



# 基于开源情报的科技人才 信息聚合与场景化推荐方法研究

汇报人：李怡然

作者：李怡然 何艳菲 李峥 胡天阁 彭海清 宋培彦\*

项目资助：国家社科基金项目“语义驱动的科研人员学术专长发现研究”

天津师范大学管理学院  
2023年7月12日

# 目录 / CONTENTS



研究背景



研究方法



研究过程



总结展望



# 研究背景

研究背景

研究方法

研究过程

总结展望



中华人民共和国中央人民政府

www.gov.cn



首页 | 繁体

习近平出席中央人才工作会议并发表重要讲话



中华人民共和国中央人民政府

www.gov.cn



首页 | 繁体 | 英文EN |

中华人民共和国国民经济和社会发展第十四个五年规划和  
2035年远景目标纲要

针对当前科技人才信息分散造成的“信息孤岛”以及人才岗位不匹配两个突出问题，采用开源数据对其信用和影响力进行识别，以在一定程度上提高人才评价与情报服务能力。



# 相关研究

研究背景

研究方法

研究过程

总结展望



## 人才信息关联模型

- 使用分散式语义Web技术构建的FOAF模型
- 通用欧洲科研信息格式CERIF模型

## 科技人才画像

- 上海“全球高层次专家信息平台”

