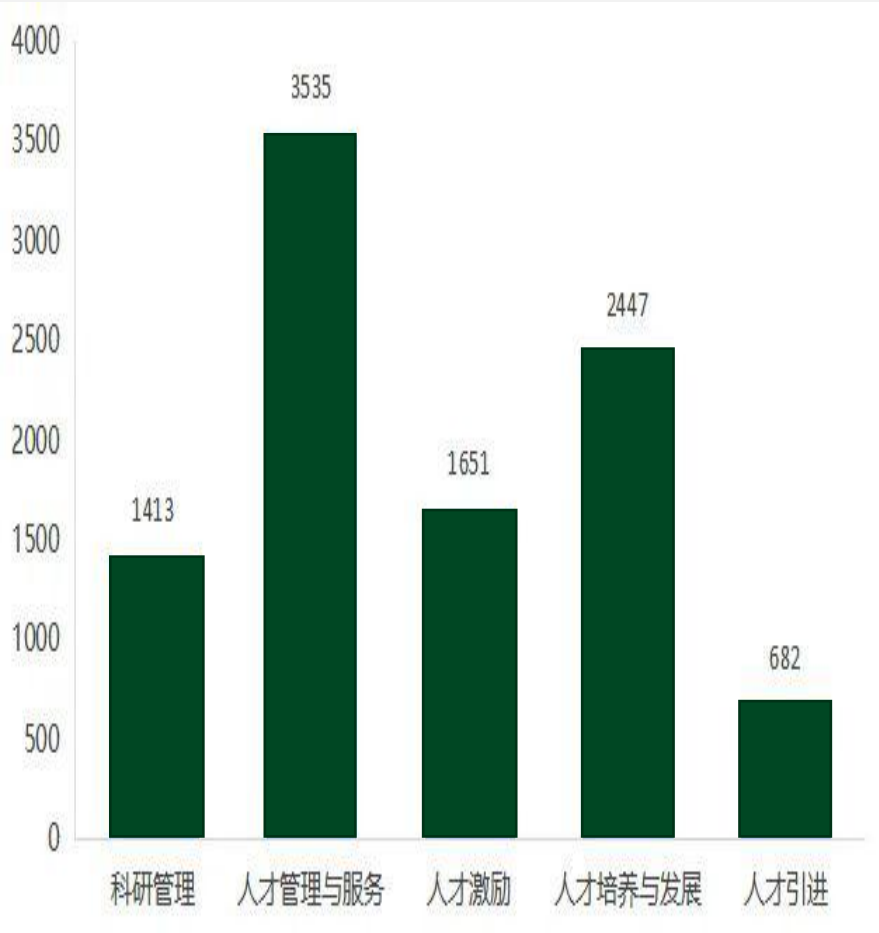
序号	字段名称	数据类型	备注
1	政策标题	文本	政策法规的名称
2	发文部门	文本	政策法规的具体发文机构
3	发布时间	Date	政策法规的具体发布时间
4	发布地区	文本	政策法规的具体发布地区
5	发文字号	文本	每条政策法规独有的标识
6	时效性	文本	政策法规目前是否有效的表征
7	效力位阶	文本	政策法规的效力等级
8	原文链接	文本	每条政策法规对应的URL

