**亿级科技知识图谱引擎要求**

亿级科技知识图谱引擎依托数据融合中枢，整合元数据与精加工数据，构建结构化知识单元，并通过抽取、关联与学习技术构建超大规模科技情报图谱。结合动态本体与图谱推理算法，提供内容理解、语义检索、智能推荐、知识问答和可视化分析，支撑科创知识的积累、传播与复用，并向智能系统输出高效精准的推理结果。

**知识图谱引擎**需提供的功能包括知识建模、图谱构建、知识融合、知识计算、图谱应用、多模态知识图谱、面向多模态知识图谱与问答能力建设。基于知识图谱引擎，需提供知识图谱构建服务。具体技术要求如下：

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **功能模块** | **功能项** | **招标技术要求** |
| 1 | 知识建模 | 概念抽取 | 需提供的功能包括科技领域种子术语生成、候选术语生成、术语向量生成、科技领域本体、基于置信度图传播的术语排序。 |
| 上下位关系生成 | 需提供的功能包括依存关系图构建、子图生成、关系图注意力网络、上下位关系生成、上下位关系管理、上下位关系智能判定、上下位关系预测。 |
| 属性识别 | 需提供的功能包括属性管理、文本语义抽取。需支持基于规则映射和基于统计学习方法的抽取策略，获取概念的候选属性。 |
| 规则建模 | 需利用Rete算法、Leaps算法开展规则建模。需提供的功能包括规则定义与管理、概念与属性规则建模、约束规则建模与校验、条件驱动结果推理、规则执行与推理引擎集成、规则可视化与结果反馈、规则调试与测试工具。 |
| 事件建模 | 需支持通过依存句法分析抽取领域术语与事件，实现基于术语/事件粗抽取、人工参与的术语/事件优化、领域知识概念层次的构建。 |
| 时空建模 | 需支持从结构化、半结构化和非结构化时空数据中进行实体、关系和属性的信息提取，通过实体对齐和指代消解实现对时空数据的知识融合并将知识存储到知识库中。需提供的功能包括：时序信息嵌入、实体动态、时序关系依赖、时序逻辑推理。 |
| 文献知识建模 | 通过自然语言处理与图谱分析技术，构建从知识单元到研究范式的结构化映射，支持研究者快速定位核心成果与研究脉络。  1.多维度解析引擎：系统自动提取文献中的研究对象、方法、结论等核心要素，生成结构化知识卡片。支持对实验步骤、数据图表、数理模型的智能解析，构建可追溯的知识链路。  2.知识脉络可视化工具：提供研究领域的时空知识图谱，直观展示技术演进路径、学派关联与学术交叉点。支持通过基因编辑、机器学习等主题词动态追踪领域突破点。  3.协作验证平台：用户可对文献解析结果进行标注、纠错或补充数据，系统通过跨用户共识算法实现知识可信度分层，形成持续优化的动态知识库。 |
| 专利知识建模 | 专利知识建模系统聚焦技术方案的权利要求解析与产业链关联，打通从科研突破到商业应用的转化路径。  1.技术要素解构框架：深度解析专利权利要求书中的技术特征、创新点与应用场景，自动生成技术功效矩阵。支持对化学分子式等专业内容的语义化标注。  2.跨域知识融合：将专利技术方案拆解为标准化的功能模块，支持与学术文献中的理论模型、实验数据进行自动匹配，揭示“基础研究-技术开发”的转化瓶颈。 |
| 3 | 知识融合 | Schema匹配 | 需提供的功能包括跨本体实体模糊关系与逻辑关系构建、本体匹配关系智能构建、本体匹配主体、本体匹配流程组件。 |
| 实例匹配 | 需提供的功能包括文本实例匹配、结构实例匹配、机器学习实例匹配、多策略融合匹配。 |
| 知识链接 | 需建立基于词和实体联合表示的无监督概率模型、基于多任务学习的实体发现与链接联合优化模型、基于词和实体联合表示的实体链接模型。需提供的功能包括文本实体识别、候选实体生成、实体链接判断、链接标注与映射生成、文档实体高亮与交互展示。 |
| 跨语言知识融合 | 需提供的功能包括本体实例匹配、查询驱动的实例匹配、跨语言属性对齐、跨语言知识库生成与对齐。 |
| 数据语义链接 | 需提供的功能包括语义关联与消歧、实体匹配消歧、图谱融合、知识存储。 |
| 4 | 知识计算 | 知识表示学习 | 需支持表示空间、打分函数、编码模型和辅助信息，服务于知识获取和下游应用。 |
