

# 基于Bert模型的创新知识图谱构建研究

金宏轩, 陆星家\*

宁波工程学院统计与数据科学学院, 浙江 宁波

收稿日期: 2025年6月7日; 录用日期: 2025年6月29日; 发布日期: 2025年7月8日

## 摘要

本文提出了一种基于BERT模型的创新知识图谱构建方法, 旨在从大规模文本数据中自动抽取高质量的实体及其关系。首先, 利用BERT强大的上下文语义表示能力, 提取论文、专利中的关键信息, 通过核心实体识别、关系抽取与链接, 构建创新知识图谱, 并开展创新指数分析。通过构建创新知识图谱, 实现对技术创新信息的结构化表达与语义关联, 方便后续的知识推理与应用。结果显示, 基于Bert模型的创新知识图谱能够有效地提取创新指数, 并分析产业上下游联系。

## 关键词

BERT模型, 知识图谱, 创新知识指数, 实体识别, 关系抽取

# Construction of an Innovation Knowledge Graph Based on the BERT Model

Hongxuan Jin, Xingjia Lu\*

College of Statistics and Data Science, Ningbo University of Technology, Ningbo Zhejiang

Received: Jun. 7<sup>th</sup>, 2025; accepted: Jun. 29<sup>th</sup>, 2025; published: Jul. 8<sup>th</sup>, 2025

## Abstract

This paper proposes a method for constructing an innovation knowledge graph based on the BERT model, aiming to automatically extract high-quality entities and their relationships from large-scale textual data. Firstly, leveraging BERT's powerful contextual semantic representation capabilities, key information is extracted from papers and patents through core entity recognition, relation extraction, and linking to build the innovation knowledge graph, followed by innovation index analysis. By constructing the innovation knowledge graph, the method achieves structured representation and semantic correlation of technological innovation information, facilitating subsequent knowledge

\*通讯作者。

reasoning and applications. The results demonstrate that the BERT-based innovation knowledge graph can effectively extract innovation indices and analyze upstream and downstream industrial connections.

## Keywords

BERT Model, Knowledge Graph, Innovation Knowledge Index, Entity Recognition, Relation Extraction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

21 世纪以来, 以知识、技术与人才等智力资本为核心的生产要素逐步取代自然资源与劳动生产, 随着生成式人工智能技术(AIGC)的出现, 世界政治格局和经济格局正经历着划时代的改变。对智力型人力资本、高科技产业布局的争夺掀起了国际上新一轮竞争, 推动了全球生产系统迭代与社会经济形态变革[1]。伴随人工智能技术为代表的新一代科技革命的高速发展, 对传统经济带来了翻天覆地的改变, 国内外相继出现了以高科技产业链为核心的城市群[2]。这些城市群不仅成为技术创新的孵化器, 引领全球产业格局的深刻变革。以硅谷为例, 它通过集聚大量高科技企业、科研机构 and 风险投资, 形成了完整的创新生态系统, 催生了无数全球领先的科技企业, 并推动了产业链的纵深发展与优化。北京、上海、深圳等中国城市也逐渐在高科技领域崭露头角, 构建起以人工智能、大数据、5G 技术等为核心的科技创新集群。这些城市的高科技产业链不仅促进了本地经济的增长, 也为全球科技创新提供了强大的动力。随着人工智能、大数据、云计算等技术的日益成熟, 产业链的上下游关系变得愈发紧密, 跨领域、跨区域的合作与融合成为推动经济高质量发展的重要引擎[3]。

知识图谱(Knowledge Graph)最早由谷歌(Google)在 2012 年提出[4], 知识图谱的引入标志着信息检索领域的一个重要转变, 通过将知识与数据连接起来, 谷歌能够在其搜索结果中提供更加智能化、结构化的信息。知识图谱不仅改善了用户的搜索体验, 还推动了人工智能技术在各行业的应用。知识图谱作为知识互联与应用的有效工具, 特别是在面对海量多元异构数据中有着广泛的应用空间。它通过构建实体之间的语义关联和层次结构, 能够帮助我们从海量数据中提取出有价值的信息, 提供深刻的洞察力。知识图谱的应用也延伸至多个领域, 包括自然语言处理(NLP)、推荐系统、语义搜索、机器学习等[5]。通过语义层面的深度理解, 知识图谱能够将不同领域的知识整合, 促进跨学科的知识发现和智能推理。

### 1.1. 全球创新指数

全球创新指数(Global Innovation Index, 简称 GII)由世界知识产权组织(WIPO)、康奈尔大学和欧洲工商管理学院联合发布, 旨在衡量各国创新生态系统的表现[6]。该指数自 2007 年起发布, 每年对全球多个经济体进行评估, 涵盖创新投入、创新产出、技术采用等多个维度, 为政策制定者、企业和学术界提供了重要的决策依据。全球创新指数从 PCT (专利合作条约)申请量和科学出版物(主要包括科学论文)两个维度衡量科技产出和创新质量, 指数越高, 代表创新成果越多, 也意味着这一地区创新力非常活跃。

