

指标体系	指标项（权重）	指标子项
管理层面（15分）	组织保障（6分）	<ul style="list-style-type: none"> <li>● 专门信息化建设部门</li> <li>● 专门项目管理团队</li> <li>● 后期维护部门规划</li> </ul>
	质量管理能力（5分）	<ul style="list-style-type: none"> <li>● 信息收集</li> <li>● 系统日志收集</li> <li>● 日志分析、处理和归档与反馈</li> </ul>
	过程管理能力（4分）	<ul style="list-style-type: none"> <li>● 项目管理制度</li> <li>● 项目管理经验</li> <li>● 项目管理资质情况</li> </ul>
社会层面（10分）	法律风险与责任（5分）	
	知识服务市场接受度（5分）	

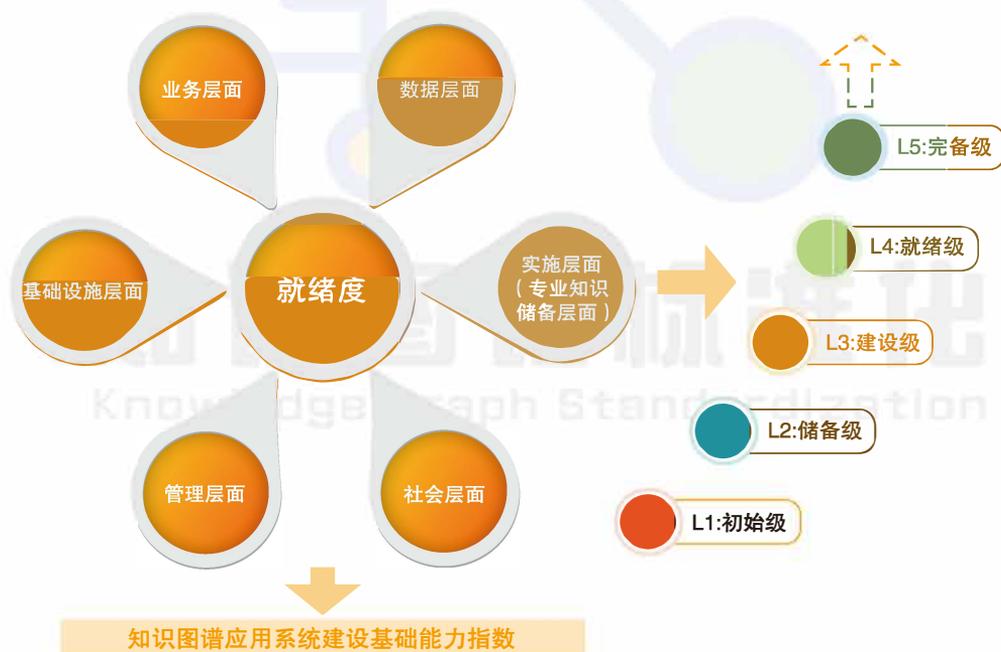


图5-3 知识图谱应用系统建设就绪度等级划分

根据上述层面各指标项所获分值的差异，可得出....。进一步，根据所计算指数，基础能力可分为：初始级、储备级、建设级、就绪级、完备级等五个等级，如图5-3所示。其中：

- 初始级是指不具备建设知识图谱应用系统的条件，应用场景、资源投入等尚不明确，有待进行战略规划。

- 储备级是指已进行战略规划，但不具备建设知识图谱应用系统的条件，有待在规划完善的基础上进行人力资源储备、基础设施储备、数据储备等。

- 建设级是指初步具备知识图谱应用系统的建设能力，部分维度仍存在较大弱项，有待全面提升各维度能力等；可针对性地开展小型知识图谱应用系统建设及研究工作。

- 就绪级是指已具备良好的知识图谱应用系统建设能力，可进一步对各层面欠缺指标项对应能力进行提升；有能力开展较大规模的知识图谱应用系统的建设及研究工作。

- 完备级是指已具备完备的知识图谱应用系统建设能力，各维度能力较为全面，可支持后续系统的建设、实施与部署；可依据业务需求开展大规模的知识图谱应用系统的建设及研究工作。

表5-2针对19项指标给出了各项指标自我评估判据，每项指标分值1至5分，总分19至95分。可对每项指标水平进行自评，根据总分做出相应决策。

- 总分位于35分及以下，初始级；
- 总分位于35至45分（含）区间，储备级；
- 总分位于45至60分（含）区间，建设级；
- 总分位于60至75分（含）区间，就绪级；
- 总分位于75分以上，完备级。

### 三、应用系统建设基础能力自我评估表

表5-2 应用系统基础能力自我评估表

能力子项	一级	二级	三级	四级	五级	
业务层面	战略愿景	对 AI 感兴趣	AI 潜在价值	AI 产品或服务已提供商业价值	AI 已广泛应用并带来利润	AI 已成为公司业务发展的组成部分
	用户需求驱动	在当前业务中知识图谱没有明确需求。	知识图谱具备潜在应用需求，具体应用需求有待挖掘。	知识图谱具有明确应用需求，基于知识图谱应用可满足部分用户需求。	知识图谱具有明确应用需求，且知识图谱可满足用户需求。	知识图谱可完全满足用户需求。
	资源安排	几乎没有预算	开展知识图谱项目 POC	已启动部分知识图谱项目并持续进行实验和 POC	将知识图谱项目建设纳入预算体系	有充足的 AI 及知识图谱建设预算
	行业资质	不具备行业资质。	具备行业资质，但行业壁垒低。	具备行业资质，但行业壁垒中等。	具备行业资质，但行业壁垒较高。	具备行业资质，且具有极高壁垒。
	行业趋势和产业政策	行业发展趋势不好，没有产业政策支持。	行业发展进入瓶颈，没有产业政策支持。	行业发展势头向好，没有产业政策扶持。	行业发展势头向好，产业政策扶持逐渐变多。	行业发展趋势火热，产业政策扶持力度大。
数据层面	数据储备	数据基础较差	数据基础较差但有部分可用数据	数据整体情况一般但部分数据情况较好	数据整体情况较好	有充足的数据积累及规范的数据存储体系
	领域专业数据基础	不具备行业语料、术语或者标准基础。	具备少量行业语料。	具备少量行业语料和术语。	具备少量行业语料和术语，并形成小范围行业标准。	具备大量行业语料、术语和基础标准。
	数据治理能力	不具备数据治理能力	数据治理水平低下，缺乏数据安全管理与保障能力	数据治理水平中等，初步具备数据安全管理与保障能力	数据治理水平较高，数据安全管理与保障能力中等	数据治理能力强，数据安全与管理与保障能力强，具备完备的数据治理工具。

能力子项	一级	二级	三级	四级	五级	
基础保障层面	人员储备	缺少领域知识专家和运维人员	具备运维人员，不具备领域知识专家	具备运维人员和领域知识专家。	具备运维人员、领域知识专家和知识图谱构建人员。	具备运维人员、领域知识专家、知识图谱构建人员、以及知识图谱应用人员。
	硬件资源储备	不具备存储能力、计算能力，不支持数据处理与存储	具有有限的存储能力、计算能力，支持小规模数据处理与存储。	具有一定的存储能力、计算能力，支持中等规模数据处理与存储。	存储能力、计算能力充沛，支持大规模数据处理与存储。	存储能力充沛，且支持动态扩展。支持超大规模数据处理与存储。
	专业知识储备	不具备专业知识储备。	具有有限的专业知识认知，对相关专业知识有粗浅的了解。	具备一定的专业知识，并能够支撑进行较为简单的应用和研究。	具备较为全面的专业知识，可以支持大规模系统的应用和研究。	全面深入的掌握了专业知识，支持大规模系统应用和研究，并能引领行业创新。
	信息化建设基础	无配套业务系统。	配套业务系统，少量业务流程实现信息化。	配套业务系统，部分业务流程实现信息化。	配套业务系统，大部分业务流程实现信息化。	配套业务系统，业务流程完全实现信息化。
	资格资质	不具备任何资质，不支持知识图谱应用系统集成和应用。	有一定系统集成或管理经验，取得少部分资质，但不支持知识图谱应用系统集成。	具备基础的资质和一定程度上进行系统集成。	具备全面的系统集成资质，可以进行大规模系统集成。	具备全面的资质，且有颁发资质的资格，可以制定系统集成资质。

	能力子项	一级	二级	三级	四级	五级
实施层面	验证评估能力	没有任何验证能力。	有一定的验证经验，可以进行简单的评估	具备一定的验证评估能力，可以对较为常见和简单的性能进行验证和评估。	有较为全面的验证评估能力，可以对知识图谱应用系统性能进行全面的验证和评估。	拥有权威的知识图谱应用系统验证和评估能力，其验证和评估结果可以作为行业标准。
	运维能力	在当前业务中知识图谱没有明确应用需求。	知识图谱具备潜在应用需求，具体应用需求有待挖掘。	知识图谱具有明确应用需求，基于知识图谱可满足部分用户需求。	知识图谱具有明确应用需求，且知识图谱可满足用户需求。	知识图谱已经满足用户部分需求，且知识图谱可完全满足用户需求。
	信息化建设能力	在知识图谱应用系统建设业务中，没有相关的信息化建设能力。	有知识图谱建设的规划能力，具体建设能力有待挖掘和加强。	具有一定的信息化建设能力，可以支持知识图谱的简单建设和应用。	具备知识图谱应用系统的全面建设能力，支持大规模系统的建设和应用。	具备满足知识图谱建设所需的所以信息化建设能力，并支持动态扩展。
管理层面	组织保障	未形成组织机构或团队	初步形成组织团队，能够组织开展。	形成严密的组织机构，并具备初步的组织措施。	形成严密的组织机构，具备规范的组织措施，初步形成规章制度。	形成严密的组织机构、完备的组织规章、规范的组织措施。
	质量管理能力	未针对项目进行质量管理	形成初步的质量目标和职责	制定严格的质量方针，明确质量目标和职责。	制定严格的质量方针，明确质量目标和职责，初步形成对项目各阶段活动质量的控制。	制定严格的质量方针，明确质量目标和职责，严格控制项目各阶段活动质量。
	过程管理能力	未形成对项目过程的管理	初步形成项目过程策划	具备完备的过程策划，并在实施过程中，初步形成对过程的检查。	具备过程策划、过程实施、过程检查步骤，并初步形成对过程的改进。	具备完整的过程策划、过程实施、过程检查、过程改进阶段，并严格按照阶段执行。