科技人才政策结构化文本数据库数据字典

政策标题	时效性	效力位阶	发文机构	发文地区	发文字号	公布日期	施行日期	原文链接
福州市公务员局关于填报2013—2014年度百千万工程人选、博士后和青年高层次人才访学进修需求情况的通知	现行有效	地方工作文件	福州市公务员局	福州市	榕人引〔2013〕13号	2013.03.25	2013.03.25	<a href="https://www.pku-law.com/lar/2af7438f20811d5dc5ef5fabf5389ceabdfb.html">https://www.pku-law.com/lar/2af7438f20811d5dc5ef5fabf5389ceabdfb.html</a>
苏州市经济和信息化委员会关于认真做好2013年度企业高层次人才培养工作的通知	现行有效	地方工作文件	苏州市经济和信息化委员会	苏州市	苏经信人培〔2013〕3号	2013.03.08	2013.03.08	<a href="https://www.pku-law.com/lar/b290a2a155147cbaa8b82238feb7a03fbdfb.html">https://www.pku-law.com/lar/b290a2a155147cbaa8b82238feb7a03fbdfb.html</a>
.....	.....	.....	.....	.....	.....	.....	.....	.....
北京市教育委员会关于印发《北京市属高等学校高层次人才引进与培养三年行动计划(2013年—2015年)项目管理办法》的通知	现行有效	地方工作文件	北京市教育委员会	北京市	京教人〔2013〕2号	2013.02.19	2013.04.01	<a href="https://www.pku-law.com/lar/04b774a73ddc54abe842e660e13c0c73bdfb.html">https://www.pku-law.com/lar/04b774a73ddc54abe842e660e13c0c73bdfb.html</a>