GII 指标涵盖了七个主要领域: 制度环境, 包括政治稳定性、政府效能、法治等因素。人力资本与研

究, 衡量教育水平、研究人员数量、高等教育质量等。基础设施, 包括信息通信技术、能源供应、交通等基础设施建设情况。市场成熟度, 评估市场规模、企业竞争力、消费者支出等。商业成熟度, 衡量企业创新能力、研发支出、知识产权保护等。知识产出, 包括专利申请、科技论文发表、软件著作权等。创意产出, 评估品牌价值、设计创新、创意产品等。

GII 不仅关注高收入国家的创新表现, 也重视中低收入国家在创新领域的进展。近年来, 印度、越南、埃及等国在创新指数中取得了显著进步, 显示出全球创新格局的多元化趋势。GII 为全球创新活动提供了量化的评估工具, 促进了各国在创新领域的交流与合作。随着全球经济环境的变化, 创新已成为国家竞争力的重要体现, 全球创新指数的持续发布为各国在创新驱动发展战略的制定与实施中提供了宝贵的参考。

## 1.2. 知识图谱与 GII

知识图谱能够与 GII 结合, 更加有效地表现创新指数。通过构建包含专利、科研成果、技术路线、市场趋势等多维数据的知识图谱, 可以为政策制定者、企业和研究机构提供精准的创新态势分析, 从而促进创新资源的合理配置和优化。通过分析专利数据和科研成果, 构建技术演化图谱, 识别技术发展的趋势和潜在的技术突破点, 为企业的研发方向提供参考。通过整合产业链上下游企业、科研机构和政府部门的数据, 构建创新生态系统图谱, 识别关键节点和合作关系, 促进资源的协同创新。通过对政策实施前后的创新指数变化进行分析, 评估政策的效果, 为后续政策的制定提供依据。构建不同地区的创新能力图谱, 比较各地区在技术研发、市场应用、人才培养等方面的优势和不足, 为区域发展战略的制定提供数据支持。随着人工智能、大数据等技术的发展, 知识图谱将在创新指数的表现中发挥越来越重要的作用。未来, 知识图谱将与全球创新指数深度融合, 形成动态更新的创新态势分析平台, 为全球范围内的创新活动提供实时、精准的数据支持, 促进全球创新资源的共享与协同发展。

## 2. 创新测度体系的构建

### 2.1. 创新指数综合评价

GII 是衡量产业创新的重要指数, 受到全世界主要发达经济体的认可。中国的创新指数(GII)排名从 2020 的 14 位, 逐步递增到 2024 年的 11 位, 近五年稳步上升, 美国的创新指数(GII)则由第 3 位滑落到第 5 位(图 1)。

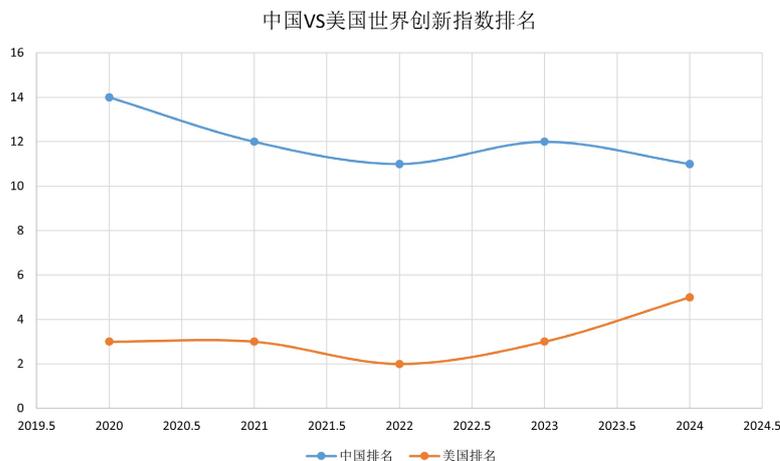


Figure 1. China vs. the United States-global innovation index ranking  
图 1. 中国 vs. 美国 - 世界创新指数排名

GII 采用多层指标体系, 其中一级指标包括创新投入指数、创新产出指数; 创新产出指数则反映创新活动产生的成果, 包括知识与技术产出及创意产出。在一级创新投入指数下, 则包含制度环境、人力资本与研究、基础设施、市场成熟度、商业成熟度 5 个二级指标; 在一级创新产出指数包括知识与技术产出、创意产出 2 个二级指标。全球创新指数体系覆盖了从创新体系的制度基础建设到具体创新成果展示的全链条, 是全面评估各国创新能力和表现的重要参考指标。除了全球创新指数(Global Innovation Index)之外, 还有一些其他的创新指数, 美国创新指数(U.S. Innovation Index)、波士顿咨询集团(BCG)发布的全球创新领先指数等。这些创新指数从不同角度评估国家或地区的创新能力与表现, 补充和丰富了全球创新评估体系。美国创新指数主要关注美国各州及主要城市的创新环境, 评估因素包括研发投入、专利产出、风险投资、创新人才分布等, 帮助政策制定者和企业理解区域创新优势及不足, 促进地方创新生态系统的发展。欧洲创新指标(European Innovation Scoreboard)、世界经济论坛发布的全球竞争力报告中的创新能力评估, 也在国际创新评价中占据重要地位。这些指数各有侧重, 有的强调技术创新, 有的关注社会创新和制度创新, 形成了多维度、多层次的创新评价体系。