	初阶	进阶	中阶	高阶	专家级
法律风险与责任	极高法律风险和责任	较高法律风险和责任。	中等法律风险和责任。	较低法律风险和责任。	无法律风险，无责任风险。
知识服务的市场接受度	市场不了解且不接受其知识服务。	市场对知识服务有一定的了解，接受极少的服务。	市场接受知识服务，但其程度有限。	市场全面接受其知识服务，对其服务有较强的认同感。	全面的接受知识服务，且用户偏向于使用其服务。

## 四、应用系统建设基础能力提升

根据知识图谱应用系统建设基础能力的就绪度等级和评估结果，可从人员、技术、资源、战略定位为切入点出发制订能力提升实施方案，并从业务、数据、基础保障、实施、管理、社会等六个层面采取措施，进而迭代通过能力提升效果评估方法进行效果评价，从而形成一整套闭环的能力改进提升方案，如图5-4所示。

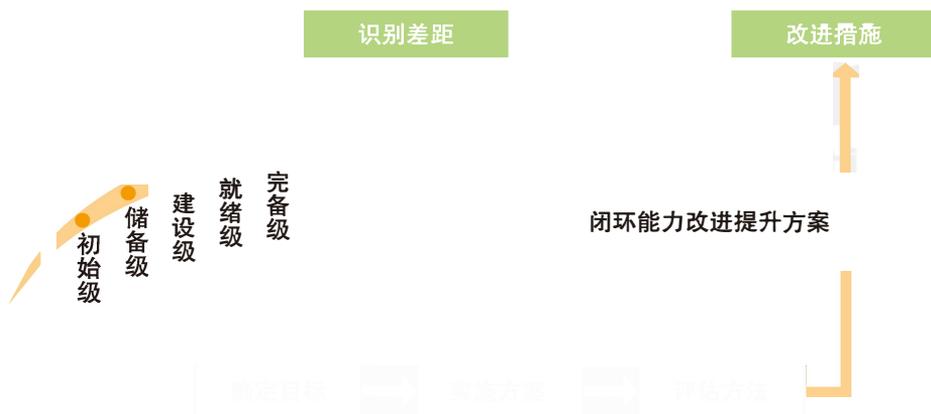


图5-4 知识图谱应用系统建设能力提升路径示意图

### 业务层面：

战略方面，重视人工智能技术在当前业务流程中的应用；业务需求方面，提升知识图谱在现有业务流程中的有效性，并不断挖掘潜在应用需求，基于部分应用场景构建图谱应用案例，以小范围样例作为业务应用试验验证。资源安排方面，提升人工智能开发及知识图谱建设预算，增加人工智能开发团队人员数量。行业资质方面，开展资格认证和评级，不断提高行业资格水平。行业趋势和产业政策方面，紧跟行业应用趋势和产业政策方向，在行业内开展相关知识图谱的建设和业务推广。

### 数据层面：

数据储备方面，开拓数据获取渠道，加大数据储备，对内企业组织整理内部业务流程数据、整合已有的数据存储资源；对外企业获取第三方组织机构的数据（例如：购买领域数据），从互联网获取开源的数据资源等。领域专业数据基础方面，积累行业术语、标准数据集。数据治理能力方面，提升现有的数据治理水平，增强数据清洗、挖掘能力，加强数据的安全防护等。

### 基础保障层面：

人员储备方面，优化企业内部的人力资源结构，拓展专业领域人员数量，进行知识图谱领域的研发、实施、运维、业务人才储备，提升人员素质，以满足可持续性开发建设知识图谱的需求。硬件资源储备方面，根据业务市场发展需求，不断丰富计算设备资源，提升云计算能力，满足大数据治理的需要。信息化建设基础方面，提升软件基础工具和环境条件，不断提高信息化系统在企业业务流程中的应用占比，建立企业知识库，逐步积累行业数据。

### 实施层面：

验证评估能力方面，持续拓展和丰富业务专家库，引入国内外顶尖专

家，定期开展硬件资源和建设能力评价，不断积累和提升专业知识水平。实施能力方面，通过实施不同项目案例种类、数量和规模的积累，进一步提高IT资质等级。运维能力方面，针对不断发展变化的业务需求和系统运行中的问题，不断寻求更优的解决方案，加快知识更新的效率。信息化建设能力方面，通过对业务系统的分类梳理，不断细化业务场景，完善知识可视化、知识问答和联想等应用，提高业务人员对知识图谱系统的理解。

### **管理层面：**

组织保障方面，形成严密的组织机构、完备的组织规章、规范的组织措施。质量管理方面，制定严格的质量方针，明确质量目标和职责，严格控制项目各阶段活动质量。过程管理方面，具备完整的过程策划、过程实施、过程检查、过程改进阶段，并严格按照阶段执行。

### **社会层面：**

法律风险与责任方面，不断完善业务流程、规范化经营，降低法律风险与责任。知识服务市场接受度方面，紧密跟踪行业趋势和产业政策，积极分析市场需求和痛点；不断提升企业自身行业资质，不断提高自身竞争力形成竞争壁垒。

此外，在研发方面，加强知识图谱相关的人员储备（包括：运维人员、知识专家等）；加强软件、硬件等基础设施建设，提升数据存储、数据处理的能力。

## 第六章 知识图谱应用系统选型

### 一、知识图谱应用系统选型准则

根据图4-1所示，知识图谱应用系统的模块根据功能定位的不同，可以划分为面向知识图谱构建前的数据接入与处理相关模块、面向知识图谱构建中的知识获取、融合、存储等相关模块、面向知识图谱构建后的知识服务与管理相关模块。由于各模块功能模块定位不同，所涉及的选型准则存在差异。

#### （一）数据处理相关模块的选型准则

面向知识图谱构建前的数据接入与处理相关模块，可聚焦数据准备相关需求，从数据类型、数据接入、数据治理、数据组织、数据服务、数据处理等维度进行选型，并根据业务需要对准则进行裁剪，如图6-1所示：

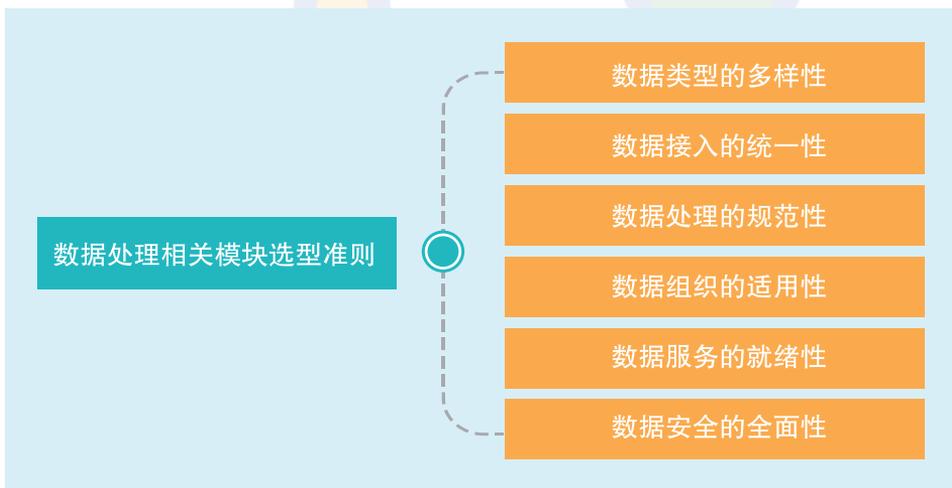


图6-1 数据处理相关模块的选型准则

**数据类型的多样性准则：**支持多样化的数据类型，包括结构化数据、半结构化数据、非结构化数据，并可支持多种模态数据类型，如：文本、图片、视频、音频等。

**数据接入的统一性准则：**具备统一的接入工具实现异构数据资源汇聚的过程；支持数据采集、汇聚等多种方式，为数据接入提供标准化、模块化、可适配的多源异构数据资源接入模式。

**数据处理的规范性准则：**支持数据的标准化、专业化管理，如元数据管理、数据质量管理、数据标准处理、数据血缘管理、数据资源目录管理等；支持对数据清洗、存储等治理流程的标准化，实现数据资产的有效组织和管理，为知识图谱上层各类业务应用提供支撑。