## 人才按需推荐

- 内容特征、方法设计、用户行为



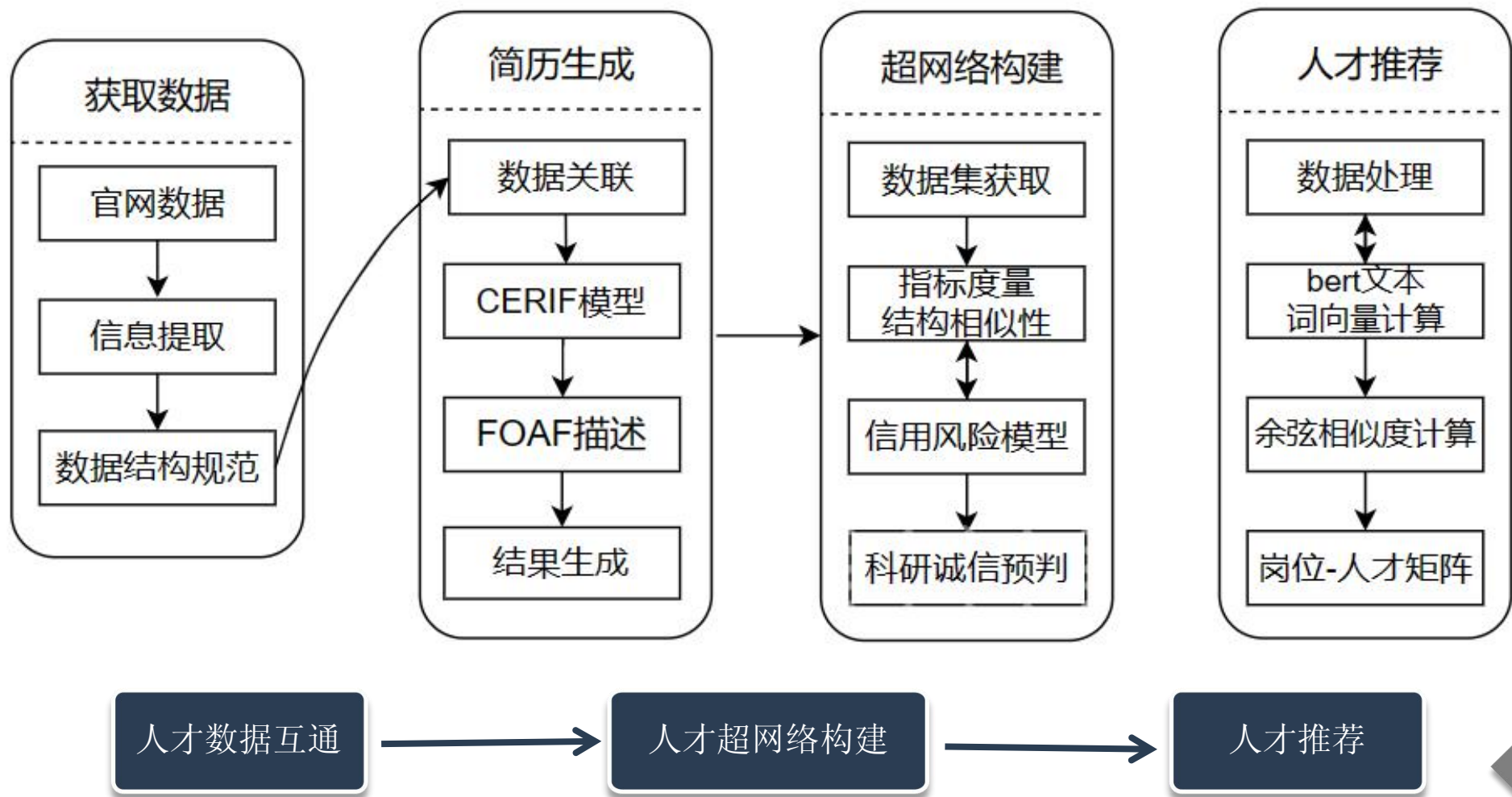
# 科技人才画像与推荐框架

研究背景

研究方法

研究过程

总结展望







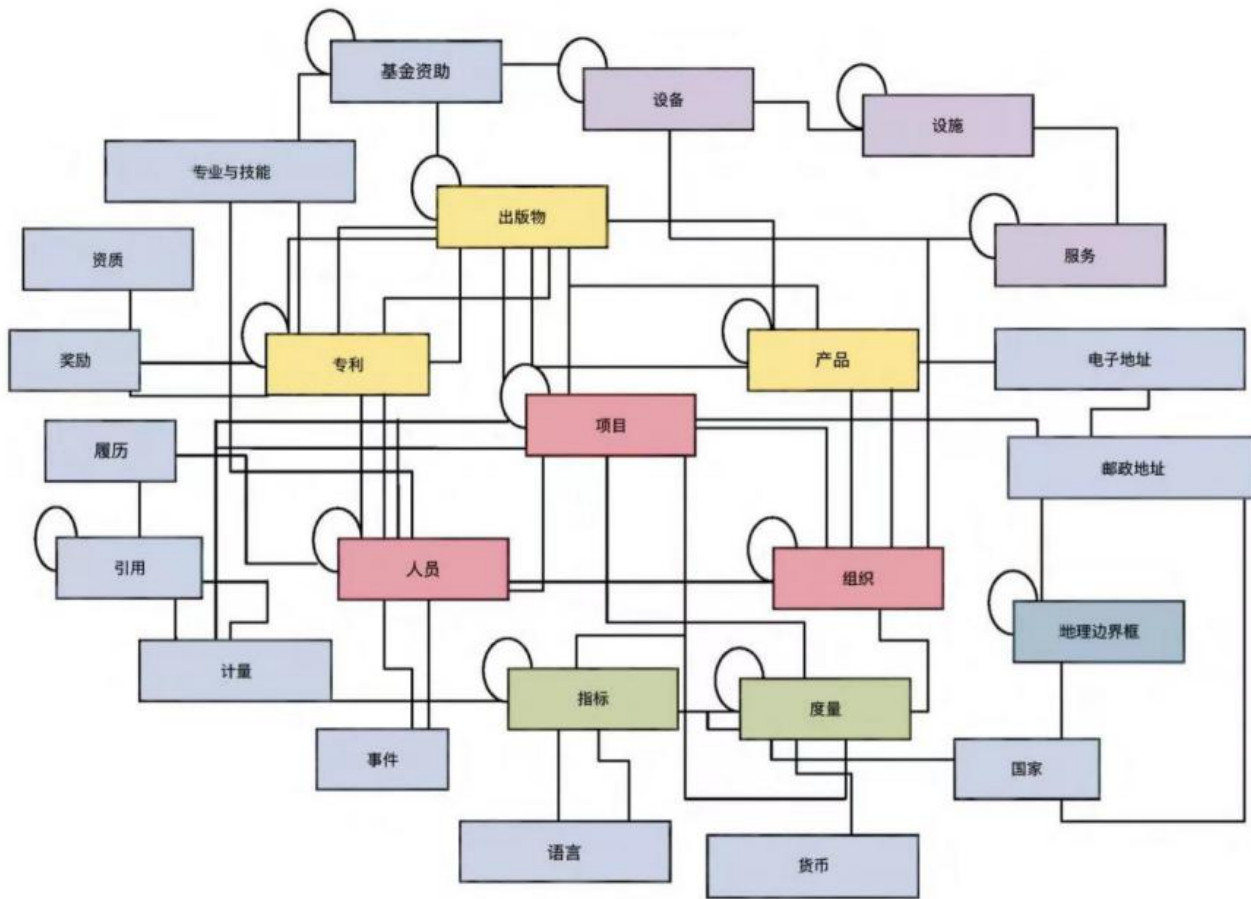
# 人才数据互通

研究背景

研究方法

研究过程

总结展望



CERIF数据模型中的实体及其关系

欧洲通用科研信息格式

**CERIF(Common European Research Information Format)**

是一种科研信息组织方法。