在充分借鉴国内外成熟创新测度体系, 本文结合国内一线、二线城市的发展特征, 构建了一套科学、可量化的创新指数体系。该体系包含五个一级指标: 创新资源与要素、创新环境与支持、创新活动与过程、创新产出、创新绩效与影响和 14 个二级指标[7], 如图 2 所示[7]。

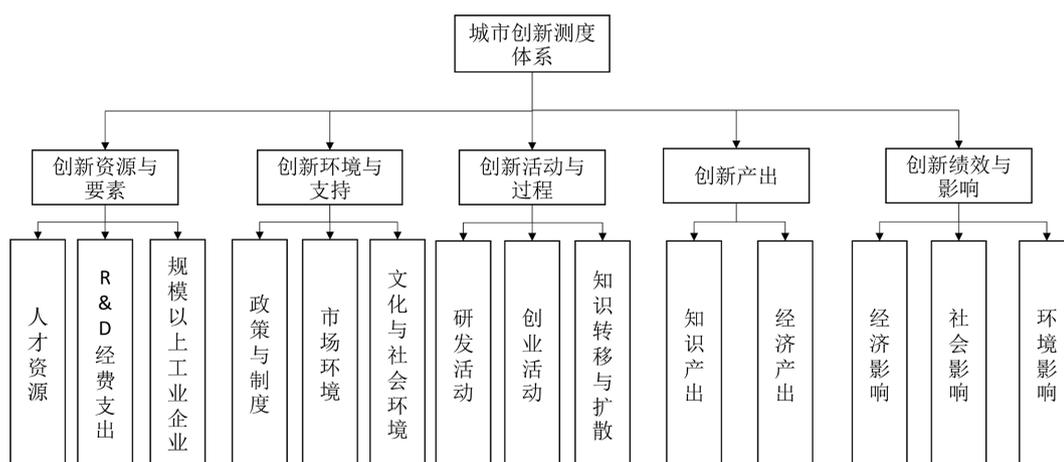


Figure 2. Innovation measurement indicator system

图 2. 创新测度指标体系

在创新资源与要素一级指标下, 是人才资源, R&D 经费支出, 规模以上工业企业 3 个二级指标。创新环境与支持一级指标下是政策与制度, 市场环境, 文化与社会环境 3 个二级指标。在创新活动与过程一级指标下, 是研发活动, 创业活动, 知识转移与扩散 3 个二级指标。在创新产出一级指标下, 是知识产出, 经济产出 2 个二级指标。在创新绩效与影响一级指标下是经济影响, 社会影响, 环境影响 3 个二级指标。

## 2.2. Bert 模型

BERT (Bidirectional Encoder Representations from Transformers)模型是一种给予 Transformer 的双向编码器表示的语言表示模型, Bert 通过联合考虑前序和后序上下文, 在所有层次上对无标签文本进行深度双向表示的预训练, Bert 模型只需要添加一个额外的输出层就可以进行微调, 无需进行大量的任务特定架构的修改[8]。在构建创新知识图谱的过程中, BERT 具备深层次理解自然语言语义的能力, 能够显著

提升从非结构化文本中抽取实体、关系及事件的准确率。通过微调 BERT 模型, 可以有效地完成命名实体识别(NER)、关系抽取(RE)和事件抽取等核心任务, 进而构建高质量、动态更新的知识图谱。在创新领域, 相关信息往往分散在科研论文、专利文献、技术报告和新闻资讯等多源异构数据中, 这些文本语义复杂、表达多样。传统的规则或浅层学习方法在面对术语新颖、上下文依赖性强的文本时往往力不从心。相比之下, BERT 通过双向编码器能够充分捕捉词语间的上下文依赖, 精准识别出如“技术路线”、“核心专利”、“研究热点”等创新要素。在实际构建流程中, BERT 模型可作为底层语言理解模块, 配合知识融合、实体消歧和图谱存储等组件, 逐步形成以技术、机构、人物和成果为节点的多维知识图谱, 为科技情报分析、技术趋势预测和创新决策提供数据支撑。借助 BERT, 创新知识图谱不仅实现了从“结构化构建”到“语义驱动构建”的转变, 也显著提升了对新兴技术和交叉学科知识的感知能力, 加速了科技信息的组织、整合与应用。

Bert 模型分为几个部分:

#### 1) 输入表示(input representation)

Bert 接受的输入是由多个词汇构成的句子构成的序列, 每个词汇通过词嵌入(word Embedding)转换为向量表示,

$$X = E + S + P \quad (1)$$

在公式(1)中,  $E$  表示词嵌入, 即每个词的固定词向量,  $S$  表示段落嵌入, 区分不同句子,  $P$  表示位置嵌入, 编码词汇在句中的位置。

#### 2) 自注意力机制(Self-Attention)

在 Transformer 机制中, 自注意力机制用于计算序列中每个词对其他所有词的注意力权重。

$$A(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (2)$$

在公式(2)中,  $Q$  表示查询向量(Query),  $K$  表示键向量(Key),  $V$  表示值向量,  $d_k$  表示键的维度。

#### 3) 遮掩语言模型(MLM)

Bert 在预训练阶段, 通过(Masked Language Model) MLM 来进行训练, 随机遮掩部分词汇并让模型预测这些被遮掩的词汇。

$$P(w_i | w_1, \dots, w_i, w_N) = \text{softmax}(W \cdot h_i) \quad (3)$$

在公式(3)中,  $w_i$  是目标词汇,  $h_i$  是经过自注意力计算得到的上下文向量,  $W$  是词汇表的权重矩阵。

### 2.3. 知识图谱构建过程

知识图谱构建过程如图 3 所示, 前期包括问题域界定, 并进行数据收集与整合, 对敏感信息进行脱敏, 利用特征工程对数据进行整合[8]。基于 Bert 模型的知识图谱构建对整合后的数据进行核心实体识别, 并进行关系抽取与链接, 完成知识图谱的构建。Bert 模型作为一种预训练的深度双向语言表示模型, 因其强大的语义理解能力, 在自然语言处理任务中表现出色, 适合应用于知识图谱构建的核心环节。在知识图谱的实体识别阶段, Bert 模型可以通过微调预训练模型, 识别文本中的命名实体, 实现准确的实体抽取。相较于传统的基于规则或统计的方法, 基于 Bert 模型的实体识别模型能够更好地捕捉上下文信息, 提高识别的精度和召回率。

在关系抽取和链接环节, Bert 模型能够用于识别实体之间的语义关系。将包含两个实体的句子输入 Bert 模型, 通过添加特殊标记突出实体位置, 借助 Bert 模型的深度编码能力, 提取代表实体对关系的特征向量, 继而通过分类器判定实体间的关系类型。Bert 模型能够有效捕捉复杂的语言表达和长距离依赖。

在实体链接阶段, Bert 模型生成的高质量语义向量, 将识别的实体与知识库中的标准实体进行匹配, 解决同名异物或多义词的问题, 确保知识图谱中实体的唯一性和准确性。根据构建的知识图谱完成创新指数的分析, 并利用可视化技术对知识图谱进行展示。

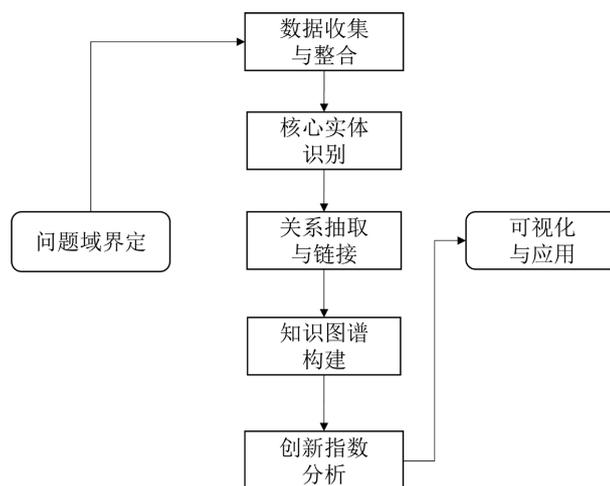


Figure 3. Knowledge graph construction process

图 3. 知识图谱构建过程

### 3. 创新指数知识图谱构建

创新指数构建需覆盖研发投入、技术产出、人才培养等多个环节。环节的有效整合促进技术创新路径的清晰描绘, 路径分析揭示产业升级的关键节点与趋势演变。趋势演变反映市场需求、政策支持和科技进步的共振效应, 效应催生新的产业生态和创新模式。模式创新带动跨界融合, 融合推动知识图谱构建, 实现创新资源的系统整合与智能分析。分析结果指引决策优化, 优化策略提升产业核心竞争力, 竞争力强则推动经济高质量发展, 发展持续依赖于创新驱动和协同合作。

创新指数分析的知识图谱需要收集产业相关的多源数据, 包括与企业相关的专利数据、科研文献, 分析企业的创新能力。收集行业的市场调研报告, 了解行业的发展动态。使用文本挖掘和自然语言处理技术, 对以上数据进行清洗和整理, 去除噪声信息, 确保数据的准确性和完整性。