**数据组织的适用性准则：**支持数据组织，如组织原始库、主题库等。原始库是描述各种数据资源、支撑各项业务工作的公共数据集合，对不同来源的数据，按照数据的原始格式进行存储；主题库是根据数据的主题及应用需求建立，如车辆库、人员库等。

**数据服务的就绪性准则：**支持其他业务模块对数据的便捷访问，畅通数据的获取路径。

**数据安全的全面性准则：**保障数据处理过程中有效的身份识别和权限控制，确定其身份安全可信后，才可访问数据、应用和服务。

### （二）知识图谱构建相关模块的选型准则

面向知识图谱构建过程的各功能模块，可聚焦知识图谱构建技术路径各环节和任务，进行模块选型并根据业务需要对准则进行裁剪，可包括知识表示模块、知识建模模块、知识获取模块、知识存储模块、知识融合模块、知识计算模块等如图6-2所示。其中，每一个模块所包含的选型准则定义如下：

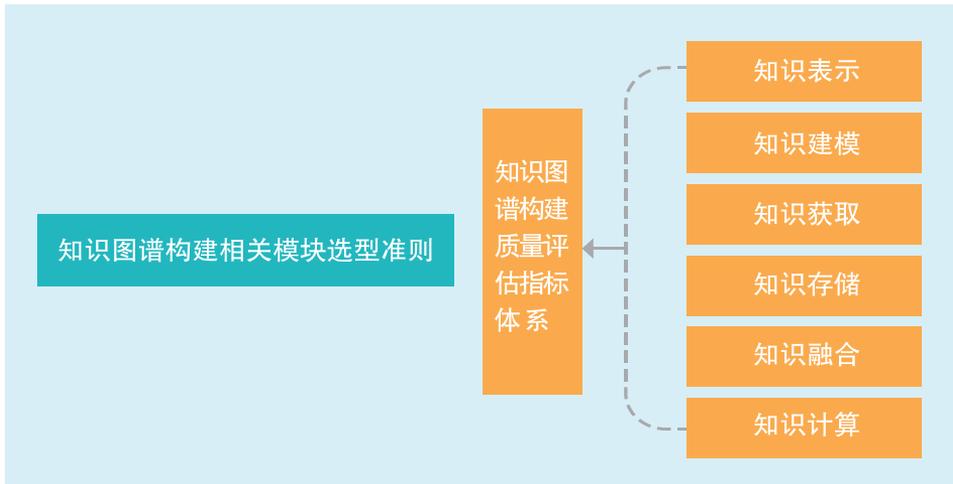


图6-2 知识图谱构建相关模块的选型准则

### 1、知识获取模块选型准则

知识获取模块可基于接入数据和已有本体模型，完成实体信息、实体间的关系信息、实体的属性信息、本体模型中缺失信息等信息的自动提取。面向的数据类型可包括：结构化数据，如存储于关系型数据库中的业务数据；半结构化数据，如百科数据等；非结构化数据，如文档、图片、视频、音频等。

### 2、知识表示模块选型准则

知识表示模块可基于知识图谱应用需求、应用约束、质量要求等，实现对知识表示模型的设定，包括知识表示框架、知识表示元素、适用范围等。知识表示形式可采用RDF图和属性图等形式：属性图中顶点表示实体和事件，顶点间的有向边表达实体/事件之间的语义关系；RDF图中，顶点是具有唯一标识符的资源，顶点间的有向边也称为谓词或属性。

### 3、知识建模模块选型准则

知识建模模块可基于知识图谱应用需求、业务规则、专家或行业知

识、企业数据现状等，形成涵盖实体类型体系、实体类型属性、实体类型间关系等的本体模型及其图式（Schema）。

#### 4、知识存储模块选型准则

知识存储模块需完成本体模型，知识获取、知识融合、知识计算、知识演化等模块形成的知识单元，业务规则、约束及算法模型等内容的存储，并为后续的知识查询、调用等服务。当知识图谱的规模庞大时，应考虑结合使用分库、为同类顶点聚合等方式作为知识图谱存储设计的准则。

同时，知识存储后数据查询的DSL表达能力也应纳入选型的范畴中。DSL应足够简洁以降低使用者的技术水平要求，应足够灵活以应对多种使用场景，应足够强大以满足知识图谱专家和复杂分析计算对数据查询的要求。在知识存储架构设计和选型阶段，图数据库及其运行环境（如操作系统、硬件和指令集）的国产化也是衡量的重要方面。

#### 5、知识融合模块选型准则

知识融合模块可基于本体模型、知识获取输出的知识单元及外部知识图谱等，整合形成具有全局统一知识标识的知识单元，并识别本体模型中缺失的信息。

#### 6、知识计算模块选型准则

知识计算模块可基于知识表示模型、本体模型、已存储的知识单元、应用场景中的计算需求，输出计算获得的新知识并为下游任务提供知识计算服务。例如：通过统计分析对知识图谱蕴含知识结构及其特征进行统计与归纳；通过推理计算从中已有的事实或关系进行隐性知识的发现与挖掘。

作为知识图谱的重要应用，知识计算需要大规模地抽取顶点和边的数据到内存中进行计算，这对计算效率提出了非常高的要求。因此，是否能

并行地、分布式地进行知识计算，以及计算性能的垂直扩展和水平扩展、是否能线性地水平扩展是知识图谱应用系统选型的关键。

### （三）知识服务与管理相关模块的选型准则

面向知识图谱构建完成后的知识服务与管理相关模块，可聚焦知识服务、业务集成、知识管理与维护、系统安全保障等任务，进行模块选型并根据业务需要对准则进行裁剪，如图6-3所示。选型准则可包括：安全性、可靠性、响应性、可移植性和易用性。

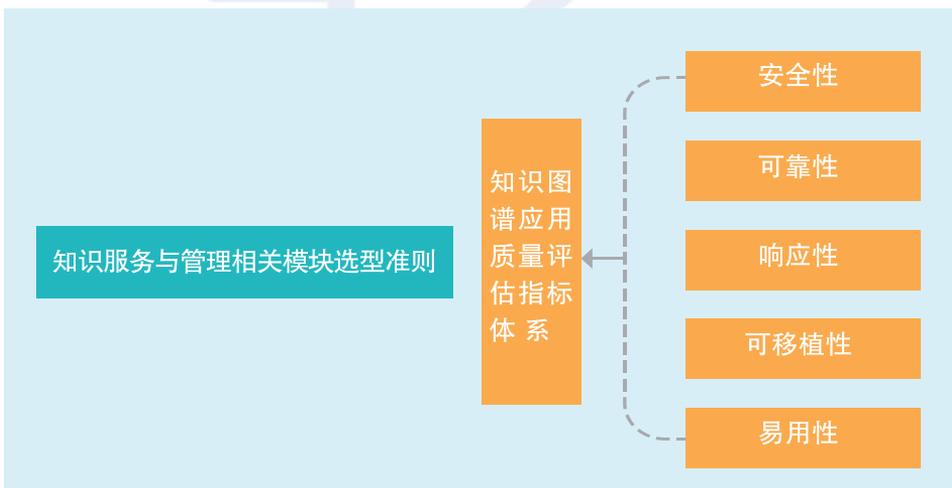


图6-3 知识服务与管理相关模块的选型准则

#### 1、安全性

知识图谱应用系统的安全性要求主要包含功能域安全要求和业务数据域安全要求：

功能域安全要求是指在应用系统的业务流程中涉及功能的安全要求，主要针对工作流中包含的各项功能流转的合法合规性，确保用户所属的角

色能访问的系统功能和适用的安全策略不越界。

业务数据域安全要求是对用户、用户组、用户角色与数据条目访问授权的多对多映射，目的是确保每个用户能够且仅能够访问到对应密级的数据条目。

知识图谱应用系统的安全性要求不应只体现在应用系统中，还需贯穿知识图谱构建中和构建后的各环节。例如，在知识存储、知识获取的环节中，就需使用符合数据域安全的限制要求和过滤请求的数据条目。进而，保证在知识融合、知识计算、知识演化和业务整合等后续步骤中不会将超出访问权限范围外的数据带入融合、计算、演化和整合结果中。

### 2、可靠性

可靠性是指衡量知识图谱应用系统在指定的场景和时间等条件下可稳定完成用户需求的程度，如：是否可以在规定时间内完成预定功能，是否支持规定场景的查询、推理、计算等。

除上述知识图谱应用系统对可靠性的要求外，可靠性的一般性要求还有成熟性、容错性、易恢复性等若干特性和指标，可参考国家标准GB/T 29832.1-2013《系统与软件可靠性 第1部分：指标体系》中的定义。

### 3、响应性

响应性是系统高效完成用户业务目标的基本保障。用于度量形成的知识图谱应用系统在收到请求后返回的结果及返回结果的过程所表现出来的能力，既包括非耗时请求的即时响应时间，也包含耗时查询、计算的响应时间、反馈结果的质量等。此外，耗时查询、计算过程应通过进度条、预估剩余时间或状态更新来降低用户等待焦虑。

### 4、可移植性

可移植性是用于度量形成的知识图谱应用系统在不同软硬件环境间移植的能力，主要目标是规范业务流程和接口定义，使得第三方应用接入和

迁移成本相对较低。

知识图谱应用系统的外部开放接口在设计和实现上需能达到足够的抽象程度，以适应不同的业务领域和使用场景。进而，避免第三方应用的重复开发，也能够确保同一个第三方应用在不同知识图谱应用系统迁移对接时的兼容性。