科技人才政策文本结构化数据示例

3.5 细粒度知识抽取



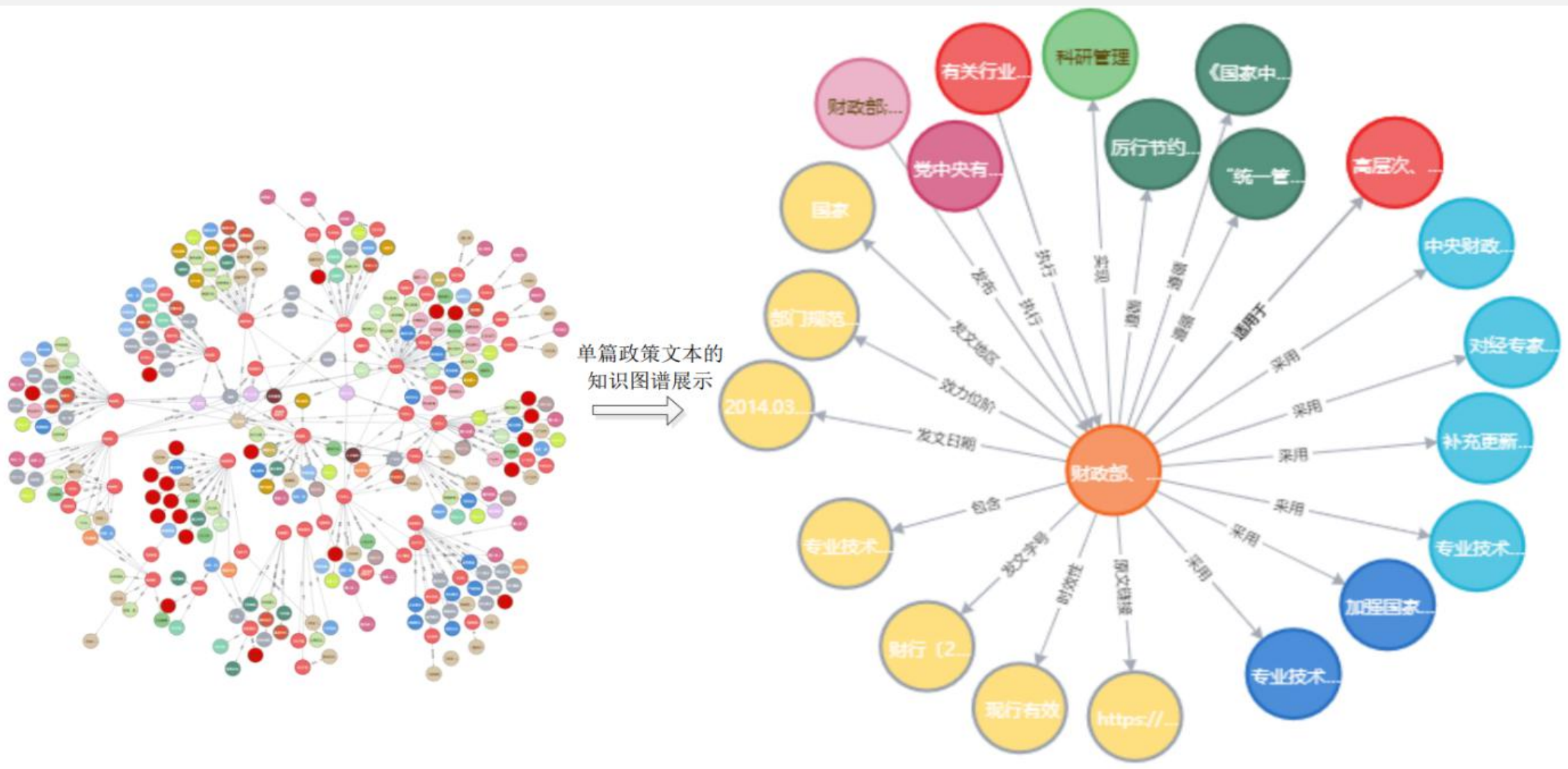
基于Bert模型的政策目标分类结果

政策条文内容	模型抽取结果 (头实体类别，头实体值，关系，尾实体类别，尾实体值)
<p>我委《关于申报国家文科基础学科人才培养和科学研究基地本科学科点的通知》（教高〔1994〕9号）下发后，有关高等学校领导和教师非常重视，按照文件要求，认真研究，提出文科教学改革思路、措施和建设“文科基地”的目标，积极申报“文科基地”。为做好这项工作，我委于1994年11月25日至27日在北京召开了国家文科基础学科人才培养和科学研究基地评审会，邀请有关专家对38所高等学校所申报的72个学科点进行认真评审。经研究，我委批准国家教委直属高等学校33个基地点，中央有关部委和省市所属高等学校16个基地点（见附件）。</p>	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，适用于，专项型科技人才，国家文科基础学科人才
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，引用，政策法规，《关于申报国家文科基础学科人才培养和科学研究基地本科学科点的通知》
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，遵循，指导思想，文科教学改革
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，遵循，指导思想，建设“文科基地”
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，采用，科技基础设施建设，积极申报“文科基地”
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，采用，产学研合作，国家文科基础学科人才培养和科学研究基地评审会
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，采用，科技基础设施建设，批准国家教委直属高等学校33个基地点
	政策法规，国家教委关于批准国家文科基础学科人才培养和科学研究基地学科点的通知，采用，科技基础设施建设，中央有关部委和省市所属高等学校16个基地点