#### 3.1. 核心实体识别

在创新知识图谱的构建中, 语义网三元组(Subject-Predicate-Object)作为一种结构化的知识表示方式, 能够有效地表达语义数据中的多维和复杂关系。通过构建三元组, 可以将“企业-生产-产品”、“产品-基于-技术”、“技术-包含-专利”等相关语义进行形式化表达, 形成一个多层次、多维度的知识网络, 从而支持对创新指数知识图谱的构建。为了保障知识图谱的语义一致性与推理有效性, 构建一套完备的创新指数分析体系至关重要。本体作为知识图谱的“语义支柱”, 定义了概念类(Classes)、属性(Properties)以及类之间的层级关系和约束规则。在创新指数本体中, 可以定义如下内容: 类层次结构(Class Hierarchy), 属性约束(Property Constraints), 关系约束(Object Properties), 数据属性(Data Properties)。在创新指数知识图谱的构建过程中, 语义网三元组的生成依赖多源异构数据的融合处理, 借助自然语言处理(NLP)技术, 可以从文本中抽取实体与关系, 再经由本体映射统一语义标签, 最终形成结构化的创新指数知识图谱。

### 3.2. 关系抽取与链接

在知识图谱构建中, 关系抽取和链接是将实体之间的语义联系结构化、规范化的核心环节。关系抽取是识别和提取实体间相关语义关系的过程, 常见的关系抽取主要包括基于规则的抽取和基于监督学习的抽取, 其中基于规则的抽取是利用专家规则或模板, 从结构化文本或半结构化数据中识别实体和关系。Bert 模型关系抽取是基于监督学习的知识抽取, 可以使用标注好的语料库自动抽取关系, Bert 模型通过监督学习, 捕捉“创新”与“企业”等不同实体间的关系类型。当缺乏标注数据时, 也可以采用无监督学习的方式进行关系抽取。使用 BERT 模型进行文本数据的关系抽取和关系链接, 首先需要在文本中识别出实体, 并通过特殊标记明确实体边界, 将标注后的文本输入 BERT 模型以获得上下文的深层次语义表示。然后, 从 BERT 的输出中提取实体对应位置的向量表示, 通常将两个实体向量拼接后输入一个全连接层进行关系分类, 从而判断实体对之间的具体关系。对于关系链接, 通常将抽取得到的关系文本映射到知识库中的标准关系, 方法包括对关系文本和知识库中关系进行编码, 再通过向量相似度匹配实现标准化, 也可以采用端到端的微调模型直接输出知识库关系 ID。整个流程可以通过预训练的 BERT 微调完成, 有效整合实体上下文信息, 实现精确的关系抽取与链接。Bert 模型处理复杂的语义关系, 通过双向上下文建模、自注意力机制、预训练和微调机制, Bert 模型通过双向建模, 同时考虑前后文关系。

## 4. 实验与结果分析

### 4.1. 数据的爬取

选取中国知网(CNKI)、WOS (Web of Science)的专利数据、论文数据, 选择“宁波”作为城市区域关键词, 下载基础数据(专利和论文)。在数据预处理阶段, 对专利和论文文本进行清洗与标准化处理, 确保实体信息和关系信息的准确性。随后, 利用 BERT 深度编码文本, 抽取关键实体如发明人、机构、技术领域等, 并通过关系抽取模型识别它们之间的关联。接着, 整合多源异构数据, 实现实体对齐与关系链接, 完善知识图谱节点和边的丰富性。基于构建完成的知识图谱, 进一步计算创新指数, 通过分析技术演进路径、合作网络和专利影响力, 辅助评估创新指数和趋势。

通过对大量文本数据进行训练, BERT 模型能够准确识别出其中的关键实体, 如技术术语、发明人、专利、技术分类、行业趋势等。此外, BERT 模型还能够识别文本中的关系, 例如技术领域之间的关联、企业之间的合作关系、技术创新与市场应用之间的联系等。通过对这些实体和关系的抽取, BERT 模型通过实体链接关系创建新的知识图谱, 呈现出创新活动中各要素之间的复杂互动关系。进一步地构建基于知识图谱的创新指数。

### 4.2. 产业分布知识图谱分析

在 BERT 模型的实验中, 选择宁波市作为样本, 收集宁波市各产业行业企业的各类经济文本数据。对原始文本数据进行预处理, 包括去除噪声数据、剔除无关符号及 HTML 标签等干扰信息。随后, 将文本拆分为具有实际意义的词汇单位, 针对中文采用分词处理, 针对英文则使用 tokenization 技术。接着, 去除停用词, 如“的”、“是”、“在”等无实质意义的词汇, 以提升文本质量。最后, 进行文本规范化操作, 包括统一大小写格式和词形还原。利用 Python 语言、可视化 Geopandas 包, 对宁波市产业分布数据进行可视化展示(图 4)。通过可视化的宁波产业分布地图可知, 宁波市形成了包括高端装备制造、新材料、生物医药、新一代信息技术、节能环保和新能源汽车等领域的支柱产业, 以上产业依托宁波的制造基础和港口优势, 形成了较为完整的产业链和集聚效应。