### 5、易用性

易用性是用于知识图谱应用系统在指定条件下使用时，知识图谱被理解、学习、使用和吸引用户的能力，并衡量使用者在利用系统解决业务问题时的难易程度。其中，既包括知识图谱应用系统各项功能对使用者的友好程度、 workflow 复杂程度，也包含知识图谱本身以及业务建模对使用者所拥有领域知识掌握程度的要求。

## 二、知识图谱应用系统的关键性能构成

### （一）数据处理相关模块的关键性能构成

**围绕数据接入的统一性，相关模块的指标构成可包括：**

- 具备支持结构化数据、半结构化数据、非结构化数据的数据接入能力；
- 数据接入工具符合统一的标准规范；
- 支持多种标准化、模块化、可适配的数据接入模式，满足多源异构数据的采集与汇聚；
- 支持数据批量接入。

**围绕数据处理的规范性，相关模块的指标构成可包括：**

- 支持多源异构数据的处理；
- 支持多模态数据的处理（可选）；

- 保障数据处理过程的模块化、规范化；
- 保障数据的归一化、标准化；
- 支持噪声数据的处理或补全；
- 支持数据的批量处理。

**围绕数据组织的适用性（可选），相关模块的指标构成可包括：**

- 支持多样化数据组织方式，如三元组、属性图等；
- 支持数据的多样化展现方式，如图、表格等。

**围绕数据安全的全面性，相关模块的指标构成可包括：**

- 从数据接入、数据传输、数据存储、数据处理等层面进行身份识别和权限控制；
- 根据知识溯源的要求，进行数据的查阅和跟踪；
- 保障数据采集、存储、处理、销毁等数据处理周期安全管控；
- 支持访问控制、数据加密、数据备份、数据脱敏、数据水印等数据安全能力。

## （二）知识图谱构建相关模块的性能构成

**知识表示模块的性能构成可包括：**

- 支持图形化的知识呈现方式；
- 支持三元组、向量和的知识表示形式等。

**知识建模模块的性能构成可包括：**

- 支持以可视化、拖拽等方式构建本体模型；
- 支持以增量方式构建本体模型；
- 支持手动添加实体类型、关系类型、属性等；
- 支持从外部文件中增量或批量导入图式（Schema）；

- 支持对图式（Schema）进行维护管理，包括新增、修改、删除等；
- 支持增量或批量导出图式（Schema）。

### 知识获取模块的性能构成可包括：

- 支持从数据库表等结构化数据中抽取知识；
- 支持从表格、网页、表单等半结构化数据中抽取知识；
- 支持从文本、图像视频等非结构化数据中抽取知识；
- 支持用户自主选择知识抽取方式，如单模型抽取、多文件多模型抽取；
- 支持用户自定义知识抽取模型；
- 支持查看知识抽取结果报表（包括当前及历史）及数据详情；
- 支持查看知识生成结果（包括当前及历史）；
- 支持人工配置知识映射规则；
- 支持用户自主配置数据源优先级；
- 支持实体、关系、属性的编辑，如新增、修改、删除、查询等。

### 知识存储模块的性能构成可包括：

- 支持实体、属性及关系的检索；
- 支持自定义实体、属性、关系；
- 支持实体、关系、属性的更新、删除；
- 支持按照数据权限或知识图谱主题授权用户查看功能；
- 支持数据审批流程可视化管理；
- 支持系统操作日志管理；
- 支持对删除后的实体、属性及关系的恢复；
- 支持按照版本号对同一应用场景下的知识图谱进行管理；
- 支持对算法模型的维护管理；
- 支持元数据管理，如：知识图谱在存储时在实体和关系上添加描

述密级或权限规则的元数据，在有多个业务来源时也需要考虑将子图隔离存储。

### 知识融合模块的性能构成可包括：

- 支持实例级（实例和属性值）的对齐；
- 支持知识映射过程中的异常发现，如：给出映射失败数据并给出映射失败原因；
- 支持用户自主配置知识融合、消歧规则；
- 支持不同应用场景下知识图谱间根据规则的融合。

### 知识计算模块的性能构成可包括：

- 支持基础图计算功能，如：社区发现、度中心性计算、紧密中心度计算、实体节点排名、中介中心度、最小生成树等；
- 支持给出起始节点和目标节点的知识图谱路径集合；
- 支持通过输入外部数据等人工干预方式训练知识抽取模型；
- 内置语料库及语义分析模型；
- 支持从节点属性推理补全知识图谱关系。

## （三）知识服务与管理相关模块的性能构成

### 围绕安全性，相关模块的性能构成可包括：

- 支持权限控制和安全隐私管理；
- 支持知识图谱存储中实体和关系添加描述密级或权限规则的元数据；
- 支持有多个业务来源时的子图隔离存储；
- 支持用户角色权限管理、数据脱敏等；
- 支持查询引擎返回用户数据前，检查数据权限合规性。

### 围绕可靠性，相关模块的性能构成可包括：

- 支持在规定场景、规定时间内完成预定功能；
- 支持平均故障间隔时间小于用户或行业要求；
- 支持规定场景的查询、推理、计算等要求。

### 围绕响应性，相关模块的性能构成可包括：

- 支持用户业务操作响应时间小于用户或行业要求；
- 支持服务接口请求响应时间小于用户或行业要求；
- 支持耗时查询、计算过程中通过进度条、预估剩余时间或状态更新来降低用户等待的焦虑；
- 支持知识体量、知识图谱复杂度的统计。

### 围绕可移植性，相关模块的性能构成可包括：

- 支持基于特定业务场景切分数据，如子图拆分、多图融合等；
- 支持常用的硬件环境；
- 支持常用的操作系统；
- 支持通过接口形式的数据调用。

### 围绕易用性，相关模块的性能构成可包括：

- 支持良好的人机交互界面；
- 支持根据业务场景利用复杂查询、图算法等方式对数据二次加工并提供服务；
- 支持基于常用查询语言，提供可视化交互查询，并针对特有业务要求，提供具有行业特色的交互展示。

## 第七章 知识图谱应用系统建设与管理过程

知识图谱生命周期根据构建过程及系统开发与部署过程的差异从技术构成与工程实施视角可细分为知识图谱构建生命周期和知识图谱应用系统生命周期两个维度。其中，识生命周期由知识表示、知识建模、知识获取、知识融合、知识存储、知识计算、知识应用和维护等阶段构成；知识图谱应用系统生命周期由需求分析、方案设计、图谱构建、应用开发与集成部署、系统评估与验收、运营推广、管理维护等阶段构成，如图7-1所示。

知识图谱应用系统生命周期各阶段因定位不同，其涉及的知识生命周期环节、建设目标、输入和交付物、主要活动、参与角色、需用户参与的内容等存在差异，系统建设与管理过程中可进行关注和整体规划，以保障各环节的有效衔接及知识图谱应用完整交付和后续运营质量。



图7-1 知识图谱应用系统建设与管理过程

## 一、需求分析阶段

### （一）阶段建设目标

**需求分析阶段的建设目标主要包括：**

**目标1：获取完整准确的用户需求。**

为了全面、准确地获取业务信息，更好服务于产品设计和迭代，从多个维度调研用户需求。

**目标2：充分理解、认识和分析用户需求。**

结合业务场景，分析用户行为，充分了解实际业务过程，挖掘用户真实需求。

**目标3：通过调研和分析得出目标系统的逻辑模型。**

对用户提出的知识图谱应用系统业务需求与数据需求进行分析和转换，构筑相关业务模型与数据模型。

**目标4：评估目前所具备的开发环境和条件。**

评估内容可包括：技术能力是否支持，企业经济效益如何，与企业的经营和发展方向是否吻合，系统投入运行后的维护有无保障等。

**目标5：评估已有数据质量情况。**

评估内容可包括：各环节规划中的数据质量是否符合要求，数据范围和影响程度是否一致，是否满足业务需要，能否满足前期调研的需求等。

**目标6：系统功能分解。**

对用户的建设需求和目标进行拆解，并对各参与方的任务进行细化，确保各参与方输出的内容可集成为符合用户需求的知识图谱应用系统。

### 目标7：编写需求相关文档。

根据需求分析结果，形成业务需求说明书、软件需求说明书、服务或功能说明书、知识图谱设计说明书等文档。

## （二）阶段输入输出

### （1）需求分析阶段输入

本阶段的输入可包括但不限于：

——用户需求：知识图谱应用系统的数据来源、建设目标等，如预期知识图谱内容和可视化效果、交互界面等预期应用效果。

注：用户需求宜通过需求说明书等书面文件方式提交确认。

——业务数据：用户或生态合作伙伴在所属业务领域内，可提供的业务领域数据，包括基础数据及加工转换后的数据等。

——算法模型。

### （2）需求分析阶段输出

本阶段形成的交付物及内容简介如表7-1所示：

表7-1 需求分析阶段交付物清单

序号	交付物名称	简介	重要性	是否
			(1-5级)	可选
1	行业调研报告	通过对当前行业的长期跟踪监测，分析行业需求、供给、经营特性、获取能力、产业链和价值链等多方面的内容，整合行业、市场、企业、用户等多层面数据和信息资源。行业调研报告是用户需求整理的基础，可以促进用户对行业现状的了解。	5级	必选交付
2	可行性报告	论证项目可行性，给出是否可行结论，明确项目可行性条件、不可行的原因、暂缓可行的方案等。对于可行性项目，分析已具备的基础设施条件、技术条件、经济条件及政策规范等。	5级	必选交付