科技人才政策条文内容抽取结果示例



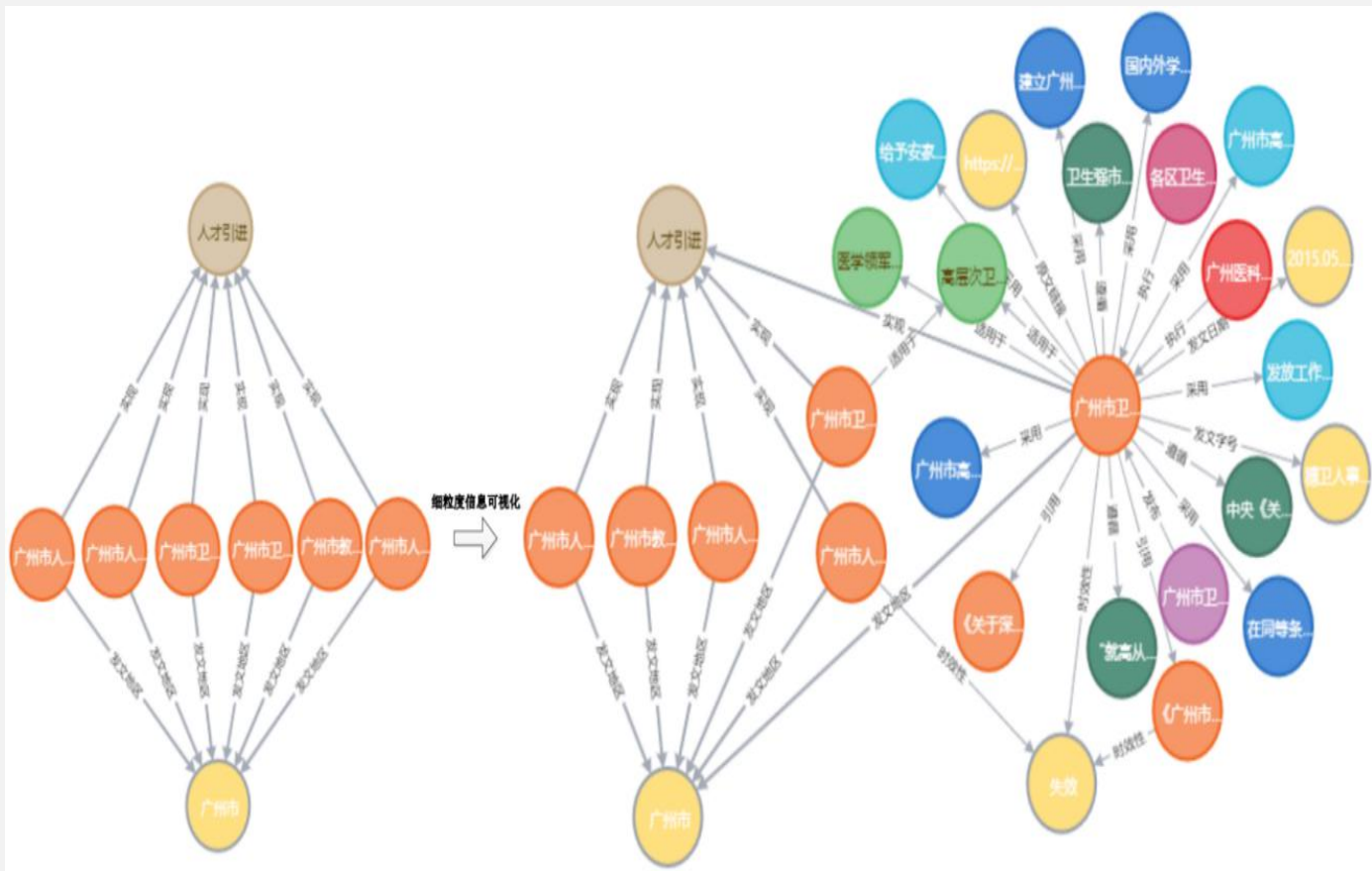
## 3.6 知识存储与可视化



## 4.1 政策搜索查询---用于人才招聘

➤ 通过科技人才政策查询功能，全面深入了解科技人才政策可以提高招聘的效率和招聘人才的质量。

- 机构可结合自身具体需求和检索结果，在科技人才引进时制定相应的引进计划，申请政策适用，制定科技人才发展规划，以提高科技人才质量和竞争力。
- 科技人才应聘时可根据自身实际情况寻求筛选各地方机构的科技人才政策信息，帮助自身在全面了解相关信息的基础上做出更好的决策。