宁波市的核心区和重点开发区依托园区平台实现企业和资源的集聚。当地重点发展临港产业和出口

导向型新兴产业, 推动产业向海洋经济和国际贸易深度融合。除主城区外, 宁波的部分县区也构建了特色产业基地, 形成多个产业发展中心, 促进区域协调发展。依托高校和科研院所, 形成创新链条, 产业园区内设有研发中心和创新平台, 推动技术进步和产业升级。

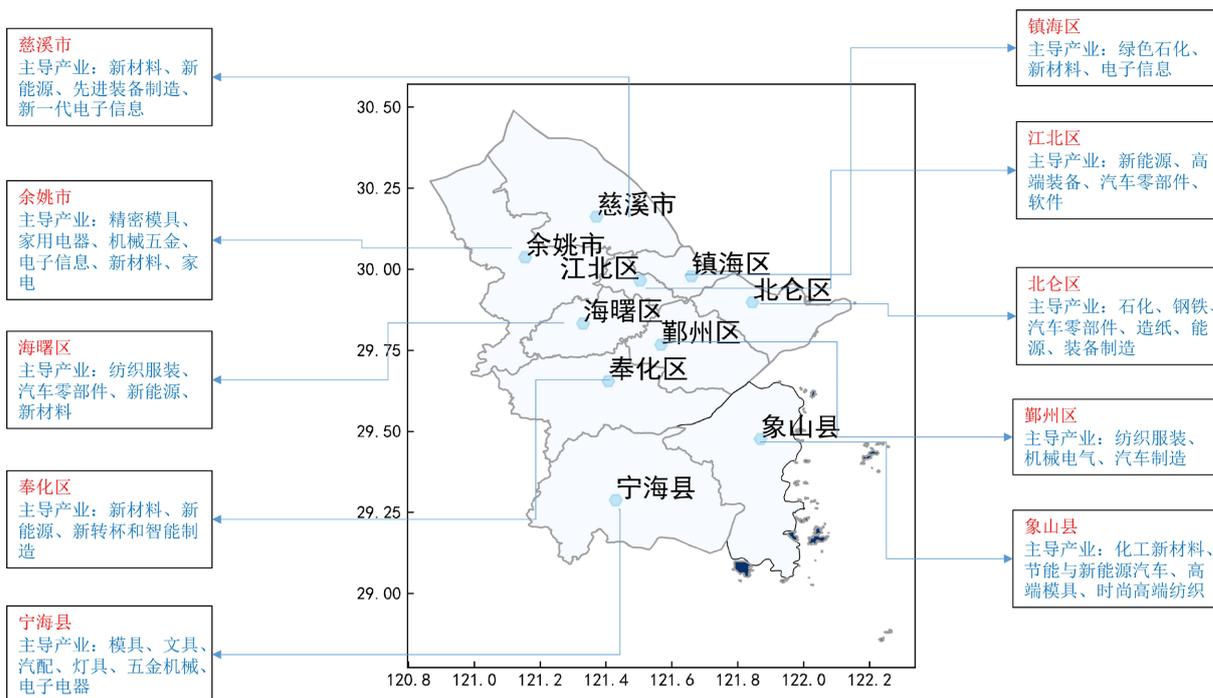


Figure 4. Knowledge graph of industrial distribution in Ningbo city

图 4. 宁波市产业分布知识图谱

### 4.3. 企业创新指数分析

通过 CNKI 的专利数据库, 下载金蝶软件公司 2000 年至今的 2650 条专利数据, 专利数据中包含专利名称、关键技术、上下游行业等重要信息。基于 Bert 模型构建的知识图谱, 知识图谱的链接关系中包括节点的源(source)、节点的目标(target)以及链接权重(weight)数据, 知识图谱分析结果如下(图 5), “装置”是最中心的节点, 与大量其他技术实体、方法和设备相关联, 显示其在专利技术体系中的关键地位。“装置”与“设备”、“计算机设备和存储介质”、“可读存储介质及程序产品”等多个节点的权重较高(如与“计算机设备和存储介质”权重 86, “设备”权重 10~23 等), 表明技术实现通常依赖硬件与存储结合, 体现软硬结合的特征。“装置”与“设备”之间的权重最高, 说明在技术实现中这两个概念密切相关, 也是创新重点方向。“装置”与“可读存储介质和程序产品”、“计算机设备和可读存储介质”等节点权重相对较高, 说明专利文本中不仅涵盖硬件还包括软件和存储介质相关技术, 技术覆盖硬软件结合。