序号	交付物名称	简介	重要性	是否
			(1-5级)	可选
3	预算文档	分析知识图谱应用系统的预算，分析项目调研、实施、部署和运维阶段的计算依据，明确预算支出科目及额度，可包括基础设施费、材料耗能费、研发劳务费、商务协作费等。	5级	必选交付
4	竞品调研文档	当前的行业内竞品的基本情况，作为判定项目是否可行的基础之一，可包括竞品行业背景、市场规模、用户群体、商务模式等，并比较分析结论。	2级	推荐交付
5	标准规范文档	对标准化对象提出的要求，规定在设计软件各个阶段中所形成的文档的格式和要求，可包括数据规范、数据集规范、专业基础标准、过程标准、质量标准、技术与管理标准、工具与方法标准等。	3级	推荐交付
6	业务需求说明书	需求文档可以使设计人员能够设计出可以满足用户需求的系统，并使测试人员能够测试该系统是否满足这些需求，可包括业务应用点、需求背景、应用场景、相关业务系统、业务规划文档、业务规则、业务术语等。	5级	必选交付
7	需求规格说明书	叙述该项软件开发的意图、应用目标、作用范围以及其他有关该软件开发的背景材料，可包括数据规约、用例规约、流程图、部署环境要求、系统指标要求、运行环境规约、需求优先级、质量属性等。	5级	必选交付
8	技术规格说明书	是一种技术文件，主要用来描述应用系统或产品的基础架构和应用关键技术及主要规格参数的文件，以便技术人员准确掌握设计意图和要求。包括法律和法规上的应用标准、产品质量的应用性、可靠性、可支持性、还包括工作制度、权责规范、数字权利、操作系统及环境、兼容性和设计约束等。	3级	可选交付
9	知识图谱应用系统功能清单	系统性展示知识图谱所有功能，以业务流程图和数据流程图的方式，整体上划分模块及逻辑关系，细节上展示每个模块内部构造及功能，包括知识图谱的构建、获取、展示和应用操作等模块。	5级	必须交付

序号	交付物名称	简介	重要性	是否
			(1-5级)	可选
10	数据集规范	根据项目背景、必要性分析、编制原则，主要技术内容说明，编制知识图谱的数据集的规范、规则，定义数据集适用范围、规范性引用文献、术语和定义、内容结构、基本数据集的元数据、基本数据集的元数据的元素的元数据等。	1级	可选交付
11	语料库分析报告	立足行业背景，明确用户语料库建设目的和语料选取，详细分析用户语料库的需求和要求，包括语料库规模设计、建设内容、实施目标和步骤等。	3级	推荐交付
12	数据样例	在数据源中提取出来的原始数据，反映出数据的特点，数据样例中应包含各个字段的描述。	3级	可选交付
13	数据规范	根据数据来源和知识应用去向，可以生成知识图谱的数据的规范、规则，定义数据集适用范围、规范性引用文献、术语和定义、数据建模内容、基本数据模型的元数据等。	5级	必选交付
14	系统原型说明书	根据需求分析的结果，设计出能最大限度满足所要求的系统原型设计，将其整理为说明书。	3级	推荐交付
15	数据调研说明书	前期需要做好数据探查工作，需要了解数据库类型、数据来源、统计学分布、数据规模、数据增量，全量数据情况及数据每年增长情况，更新机制；还需要了解数据是否结构化，是否需要清洗，是通过接口调用还是直接连接数据库，设计数据实体和库表；定义数据类型和数据结构，还包括图谱待加工数据清单。	2级	可选交付
16	风险管理计划	风险管理计划制定风险识别、风险分析、风险减缓策略，确定风险管理的职责，为项目的风险管理提供完整的行动纲领，确定如何在项目中进行风险管理活动，以及制定项目风险管理计划的过程，可包含项目排期，项目功能开发、测试和部署等计划安排等；风险摘要、风险管理任务、组织和职责、预算、工具和技术、需要管理的风险项等。	3级	可选交付

序号	交付物名称	简介	重要性	是否
			(1-5级)	可选
17	备选策略分析	作为知识图谱技术可行性、经济可行性和政策法规方面的备选策略，分析知识图谱应用系统的备选策略及实施方案，描述备选策略环境适应性、方案实施内容和步骤、规避风险因素等。	3级	推荐交付
18	工作说明文档	为了保证知识图谱项目按计划可持续推进，制定参与工作人员和事务处理规范文档，包括项目人员管理制度、工作人员安全责任制度、项目排期计划推进监督制度、项目质量管理体系等。	3级	推荐交付
19	权责规范	为了保障知识图谱项目用户方和实施方等各方的权利和义务，制定权责说明规范书，充分保证各方应该享有的知识产权、经济权利、人身安全权利等，可包括物质权利、数字权利、精神和文化权利等，以及应尽的义务。	4级	推荐交付
20	遵守的法律法规	知识图谱应用系统利益各方和从业人员，行业需遵守的法律法规，可包括知识产权法规、安全生产法、隐私保护法、商业销售和许可法等。	4级	推荐交付
21	可复用的通用图谱分析报告	调研或积累的可复用通用知识图谱，比较分析通用知识图谱优缺点，可用于垂直领域知识图谱的借鉴，以完善和优化实施方案。	3级	可选交付
22	部署环境要求	按照知识图谱的开发、测试、运维阶段，部署环境要求的文档分为三类：开发部署环境文档、测试部署环境文档、运行和维护文档，在开发阶段的部署环境文档描述图谱系统开发技术人员工作的软硬件环境，测试阶段的部署环境文档描述测试工作人员需要的软硬件、测试数据集，运维阶段的部署环境文档描述用户应用和管理需要的文档。	5级	必须交付

## (三) 主要活动

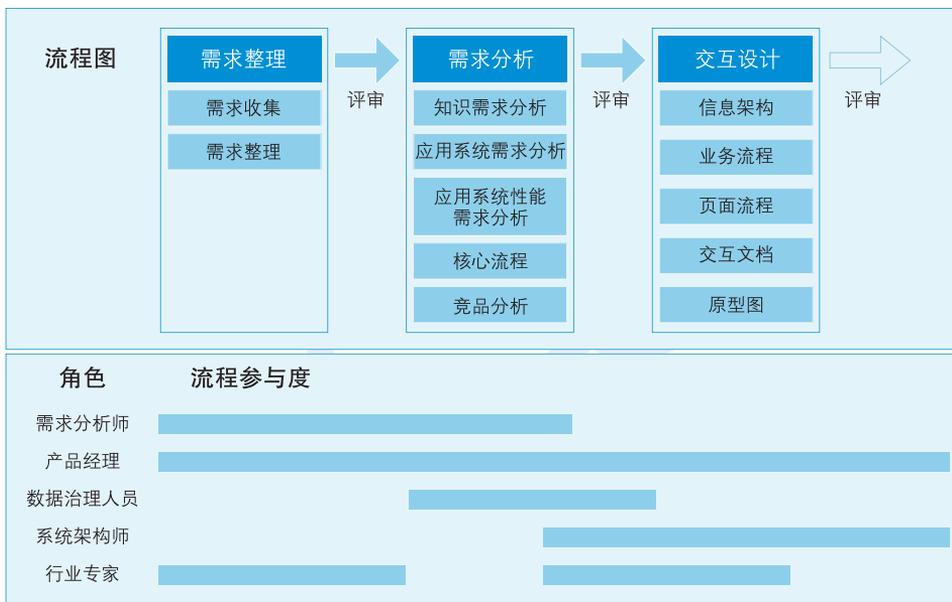


图7-2 知识图谱需求分析阶段活动流程图

需求分析阶段根据拟实现目标的不同可划分为需求整理、需求分析、交互设计等三项活动，本阶段活动流程及主要参与角色如图7-2所示。各项活动主要内容如下：

### (1) 需求整理

收集和整理用户现状和目标需求，可包括构建知识图谱应用系统的目的、约束条件、交付范围。同时，了解用户的业务数据、专家经验和已有软硬件环境，并与用户、专家进行沟通确认，形成完整的用户需求清单。

该阶段可通过行业调研、业务访谈、数据访谈等多种形式进行。其中，行业调研主要分析影响行业发展的主要敏感因素及影响力，预测行业未来发展趋势。业务访谈主要是与用户方的管理和业务人员沟通，从宏观上把握用户的具体需求方向和趋势。数据访谈主要是了解现有的数据资源、数据流程等具体情况，便于系统开发和应用的数据建模及构建数据流程图。

### (2) 需求分析

从知识、应用系统需求和应用系统性能需求等方面，进行数据分析和用户需求分析，绘制知识图谱应用系统业务架构图。同时，按照系统功能模块和实现流程进行拆分，划分各任务工作内容、边界输入/输出和依赖关系，并确定各部分的验收标准。此后，根据上述分析内容，撰写知识图谱应用系统的需求说明书，描述数据加工与转换过程、功能流程、交互逻辑等，使得各任务能按照需求说明书进行实际设计和研发工作。

该阶段可通过原型设计、需求引导或深层分析等方式完成目标。其中，原型设计是指项目组已了解具体用户方的组织架构、业务流程、硬件环境、软件环境、现有的运行系统等具体实际和客观的信息基础，结合现有的软硬件实现方案做出能够反映基本业务流程、易与用户交流的系统UI原型草图。用户通过UI原型草图可快速和形象地了解系统功能，评估整个业务流程设计的合理性、准确性，及时地提出改进意见。需求引导或深层分析是指结合以往的项目经验对用户采用引导式或启发式的分析方法和手段，与用户共同探讨业务流程设计的合理性、准确性和易用性。