| 知识相关性计算 | 需提供的功能包括基于监督学习的相似度计算、快速图节点相似度计算、关联强度量化、语义检索需支持、推荐应用需支持。 |
| 知识推理模型 | 需开展信息提取、信息推理，需提供的功能包括渐进式子图扩展、关系推理。 |
| 知识补全 | 需提供的功能包括大模型生成能力接口、分层置信标注集管理、强化学习框架集成、少样本学习三元组生成、多源人工验证数据反馈机制、知识图谱迭代修正。 |
| 5 | 图谱应用 | 语义搜索 | 需构建两种信息的语义检索方式：基于概念的语义检索和基于规则的语义检索。构建语义检索效度。需提供的功能包括知识搜索、语义搜索。支持以自然语言方式作为检索输入，检索的文献库包含1.5亿篇以上文献，其中检索得到的文献数量不少于40篇，对于检索到的文献列表，支持通过时间、期刊、影响因子等方式进行筛选，大语言模型自动对检索得到的文献进行概括总结，概括总结内容可直接索引到文献，对于检索到的文献列表，可以多选进行深入对话，可以批量保存到知识库，文献检索、问答要求提供不少于3种大模型。 |
| 演化分析 | 需提供的功能包括动态主题演化图、学科交叉分析、发展推演沙盒、竞争力雷达图、科研成果关联分析、科技发展情景模拟、跨领域影响分析。 |
| 关系分析 | 需提供的功能包括关系评分模型构建、关系度量与排序、关系分析可视化、关系分析结果导出。 |
| 推理预测 | 需提供的功能包括关联规则挖掘、规则扩展、规则筛选与应用。 |
| 决策支持 | 构建基于规则的决策支持和基于关联分析的决策模型。需提供的功能包括领域专家规则策略、聚类实体群决策、决策模型整合、决策结果评估与反馈。 |
| 图谱可视化 | 需提供的功能包括图谱关系可视化、图谱统计数据可视化、关联图分析、统计图联动关联图分析、地图联动关联图分析、系列图谱可视化分析辅助功能、需支持基于快照形成快速报告。 |
| 知识校验 | 需提供的功能包括检索设置、添加关键词关系、删除关键词、删除关键词间的关系、知识模型结构智能化校验、支持图谱统计数据可视化、支持时序可视化。 |
| 知识库构建 | 1.元数据自动标注：系统支持对上传文献进行自动化元数据标注，包括作者、出版时间、期刊、DOI等信息，提升文献管理的精确度和便捷性。  2.树状文件夹管理：支持树状结构的文件夹管理，使得文献的分类和检索更加有序、精准，用户可以轻松管理大量文献。  3.自动标签与分类：系统能够自动为上传的文献添加标签，并按分类规则进行智能归档，极大提高文献整理效率，使分类操作更加便捷。  4. 提供知识库发布广场，承载1w以上用户自行发布高质量知识库。 |
| 文献要点总结 | 1.实现高效阅读，深度理解：集成AI问答、智能翻译、笔记标注等功能。侧边栏提供实时互动支持，帮助科研人员高效阅读、理解和管理学术内容，提升科研效率。  2.支持文档内分子活性信息抽取：从复杂文献中精准识别并匹配分子结构及其活性数值，生成结构化数据表格。显著提升数据收集效率与准确性，为分子优化和药物设计提供可靠数据支持，加速科研进程。  3.支持单篇或多篇文献的AI问答功能，用户可以通过提问，获得文献中的关键信息和答案，提升文献的解读效率。  4.支持对文献知识库通过自然语言进行检索和智能问答；支持缩略图、大纲、书签、图层视图；页面支持顺时针和逆时针旋转、支持单页和双页查看；支持下划线、曲线、手绘、自由文本、插入文本、替换文本等注释方式；支持矩形、椭圆、多边形、云等标记形状；支持插入文件、盖章；支持添加备注。 |
| 学术海报与音频概览生成 | 1.根据精读内容，自动生成学术海报和音频概览，便于快速展示和分享研究成果；  2.一键生成笔记：通过智能算法提取文献中的关键信息，自动生成笔记，方便科研人员快速整理和总结文献要点。 |
| 6 | 多模态知识图谱 | 多模态知识图谱相关数据集 | 需提供的功能包括数据集构建、数据集分类、数据集评估、知识点详情查看、知识点搜索、图谱下载、知识点属性自定义。 |
| 数据的统一建模与表示融合 | 需提供的功能包括多模态完备表示、异质多模态特征融合、多模态数据集管理、多模态特征融合、知识点关联。 |
| 语义理解模型 | 需支持基于对抗迁移学习的知识图谱补全和基于元学习的少样本知识图谱推理。 |
| 复杂多样的知识定位 | 需提供的功能包括实体定位、概念定位、关系定位。 |