广州市关于人才引进的相关政策及政策关联信息

#### 4.1 政策搜索查询---用于人才培养与发展

- 通过深入了解其他相同类型的单位对相应科技人才的技能水平、历史成就等要求，可为各单位招聘相应的科技人才制定科学专业、系统化、个性化的培训方案提供参考。



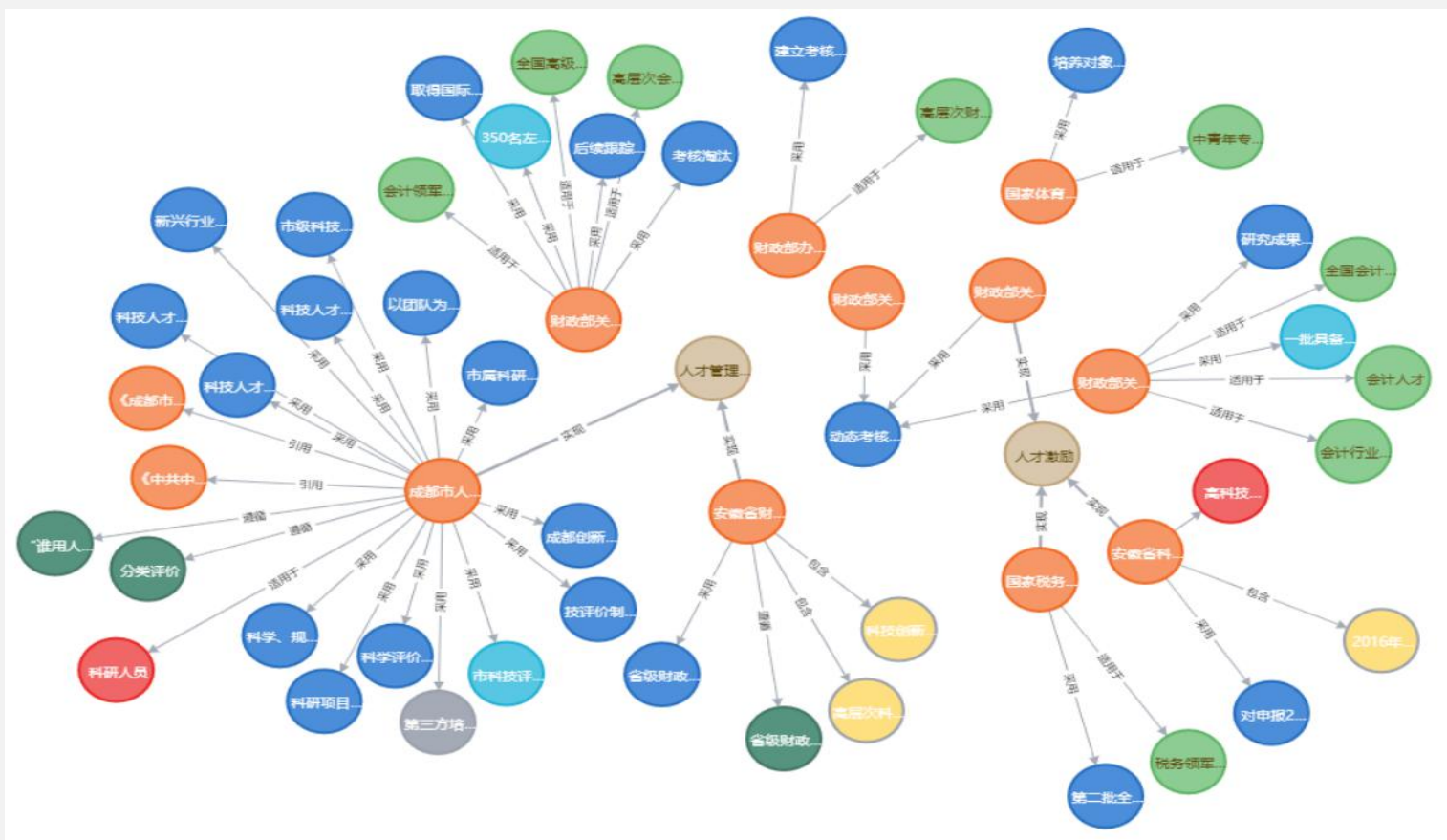
- 人才自身也可通过科技人才政策细粒度知识图谱进行职业规划，可以依据自身所属的专业领域、技能水平、历史成就等细粒度信息，了解到各地、各单位对科技人才的需求程度、重视程度、政策优惠等方面政策规定。