**CEIRF**模型中存在实体与实体之间的二元联系，以及同一实体间不同实例的一元联系，用关联实体表达实体之间的关系或实体的某种属性。



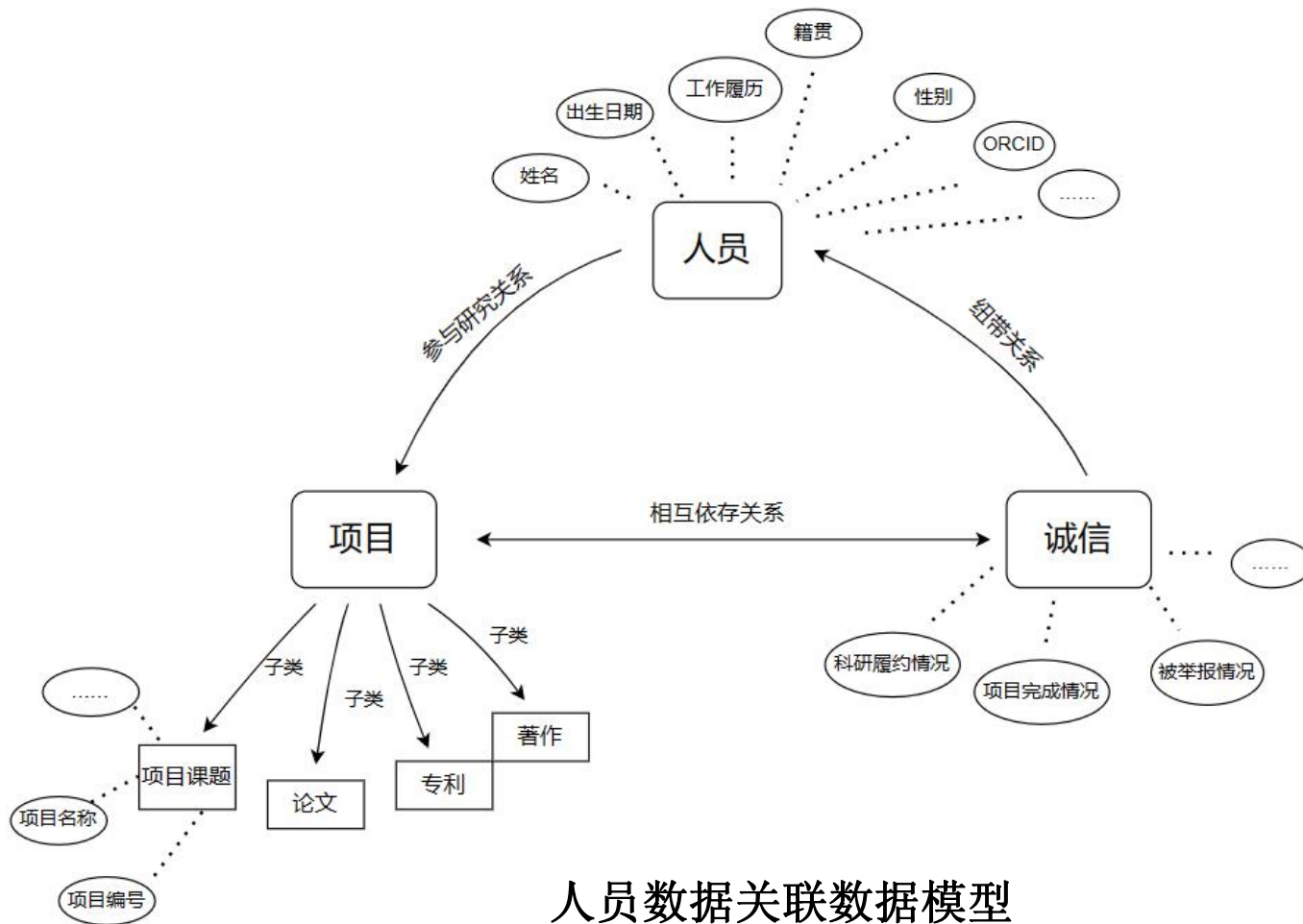
# 人才信息关联数据模型

研究背景

研究方法

研究过程

总结展望



人员数据关联数据模型

根据CERIF模型构建符合科研人员元数据规范的关联数据模型，在CREIF模型中，包括实体、属性及实体间的关系。



# 超网络模型：多维度发现人才专长、关系及影响力

研究背景

研究方法

研究过程

总结展望

建立科研人员多元合作关系，构建基于超图为底层结构的科研合作网络。用作者表示节点（vertex），用超边（edge）表示作者间的合作关系，作者通过共同撰写、发表文献建立其合作关系。