### (3) 交互设计

该活动是设计与确定用户与系统交互方式的阶段，通过交互设计可将抽象的业务需求转化成可操作的高保真系统界面。本活动不仅需产品经理和交互设计师配合完成，还需与用户方完成需求的最终确认，可包括数据确认、流程确认、业务需求确认等。

其中，数据确认是指开发方需提供数据调研说明书、数据模型设计说明书、数据项表等，并能清晰地向用户描述系统的数据使用目标。流程确认是指开发方需提供明确的业务流程报告，并能清晰地向用户描述系统的业务流设计目标。业务需求是指用户方通过审查需求分析说明书、系统原型说明书来提出反馈意见并对已可接受的报告或文档签字确认。

#### (四) 用户角色与分工

本阶段的参与人员角色及其任务简介如表7-2所示：

表7-2 需求分析阶段参与人员角色及其任务简介

序号	角色	承担任务简介	交付物名称（如有）
1	业务/ 领域专家	特定领域的专家，具有该领域的丰富知识和技能。可分析影响行业发展的主要敏感因素及影响力，预测行业未来发展趋势。	行业调研报告 竞品调研文档 可复用的通用知识图谱
2	数据治理 工程师	负责数据内容调研、数据治理标准解决方案研究。针对业务需求，设计合理的数据分析框架和方案，并提供分析所得的意见；产出分析结果、数据问题、说明、并提供解决方案。	标准规范文档 数据规范 数据样例 数据调研说明书
3	业务人员	对接知识图谱需求分析师，负责协调用户与技术人员有效沟通，参与需求调研、协调用户，参与编写知识图谱用户需求分析报告、项目可行性报告等。	业务需求说明书 需求规格说明书 权责规范/数字权利 遵守的法律法规 备选策略分析
4	需求 分析师	跟用户交流，准确获取用户需求。需求分析师是项目前期与用户方接触最多的人，对于用户来说可代表整个项目组，对于项目组成员来说其意见可代表用户方的意见，项目组内所有与用户需求相关的事宜需与其确认。	需求规格说明书 部署环境要求 可行性报告 可复用的通用知识图谱
5	项目经理	负责产品开发进度的控制、风险评估进度的把控、工作任务的分配、项目中日常事务调配、人员配置等。同时，对项目过程进行监控，对项目的进度、质量负责。	需求规格说明书 风险管理计划 标准规范文档 技术规格书 工作说明文档 权责规范/数字权利 遵守的法律法规 备选策略分析 预算文档

序号	角色	承担任务简介	交付物名称（如有）
6	售前经理	将功能需求列表转化为整体解决方案交付给用户，协调其它部门资源，推进实施方案；与用户保持技术与方案上的沟通，及时解答疑问、响应新需求、处理突发故障；配合销售项目负责人完成售前技术支持。	业务需求说明书 需求规格说明书 备选策略分析
7	产品经理	负责市场调研并依据产品、市场及用户需求，确定研发产品方向，选择业务模式和商业模式等。同时，推动相应产品的开发设计组织，依据产品的生命周期，协调研发和运营等，并确定和组织实施相应的产品策略。	技术规格书 系统原型说明书 竞品调研文档
8	潜在用户	需求与公司可提供产品吻合的用户群体，有知识图谱需求意向的用户。可能当前实施条件不成熟，但是能够描述需求和愿景，并能体现出一种趋势需求。	需求说明书
9	需求评审人员	需求评审可能涉及的人员包括：需方的高层管理人员、中层管理人员、具体操作人员、IT主管、采购主管等；服务方的市场人员、需求分析人员、设计人员、测试人员、质量保证人员、实施人员、项目经理等；第三方的领域专家等。各方人员由于所处立场的不同，可能形成互补的观点。	业务需求说明书 需求规格说明书 系统原型说明书
10	政策专家	了解国家省市有关的政策法规的更新，及时掌握并传达有关政策的调整或变动，并对政策变化对各项目开发的影响进行评估，提出建设性建议。	技术规格书 遵守的法律法规
11	法务人员	处理及收集整理资料配合律师处理公司有关法律事务；收集、分析与公司业务相关法律信息并结合公司情况提出专业意见，针对工作中发现的问题及时提出预防措施；为公司提供咨询和法务意见书，提供用户及员工的法律问题咨询，负责制订公司的各类法律文件；成立健全公司知识产权和商业秘密保护体系。	技术规格书 权责规范 遵守的法律法规

序号	角色	承担任务简介	交付物名称（如有）
12	数据分析师	针对业务需求设计合理的数据分析框架和方案，并提供分析所得的意见；产出分析结果、数据问题和说明并提供解决方案。	数据样例 数据调研说明书 数据模型设计说明书 数据集规范
13	算法工程师	解决关于自然语言处理和分析的技术挑战和问题；在自然语言处理、分析和生成等技术领域结合实际场景开展先导性应用和领域创新研究。	语料库分析报告 数据集规范 可复用的通用图谱
14	软件测试工程师	按照知识图谱应用系统设计方案，编写测试用例方案，检查软件缺陷，测试软件的稳定性、安全性、易操作性等性能；对构建的知识图谱功能进行测试和文档检查，记录测试缺陷。	测试用例
15	供应商数据调研人员	负责评估提供数据的服务方，沟通数据集需求，决定数据源的选择，检查数据质量。	数据样例 数据规范 语料库分析报告 数据调研说明书
16	系统架构师	确认和评估系统需求，编制技术开发路线，负责对知识图谱应用系统整个软件架构、关键构件、接口的设计，可包括系统架构设计、组件规约、功能模块设计等；辅助需求分析师编制需求分析说明书。	技术规格书 备选策略分析 系统指标要求 部署环境要求

### （五）需用户方参与的内容

在需求分析阶段，用户方需配合参与的工作可包括：

#### （1）业务目标的明确

业务目标是企业构建知识图谱需要梳理的核心问题。制定正确、有

效、可执行的业务目标，是知识图谱应用系统建设成功并达到用户预期的基石之一。用户需和服务方在知识图谱需求分析过程中，给予持续的关注、评估和反馈，并围绕业务目标、应用场景和应用效果，不断迭代优化。

同时，用户需积极主动地表达出尽量全面和准确的需求，并配合需求人员的引导，逐步深入挖掘业务后续的隐藏需求。此外，应用场景和业务需求的边界、知识图谱要达到的目的及需求人员整理的需求规格说明文档，均需用户最终确认其准确性，以避免在知识图谱应用系统建设中修改业务范围与需求定义可能带来的项目风险及修改代价。

### (2) 外部系统和数据的协调

知识图谱应用系统需与数据对象访问服务、业务应用系统、数据存储系统、数据处理系统等进行对接。在项目设计与建设过程中，需充分考虑知识图谱应用系统与其他系统的功能边界、数据交互接口及与可能对接的外部系统，以确保能够满足业务场景需要，并同时兼顾未来的扩展需要，支持服务的横向扩展。

### (3) 部署环境的确认

知识图谱应用系统可能涉及公有云、私有云、混合云及本地私有化多种部署方式。企业用户需要根据自身需求确认部署方式，对于部署环境有特殊要求的，用户需及时提出并和服务方沟通，比如特定国产化环境和开放应用监控接口等的要求。

### (4) 提出应用符合安全标准

为了确保知识资产安全,需根据国家、行业或企业标准要求，用户方需对应用系统提出明确的安全规范要求，并从需求、设计、验收测试等环节对应用提出安全方面的评估内容。此外，用户还需参与安全方案的建设，可包括：数据隐私与安全、数据传输、存储加密等。

### (5) 应用系统用户验证

用户方需提前确定好系统的使用人员类型，在应用需求设计环节确认好需求，并在应用开发、测试后进行验证测试，确保应用的整体使用流程和功能设置符合预期。

## 二、方案设计阶段

### (一) 阶段建设目标

方案设计阶段是根据知识图谱需求分析阶段的输出内容，依据知识图谱应用系统设计原则与方法，给出相关详细设计成果物，为后续知识图谱构建阶段提供输入的过程。具体需要达成以下目标：

#### **目标1：基于业务需求，完成知识模型结构的定义。**

充分学习、掌握业务领域知识，结合用户需求，明确知识语料库范围，完成知识图谱本体模型的设计，确定支撑业务场景所需的概念、属性及其关系。同时，建立业务术语表，消除异构信息源中不同业务词汇间的歧义，确保认知的一致性。

#### **目标2：依据设计原则，完成系统方案设计。**

依托系统设计方法、遵循系统设计原则，以用户需求为核心，在功能层面，实现系统架构设计，涵盖应用架构、技术架构、部署架构；在数据层面，结合数据特征、数据现状，实现数据存储设计；在算法层面，考虑业务、技术约束，设计实现适合系统业务特点的算法。

#### **目标3：综合考虑系统现状，完成集成方案设计。**

围绕业务需求，结合当前存量系统及数据现状，设计系统集成方案。在应用层面，按需完成表现层的页面集成、服务层的接口集成、数据层的表集成；在知识层面，完成同构及异构知识图谱之间的集成方案设计。