## 科技人才政策“卫生健康”可视化结果



## 4.1 政策搜索查询---用于人才考核与评价

- 基于本研究所构建的知识图谱，可以查询到科技人才政策的细节，如人才类别，帮助政策制定者快速分类科技人才，根据人才的不同特点进行个性化政策定制服务。



- 政策机构即可根据已发布的针对不同领域的科技人才政策考核与评价政策制定符合本地特色的政策。
- 科技人才可查询各地方机构对相关科技人才政策的考核标准与规划评价制度，根据相关标准有针对性地提升自身水平。

## 4.2 政策演化分析

- 根据本研究构建的知识图谱，可以了解到相应的政策网络的政策主体结构演化规律、政策目标演化规律、政策工具演化规律。
- 也能依据政策之间是否存在引用关系判断特定主题的政策演化路径。
  - 政策相关者可以更好的了解科技人才政策之间的关系和作用，分辨出具有核心影响力和重要地位的政策法规，进一步预测相关科技人才政策潜在的需求和政策倾向；
  - 帮助建立和完善科技人才制度，创新人才的培养、激励、使用等，以提供更好的服务。

## 4.2 政策演化分析

- 以政府部门角色为例展示“演化分析”服务。比如，某政府部门调研关于服务外包人才相关政策的演化路径；
- 政府还可结合政策时效性、发文日期等政策相关信息进行细粒度分析。