### 4.4. CRF 与 Bert 的对比分析

条件随机场(Conditional Random Field-CRF)模型是一种传统的序列标注模型, 是一种基于概率图模型的序列标注方法, 广泛应用于自然语言处理(NLP)领域, 尤其是在命名实体识别(NER)、词性标注(POS tagging)、关系抽取(RE)等任务中。CRF 通过考虑上下文信息和相邻标签之间的依赖关系来对序列数据进行建模, 从而使得它在序列标注任务中非常有效。分别将专利数据测试数据导入到 CRF 模型和 Bert 模型, 训练和测试数据相同。CRF 模型的结果如表 1 所示, CRF 模型文本分类的准确率为 62%。

**Table 1.** CRF model classification report  
**表 1.** CRF 模型分类报告

类别	精确度	召回率	F1-得分	支持度
B-PATENT	1.00	1.00	1.00	2
I-PATENT	1.00	1.00	1.00	2
B-INVENTOR	0.00	0.00	0.00	1
I-INVENTOR	0.00	0.00	0.00	2
O	0.40	1.00	0.57	2
整体准确率			0.62	8
宏观平均	0.48	0.60	0.51	8
加权平均	0.47	0.62	0.52	8

**Table 2.** Bert model classification report  
**表 2.** Bert 模型分类报告

类别	精确度	召回率	F1-得分	支持度
B-PATENT	1.00	1.00	1.00	2
I-PATENT	1.00	1.00	1.00	2
B-INVENTOR	0.00	0.00	0.00	1
I-INVENTOR	0.00	0.00	0.00	2
O	1.00	1.00	1.00	2
整体准确率			1.00	8
宏观平均	1.00	1.00	1.00	8
加权平均	1.00	1.00	1.00	8

Bert 模型(表 2)相比 CRF 模型, 在分类的精确度、召回率和 F1 得分上都接近 1, 说明 Bert 模型具有很好的泛化能力, 能够准确地识别专利文本中的各种实体。“装置”相关的方法节点种类繁多, 如“报表处理方法”、“单据查询方法”、“数据查询方法”、“权限管理方法”、“模型训练方法”等, 这表明专利技术应用领域广泛, 涵盖报表、权限、数据处理、模型训练等多个子领域。许多方法权重较小(大多为 1 或 2), 体现技术细节多样且分散。“提示语替换方法”与“装置”、“设备”、“存储介质和程序产品”的权重为 1, 表明存在基于提示词或自然语言技术的创新, 可能涉及智能交互或人机界面技术。

节点间大量指向“存储介质”、“计算机设备”, 以及“系统”、“平台”等, 反映整个创新体系涵盖数据存储、计算处理和软件平台的紧密结合。技术生态系统特征显著: 数据显示技术从基础硬件(设备、存储介质)到应用方法(报表生成、权限管理、数据加密等)组成一个完整的技术生态, 说明该技术领域注重软硬件协同创新。关系较多也指示着专利之间存在丰富的技术关联与集成。

图 6 是金蝶软件的上下游分析, 在产业链上下游关联中, “提示语替换方法”与“装置”、“设备”、“存储介质和程序产品”的权重为 1, 表明存在基于提示词或自然语言技术的创新, 可能涉及智能交互或