**目标4：完成质量评估指标及方法的制定。**

根据知识建模方案及已构建的本体模型，设计适合的知识质量评估指标和评估方法，确保知识质量评估的完整性和准确性。同时，根据系统设计方案，完成系统主要测试用例的设计。

## （二）阶段输入输出

### （1）方案设计阶段输入

本阶段的输入可包括但不限于：

- 需求规格说明书；
- 预算文档；
- 行业调研报告；
- 数据规范；
- 知识图谱基础产品或服务；
- 知识图谱生态合作伙伴接入说明；

注：如在系统建设过程中引入的生态合作伙伴数据、产品、服务能力说明。

- 方案变更请求。

### （2）方案设计阶段输出

本阶段形成的交付物及内容简介如表7-3所示：

表7-3 方案设计阶段交付物清单

序号	交付物名称	简介	重要性 (1-5级)	是否 可选
1	语料库	构建知识图谱所需的领域资料，可包括但不限于标准规范、操作手册、业务指南、存量知识库等。	5	必须交付
2	本体设计文档	描述知识图谱应用系统在概念层面的设计文档，可包括知识图谱包含的实体类型、实体类型属性及实体类型间的语义关系等。	5	必须交付
3	术语表	知识建模过程中所需的核心业务术语及术语间的层次关系。	5	必须交付
4	系统概要设计文档	描述系统包含的关键抽象、核心架构机制、业务子系统、业务子系统之间的依赖以及职责分配。	5	必须交付
5	系统详细设计文档	可包括系统功能设计、数据设计及算法设计等。其中，功能设计描述业务子系统内部的设计实现，可包括边界类、控制类、实体类型的设计及伪代码实现。数据设计描述系统事务型数据的存储方案和知识图谱网络型数据的存储方案。算法设计描述知识图谱构建所需算法及其执行流程、相关参数、性能指标等。	4	推荐交付
6	系统设计原型	具有交互操作的系统页面高保真模型，可包括页面UI设计图、页面之间的链接操作等。	5	必须交付
7	系统部署文档	描述系统部署时的网络拓扑结构及所需的软硬件资源等。	4	必须交付
8	系统集成方案	可包括功能集成及知识集成。其中功能集成描述知识图谱应用系统与第三方系统的集成方案，如集成模式、所需接口服务、涉及的数据表和需开发或嵌入的页面等。知识集成方案描述知识图谱中知识融合的方案，可包括融合的数据范围，融合的策略，演化机制及质量控制。	5	必须交付
9	测试用例	描述系统功能和性能测试的场景和指标等。	3	可选交付
10	质量评估方案	包括系统功能、性能、安全及知识质量等内容的评估报告。知识质量评估可用于评价知识图谱包含的知识完整性、准确性、一致性等质量指标。	2	可选交付

### (三) 主要活动

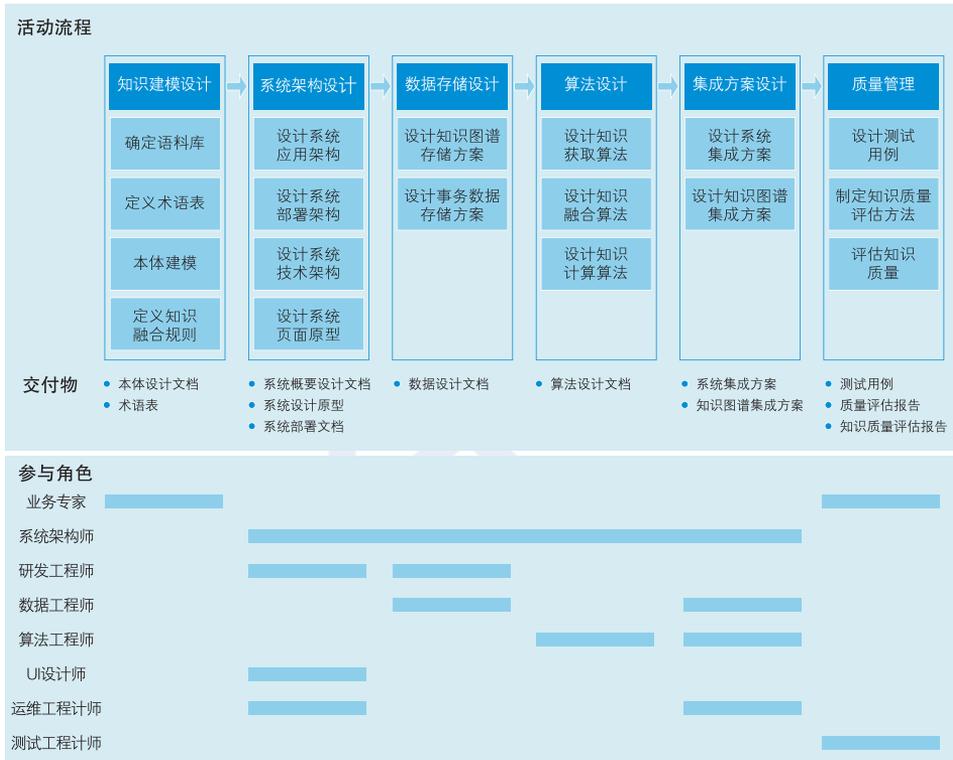


图7-3 方案设计阶段活动流程图

方案设计阶段根据拟实现目标的不同可划分为知识建模、系统架构设计、数据存储方案设计、算法设计、集成方案设计、质量管理等六项活动，本阶段活动流程及主要参与角色如图7-3所示。各项活动主要内容如下：

#### (1) 知识建模

**确定语料库：**选择可信范围内的领域知识数据，为后续活动中的知识提取提供质量保障。包括相应生产系统底层的业务数据库、相关业务操作

手册、指南文档、标准规范文档、存量知识库等。

**定义术语表：**选择或制定业务领域内的术语标准集。

**本体建模：**基于标准术语集，实现知识本体的建模，确定知识结构，可包括实体类型的识别、实体类型属性的识别及实体类型间关系的建模。此外，还需识别业务约束、规则，建立可推理的公理系统。

**定义知识融合规则：**制定同一知识体系及异构知识体系下，知识融合的人工规则，可包括实体类型的对齐规则、实体的对齐规则、属性的对齐规则及关系的对齐规则等。

### (2) 系统架构设计

**设计系统应用架构：**根据业务需求，按照工程方法实现系统的分层设计，分析和设计特定领域的业务组件及其交互协作接口，建立应用架构体系。此外，识别通用业务组件，降低依赖，并提高复用。

**设计系统技术架构：**根据业务需求、应用架构分析与设计结果，从业务组件中识别关键技术组件，完成其分析与设计任务，确定系统实现层面的技术路线及系统所需使用的中间件。最终，完成系统技术架构体系的构建。

**设计系统部署架构：**根据系统架构设计要求，设计系统在生产环境实施部署时所需的网络拓扑结构及相关的硬软件资源等基础设施配置。

**设计系统页面原型：**根据业务需求及系统设计约束，依据视觉和交互设计规范，设计具有可交互操作和高保真的系统原型页面。

### (3) 数据存储方案设计

设计知识图谱存储方案：包括选择数据模型、选择数据存储模型两部分。其中，a) 选择数据模型：结合业务需求及应用架构模式，选择或设计知识图谱的数据模型表示方式，如，RDF、属性图等。b) 选择数据存

存储模型：基于数据模型及业务需求，选择知识图谱的存储方法。

设计事务数据存储方案：有别于知识图谱存储方案设计，设计围绕系统功能特性的事务数据存储方案，可包括关系数据库业务表设计及用于性能提升的数据缓存方案设计等。

#### (4) 算法设计

**设计知识获取算法：**结合本体模型，针对非结构化、半结构化和结构化数据特点，设计相应的数据提取算法，如对非结构化数据采用基于深度学习的实体识别和关系抽取算法。

**设计知识融合算法：**基于本体模型及业务数据，设计和实现相同知识体系和异构知识体系下知识图谱中相关知识的融合算法，可包括本体层的实体类型、属性、关系的对齐算法，及实例层的实体、属性、关系的对齐算法。

**设计知识计算算法：**基于业务需求和应用场景，设计面向知识图谱的图挖掘及图推理算法，实现隐性知识的挖掘、知识图谱的补全和知识质量检测，并为基于知识图谱的上层知识应用，如智能推荐、语义检索、智能问答等提供底层技术支撑。

#### (5) 集成方案设计

**设计系统集成方案：**设计知识图谱应用系统业务功能层面的集成方案，定义知识图谱应用系统与第三方业务系统之间的集成方式（如：业务层集成、表现层集成或数据层集成）。同时，定义在不同集成方式下所需的集成资源，如：提供的接口服务、数据视图以及相关集成页面。

**设计知识图谱集成方案：**设计同构及异构知识图谱间的集成方案，可包括设计本体集成方案、设计实体集成方案、统一术语体系及知识表示方法、设计知识演化机制、定义集成质量要求等。

### (6) 质量管理

设计测试用例，如完成系统功能测试、性能测试及安全测试。同时，制定知识质量评估方法，可包括制定质量属性、制定评估方法等。其中：

a) 制定质量属性：定义和制定知识图谱质量评估指标体系，实现知识图谱知识质量的量化标准，可涉及知识的准确性、完整性、一致性、时效性和系统安全性、系统性能等方面。b) 制定评估方法：制定考虑不同领域和应用场景特点，切实可行的知识图谱质量评估方法，如基于人工的对比评估方法、基于算法的自动化一致性检测方法及半自动化的统计学抽样检测方法等。