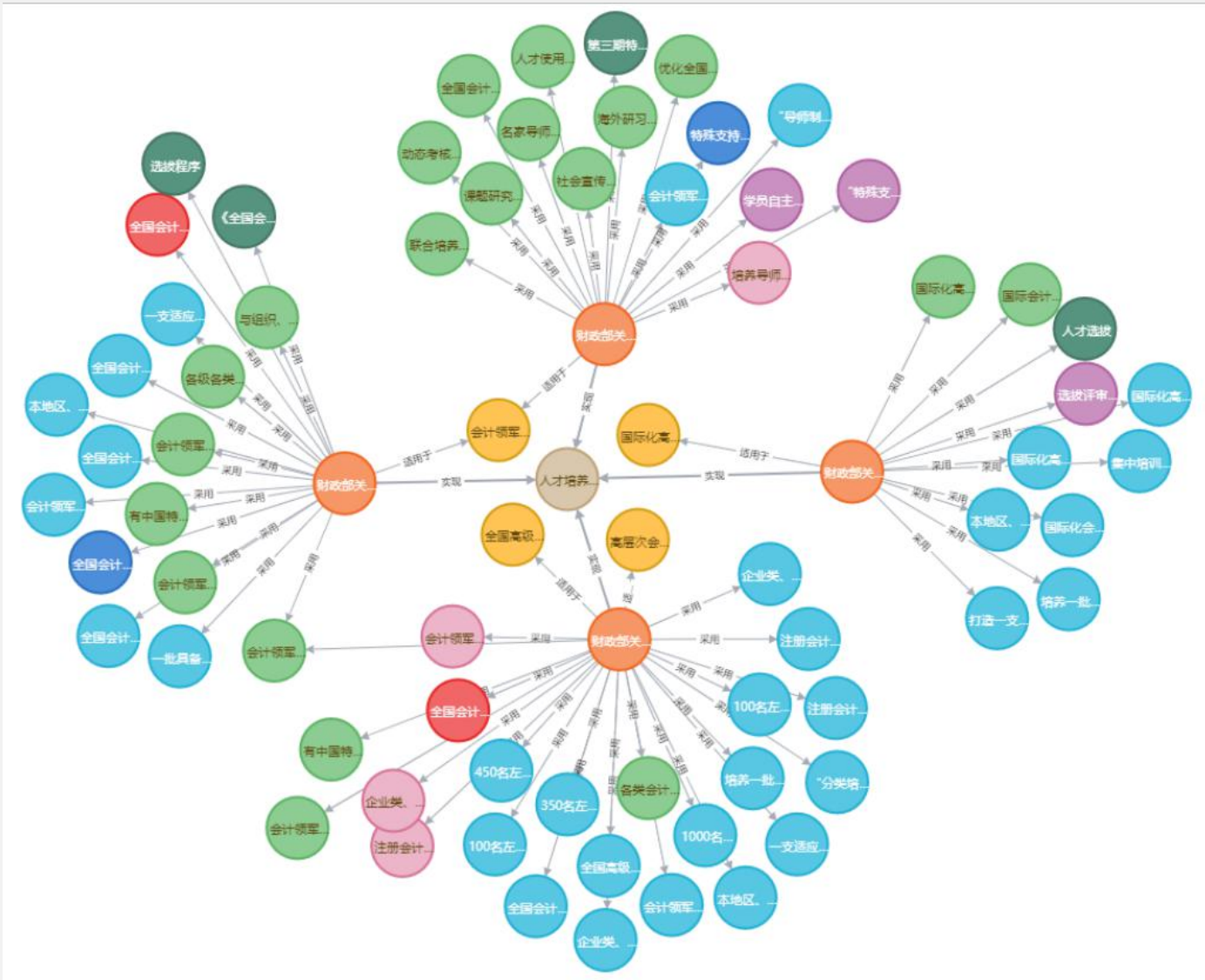


## 4.3 政策对比分析

- 对比分析科技人才政策可从政策目标、政策环境、政策来源、政策工具、政策对象、政策相关信息等多个方面进行综合比较，为政府部门制定更加符合本地实际情况的科技人才政策提供参考和借鉴。
  - 在政策目标对比分析方面，可比较不同地区、不同国家的科技人才政策的目标和理念，例如是否强调引进高端人才、是否强调培养本土人才、是否强调创新驱动等；
  - 在政策工具方面，可比较不同地区的科技人才政策措施的广度和深度，例如税务政策、人才贡献奖励、龙头企业政策、科研项目扶持等内容，看哪些政策的作用更加显著。

### 4.3 政策对比分析

- 某政府部门可对比分析国家发布的多个为实现会计领域科技人才培养与发展而发布的政策文本在政策工具、政策环境等多个方面的异同点。如各政策之间的横向对比可得出各机构所采用政策工具所属的细分实体类型的差异，纵向对比则可得出机构所采用政策工具的偏好。
- 进一步精准化制定和实施科技人才政策，创新政策工具，开发政策创新空间，提升政策的执行效能和落实水平。



# 国家关于会计领域科技人才培养相关政策信息



## 5.1 结论

1、本研究结合国家在加强“人才强国”“科技强国”建设方面的实际背景下，聚焦于科技人才政策领域，围绕政策相关者实际应用需求，提出的科技人才政策细粒度知识图谱的构建框架及技术路线是可行的；

2、科技人才政策细粒度知识图谱的应用可实现多样化、精细化服务。本研究结合政策相关者实际需求与研究现状，讨论了多种科技人才政策细粒度知识图谱的数据分析维度及应用场景，主要包含科技人才政策搜索查询、演化分析、对比分析等服务模块。

## 5.2 不足

### 研究不足

1

本研究所收集采用的科技人才政策数据只是科技人才政策的部分数据，在数据预处理和清洗的过程中难免造成部分数据丢失。

2

实体类型、属性定义和关系定义大多是依据政策计量研究、经验和统计分析确定，可能是不完善的，进而造成科技人才政策细分实体类别不够细致、关系稀疏。

3

本研究采用图数据库Neo4j可视化展示科技人才政策细粒度知识图谱，功能和展示效果目前较为简单，今后可借助更多的新兴技术扩展更多功能以符合实际需求。



谢谢!