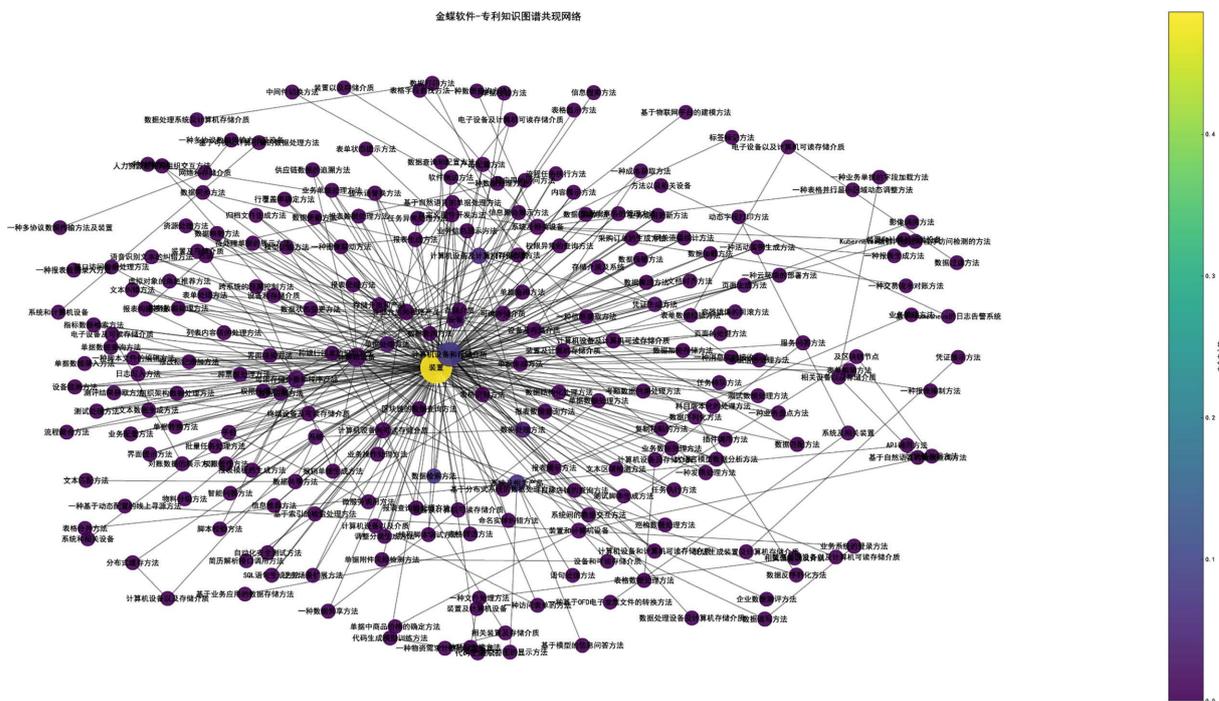


Figure 5. “Kingdee software” patent knowledge graph  
图 5. “金蝶软件”专利知识图谱

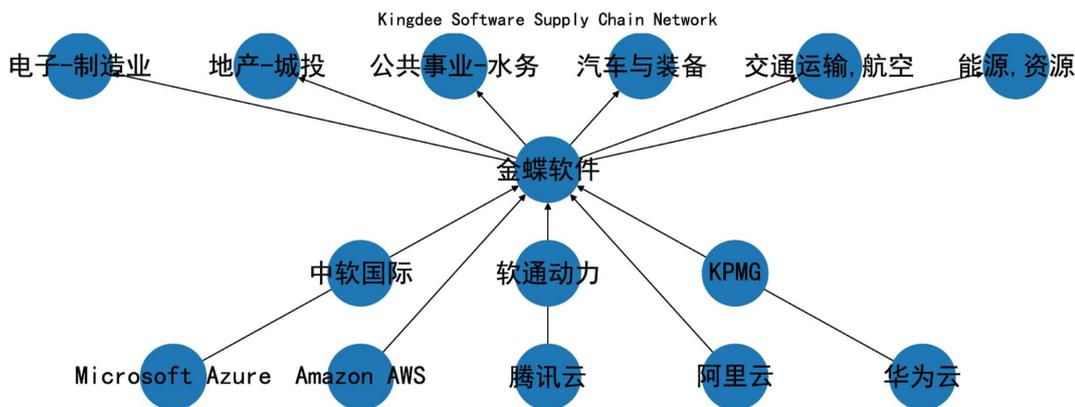


Figure 6. Kingdee software’s upstream and downstream supply chain analysis  
图 6. 金蝶软件上下产业链分析

人机界面技术。节点间大量指向“存储介质”、“计算机设备”，以及“系统”、“平台”等，反映整个创新体系涵盖数据存储、计算处理和软件平台的紧密结合。技术生态系统特征显著，数据显示技术从基础硬件(设备、存储介质)到应用方法(报表生成、权限管理、数据加密等)组成一个完整的技术生态，说明该技术领域注重软硬件协同创新。

### 参考文献

- [1] Malik, A.S., Boyko, O., Aktar, N. and Young, W.F. (2001) A Comparative Study of MR Imaging Profile of Titanium Pedicle Screws. *Acta Radiologica*, **42**, 291-293. <https://doi.org/10.1080/028418501127346846>
- [2] Nickel, M., Murphy, K., Tresp, V. and Gabrilovich, E. (2016) A Review of Relational Machine Learning for Knowledge

- Graphs. *Proceedings of the IEEE*, **104**, 11-33. <https://doi.org/10.1109/jproc.2015.2483592>
- [3] Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J. and Yakhnenko, O. (2013) TransE: Translating Embeddings for Modeling Multi-Relational Data. *Proceedings of the 26th International Conference on Neural Information Processing Systems*, Lake Tahoe, 5-10 December 2013, 2787-2795.
- [4] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., *et al.* (2014) Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 24-27 August 2014, 601-610. <https://doi.org/10.1145/2623330.2623623>
- [5] Yao, L., Mao, C. and Luo, Y. (2019) KG-BERT: BERT for Knowledge Graph Completion. *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, Macao, 10-16 August 2019, 4218-4224.
- [6] 吴江, 任馨怡, 冯定国. 中美两国创新能力的对比与趋势分析——基于 2015-2024 年《全球创新指数》的考察[J]. 西南交通大学学报(社会科学版), 2025, 26(2): 1-20.
- [7] 邱子轩, 王艺琳, 陆星家. 基于 TOPSIS 模型的人才吸引力指数研究[J]. 宁波工程学院学报, 2025, 37(2): 50-57.
- [8] 闫效莺, 靳艳春, 冯月华, 张绍武. 基于 Bert + GCN 多模态数据融合的药物分子属性预测[J]. 生物化学与生物物理进展, 2025, 52(3): 783-794.