采用**加权指标**，利用权重有效提高相似性预测的准确度，即采用**加权超图共同邻居指标WHCN**和**加权超图资源分配指标WHRA**来度量超图节点的结构相似性。

$$\text{WHCN}(x, y) = \sum_{z \in N(x) \cap N(y)} w(x, z) + w(z, y)$$

$$\text{WHRA}(x, y) = \sum_{z \in N(x) \cap N(y)} \frac{w(x, z) + w(z, y)}{d(z)}$$





研究背景

研究方法

研究过程

总结展望

# 基于语义关联度的人才推荐

## Bert预训练模型语义相关性计算

应用**BERT**对文本进行向量化表示，从多个粒度层面实现文本语义信息的动态表征。

## 余弦相似度计算

基于两份文本的词向量计算文档间相似度的一种方法，通过计算两个向量的夹角余弦值来评估它们的相似度

## 科技人才推荐

数据收集及预处理    文本词向量计算    相似度计算    实验结果测评



# 基于Ceirf模型的科技人才学术简历自动生成

研究背景

## 场景一：风险敏感型

研究方法

研究过程

总结展望

作者合作数据

王**	1	黎*	闫*	祝*龙	王*恒	李*辉													
	2	祝*龙	程*	马*	王*恒	王*	霍*亮	李*辉	刘*森	霍*亮									
	3	祝*龙	程*	张*菲	吴*涵	简*帅	闫*	田*伏	刘*森	霍*亮									
	4	李*	闫*	李*瀚	程*														
	5	李*	闫*	霍*亮	李*瀚	赵*													
祝**	6	霍*亮	祝*龙	田*伏	李*														
	7	祝*龙	刘*森	吴*涵	李*辉														
	8	祝*龙	程*	张*菲	吴*涵	简*帅	闫*	田*伏	刘*森	霍*亮									
	9	霍*亮	祝*龙	田*伏	李*														
	10	吴*涵	祝*龙	白*	张*菲	简*帅	李*	杜*	李*辉	祝*龙									
	11	吴*涵	李*辉	曹*	姜*	简*帅	李*	白*	杜*	祝*龙									
	12	吴*涵	李*辉	曹*	姜*	简*帅	李*	白*	杜*	祝*龙									
李**	13	简*帅	张*菲	李*	任*娇	杜*	祝*龙	李*辉											
	14	李*辉	简*帅	霍*亮															
	15	李*	李*辉	霍*亮															
	16	李*辉	简*帅																
程*	17	李*	闫*	李*瀚	程*														
	18	张*苗	王*睿	蓝*	蓝*慈	刘*楠	冒*琪	任*翔	马*	李*辉	史*华	吕*	严*鹏						
	19	张*苗	王*睿	蓝*慈	蓝*慈	刘*楠	刘*雁	马*	李*辉	吕*	严*鹏								
	20	马*	康*然	李*行	吉*	吝*	常*曦	李*辉	李*	吕*	严*鹏								
	21	马*	李*	樊*	吝*	吝*	常*曦	李*辉	李*	吕*	严*鹏								
王*	22	曹*琳	陈*豪	王*	刘*														
	23	史*元	姬*	霍*亮	刘*璋	许*强	傅*												

科研合作超网络及潜在传播路径

超网络	Hyper net
文献数量	23
作者节点数量	49
二元合作关系数量	260
多元合作关系数量	23

采用超网络计算的优势在于将人纳入多个维度的复杂关系网络中，既可以过往历史事件进行风险范围和路径研判，又可以对信用等级进行量化计算预警，降低用人风险。



# 基于Ceirf模型的科技人才学术简历自动生成

研究背景

研究方法

研究过程

总结展望

场景二：企业需求侧驱动的高技能人才推荐



在一定程度上有助于提高岗位—人才匹配度,找到“小同行”, 实现精准引才。  
对于当前社会分工日益精细、实现人才“按需引进”具有启发性。



# 总结展望

研究背景

研究方法

研究过程

总结展望

- 采用**CERIF**模型，有助于**人才数据关联和共享**，从而大大缓解了**信息孤岛问题**
- 以超图的方式展现科研人员关系网络，通过识别与发现科技人才的学术合作关系、专长发现、风险预警等，对**人才的发现与遴选**具有一定的积极意义。
- 以招聘需求为驱动力，通过**BERT**模型对企业岗位需求信息以及科技人才信息进行语义计算，从而实现**岗位与人才的匹配与推荐**。
- 开源数据自身也存在噪音干扰、数据质量不一等问题，可以作为人才库建设的辅助手段，与现有专家库建设相互补充。

“有数”、“管数”、“用数”一体化设计，实现了科技人才画像与推荐方法的相互结合，有助于提高岗位人才推荐的质量与的匹配程度，为人才引进、人才评价、人才政策制定提供有益的参考。



# 感谢聆听 敬请批评指正

汇报人：李怡然