| 垂直领域多模态知识图谱 | 通过关联企业、行业、人物和事件信息，实现企业经营的全方位考察。需提供的功能包括行业发展风控、  垂直领域多模态知识图谱构建、多模态知识图谱推理、  知识门户工具、主题科技知识图谱分析应用。 |
| 7 | 面向多模态知识图谱与问答能力建设 | 面向科研三大资源的数据解析与统一结构表达能力 | 针对学术文献、技术专利与科研数据库三类核心知识资源，构建专门的数据解析子系统，实现各类模态的结构化提取与标准化转换。  1.文献处理：支持 PDF、JATS XML、HTML 格式的正文、图、表、公式解析；开发段落级摘要生成、图表关系还原、公式结构识别与图注关联机制；输出内容包括：概念定义、变量关系、实验过程、引用路径等。  2.专利处理：支持结构图、流程图与权利要求内容解析；实现权利部件 - 功能 - 用途三元组抽取；建立专利图谱中 “技术方案 — 结构实现 — 应用场景” 路径模型。  3.数据库处理：对接已有结构化科研数据库（如材料性能库、反应数据库、药物分子数据库）；实现字段语义映射与实体标准化；将数据条目映射为图谱节点、边与属性值，纳入统一知识空间。 |
| 多模态融合知识图谱构建 | 建设统一本体指导下的多模态科研知识图谱构建平台，打通文本、图像、表格、公式、数据字段等不同信息形态，实现语义融合与图结构统一。构建科研通用本体体系（覆盖实体类别、关系类型、属性层级等），重点强化文献概念、专利技术实体与数据库数据实体的类别对齐；支持 “实体 - 关系 - 实体 + 属性” 四元结构的自动提取与图谱建模，开发基于跨资源共现分析的关系预测模型；实现跨模态节点融合与消歧，通过文献引证关系、专利同族关联、数据库字段关联实现三类资源实体的精准匹配，支持节点版本演化与图谱动态更新；建立图谱结构可视化、节点社群分析与关键路径抽取能力，可直观展示 “文献理论 - 专利技术 - 实验数据” 的关联网络。 |
| 多模态语义检索与路径导航 | 构建统一嵌入空间下的多模态语义检索引擎，支持从任意模态出发的科研知识图谱检索与路径发现。训练跨模态嵌入模型（支持文本、图像、图表、公式、表格向量表达），通过三元组对比学习强化文献 - 专利 - 数据库实体的语义关联；实现 “以图找文”“以表找专利”“以公式查数据库” 等多模态反向检索能力，支持跨资源类型的混合检索请求；建立路径导航机制，支持图谱中多跳路径、多解路径的智能排序与结构可视化，可自动发现 “技术问题（专利）- 解决方案（文献）- 验证数据（数据库）” 等科研关键路径；集成图结构约束与嵌入相似度，实现 “语义 + 结构” 双因子排序，优先返回跨三类资源的关联结果。 |
| 图谱增强的高质量科研问答 | 以知识图谱结构为基础，结合大语言模型能力，打造具备准确生成 + 可追溯 + 结构清晰的科研 AI 问答系统。包含问题识别模块、证据构建模块、回答生成模块、、可视化联动模块等。 |
| 8 | 知识图谱构建服务 | 知识图谱构建标准规范 | 根据项目实际需求，提供不少于2名知识图谱构建人员驻场，提供三年服务。包括：知识图谱构建标准规范 、知识构建类服务、图谱构建服务、知识融合服务、知识计算服务、科技专家/人才知识推理构建服务。  知识图谱构建标准规范。需提供知识图谱构建标准规范，包括知识构建类标准规范、图谱构建标准规范、知识计算标准规范、科技产业链分析图谱构建标准规范。  知识构建类服务，包括概念抽取服务、上下位关系服务、属性识别服务、事件建模服务、时空建模服务、献知识建模服务、专利知识建模服务。  图谱构建服务。包括实体识别与分类服务、事件知识提取服务、概念实例提取服务、事件关系抽取服务、实例属性关系抽取服务。  知识融合服务。包括Schema匹配服务、实例匹配服务、知识链接服务、跨语言知识融合服务、数据语义链接服务、多模态科研知识服务、科研本体知识图谱服务、图结构嵌入与拓扑分析服务、图推理生成与类比设计能力服务。  知识计算服务。包括知识表示模型训练服务和知识推理模型训练服务。  科技专家/人才知识推理构建服务。包括科技专家/人才直接关系、科技单节点间接关系、科技两点合作成果、科技专家同事关系、科技专家校友关系、科技专家论文合作关系、重点关注科技企业关系、科技产业链点TOP-N事件关系、科技产业链全景图。 |