评估知识图谱质量：根据制定的评估方法，对已建立的模型在本体层、实例层做人工或自动化的审核和校验，确保构建的本体和实体等能够客观真实地描述和反映领域业务知识。

### (四) 角色与分工

本阶段的参与人员角色及其任务简介如表7-4所示：

表7-4 方案设计阶段参与人员角色及其任务简介

知识图谱标准化  
Knowledge Graph Standardization

序号	角色	承担任务简介	交付物名称（如有）
1	业务专家	负责知识建模及知识质量评估，可包括确定语料库，定义术语表，完成本体建模，定义知识融合规则，制定知识质量评估方法及评估知识质量等。	语料库 本体设计文档 术语表 知识质量评估报告
2	系统架构师	负责系统设计全生命周期的把控，可包括设计系统应用架构，设计系统技术架构，设计系统集成架构，及评估数据存储设计、算法设计和集成方案设计过程中的相关活动。	系统概要设计文档 系统详细设计文档 系统集成方案
3	研发工程师	负责具体业务子系统的详细设计，可包括设计类、实现类、接口服务的设计及实体类型的设计。	系统详细设计文档
4	数据工程师	负责整体数据架构及存储设计，可包括事务型数据存储设计和知识图谱数据逻辑模型与物理模型的设计。	系统详细设计文档
5	算法工程师	负责知识图谱构建及集成过程中所需算法的设计和实现。	系统详细设计文档 系统集成方案
6	UI设计师	负责基于需求及系统设计约束，输出可交互的系统设计原型。	系统设计原型
7	运维工程师	负责系统部署架构的设计及相关配套资源的识别，并对系统集成方案中涉及的部署活动输出设计方案。	系统部署文档 系统集成方案
8	测试工程师	负责系统功能、性能、安全测试以及知识质量评估。	测试用例 质量评估报告

### （五）需用户方参与的内容

在方案设计阶段，用户方需配合参与的工作可包括：

#### （1）知识建模

用户业务人员需提供知识语料库或核实，确认知识语料库的范围，并参与术语的定义和评审。

#### （2）知识评估

用户业务人员需参与本体模型的评估，可通过抽查的方式，评估本体模型是否能够反映用户领域内的知识结构，是否能够准确表达领域知识内涵，是否能够支撑用户业务需求，以避免本体模型的设计错误。

### (3) 质量指标评审

用户业务人员需参与知识质量评估指标与评估方法的定义和评审，以降低评估指标的冗余，提高指标体系的完整性、专业性和准确性。

此外，还可考虑如下内容：

#### (1) 确认资源是否可支持设计的要求

通常情况下，服务方负责交付系统，而用户负责提供系统运行的环境资源。用户需求在设计过程中识别出对环境的配套要求会比较，并可能会超出用户的承受能力。此时，需设计人员及时与用户沟通，用户同样需参与讨论并及时反馈结果给设计人员，避免持续进行无法实现的设计甚至进入实施阶段，导致较大的资源和时间浪费。

#### (2) 确认设计的预期输出是否与需求匹配

设计过程中，功能输出的描述与用户的需求表述已存在较大的差异。而且，即便使用同样或类似的表达，在双方的语境里含义也可能不尽相同。此时，需要设计人员与用户进行充分的沟通，避免交付的并非需求所要求的内容，导致重复返工。

作为服务方，设计端的主动性通常是可预期的。但是作为用户方，参与上述活动必要性有可能难以被重视。因此，用户应注意在工作上主动参与沟通，确认设计的预期输出符合自身真实需求，避免沟通失误导致的浪费。

### 三、图谱构建阶段

#### (一) 阶段建设目标

知识图谱构建阶段是知识图谱应用系统建设的核心阶段，本阶段以业务需求为出发点，以业务数据资源为基础，以知识图谱的构建方法为指导，完成知识图谱从无到有的构建过程，最终形成知识图谱形式的业务数据存储。同时，提供知识的检索、推理和可视化等基础服务，为知识图谱的业务应用提供支撑。具体建设目标内容如下：

##### **目标1：完成业务专业术语体系的建设。**

根据业务范围参考专业术语相关标准、规范、数据集等，设计和构建业务专业术语词典及术语层级体系。

##### **目标2：完成符合业务需求的知识表示模型和本体模型的设计。**

依据本体理论结合业务需求确定知识表示形式，定义业务功能实现过程中宜遵循的业务规则及相关约束等，建立知识表示模型，定义知识图谱的本体模型，包括实体类型、关系、属性等。

##### **目标3：完成知识图谱的内容构建和质量评测。**

梳理和归集多种业务数据并进行数据治理，根据本体模型设计，结合知识抽取、知识融合等过程完成从业务数据到知识图谱内容的构建，完成知识图谱的存储，并依据评价标准体系对构建的知识图谱进行质量管控和评估。

##### **目标4：完成知识图谱应用接口的设计和开发。**

基于知识图谱的存储数据库查询语言，结合业务需求定义知识图谱的应用接口结构形式并开发接口。

##### **目标5：形成可持续的知识图谱构建流程和工具。**

基于知识图谱构建流程准则及相关的知识图谱子系统或工具，采用主

流程模板化和业务领域定制化相结合的方式，建立可持续的知识图谱构建流程和工具，便于用户后续其内部产品的知识图谱构建使用。

### 目标6：完成知识图谱运维保障机制建设。

通过建立知识图谱的维护、更新、管理等机制，保证知识图谱对系统业务功能支持的可用性和可靠性。

## （二）阶段输入输出

### （1）图谱构建阶段输入

本阶段的输入可包括但不限于：

- 业务需求；
- 业务数据；
- 辅助知识：包括已有的知识库、知识图谱等，内容为行业知识、常识、领域专业知识、专家资料等；
- 知识图谱构建保障工具、软件、技术包等。

图谱构建阶段的输入物及其简介如表7-5所示。

表7-5 图谱构建阶段输入物清单

序号	输入物名称	简介	重要性（1-5级）
1	业务数据资源	知识图谱构建相关的业务文档、数据库、已有知识图谱等各类数据。	5
2	业务需求文档	应用知识图谱预期实现的业务需求描述。	5
3	专家资料	业务领域专家提供的经验资料，如文档、典型数据案例等。	3
4	知识图谱存储数据库	用于存储知识图谱的数据库。	5
5	知识图谱构建工具包	包括构建知识图谱过程用到的各类工具、软件、技术包等。	4
6	知识图谱评价标准文档	用于知识图谱构建各活动质量评估的评价标准体系文档。	4

### (2) 图谱构建阶段输出

本阶段形成的交付物及内容简介如表7-6所示：

表7-6 图谱构建阶段交付物清单

序号	交付物名称	简介	重要性 (1-5级)	是否 可选
1	专业术语词典	满足业务需求的业务专业术语词典和术语层级关系表示。	4	推荐交付
2	本体模型	以图式 (Schema) 等形式描述的实体类型、属性、关系、约束、规则等模型文件。	5	必须交付
3	业务需求变更文档	知识图谱构建过程中进行的需求变更所形成的变更说明书、需求变更评审报告等。	2	可选交付
4	图谱算法模型设计文档	包含知识抽取、知识融合等相关算法模型设计内容的文档，及基于知识图谱的算法应用设计内容的文档。	4	必须交付
5	标注数据	用于支持知识抽取模型、知识融合模型等算法模型训练的标注数据。	4	必须交付
6	数据预处理工具	对输入的业务数据资源进行加工处理的工具，可提供数据清洗转换、数据标注、数据质量监控、数据生命周期管理等功能。	3	推荐交付
7	知识图谱	结合算法模型和人工干预等方式，基于业务数据资源构建完成的知识图谱。	5	必须交付
8	知识图谱管理工具	提高知识图谱构建、维护和分析效率的工具，可支持知识图谱的构建、维护、检索、可视化、探索等功能。	4	必须交付
9	知识图谱接口设计文档	符合业务功能需求设计的知识图谱接口文档，及运用数据库语言和开发语言编写的接口示例代码，可包括知识检索、推理、可视化接口等。	5	必须交付
10	知识图谱设计文档	指导知识图谱构建和相关模型、工具开发的设计文档，可包括本体模型设计、算法模型设计、数据标注（标签）设计、图谱管理工具设计、知识服务设计等内容。	4	必须交付

序号	交付物名称	简介	重要性 (1-5级)	是否 可选
11	研发进度记录	对系统整体研发进度的记录，为进度把控、按时完成系统交付物提供支持。	2	可选交付
12	版本控制文档	对知识图谱软件开发过程中各种程序代码、配置文件及说明文档等文件变更的管理控制文档。	1	可选交付
13	测试报告	对知识图谱构建过程中所用的相关模型和工具、接口等进行测试和联调后形成的测试报告；知识图谱的各个活动流程根据评价指标进行整体质量评测的质量评测报告。	4	必须交付
14	知识图谱使用说明文档	数据预处理工具、图谱构建与管理工具等工具或服务的使用操作说明、开发接口说明和实例代码等。	4	必须交付

### (三) 主要活动

知识图谱构建阶段是知识图谱应用系统开发生命周期中的重要和关键环节，根据拟实现目标的不同可划分为数据准备、知识表示与知识建模、知识获取与知识融合、知识图谱优化、知识应用服务等五项活动，本阶段活动流程及主要参与角色如图7-4所示。各项活动主要内容如下：

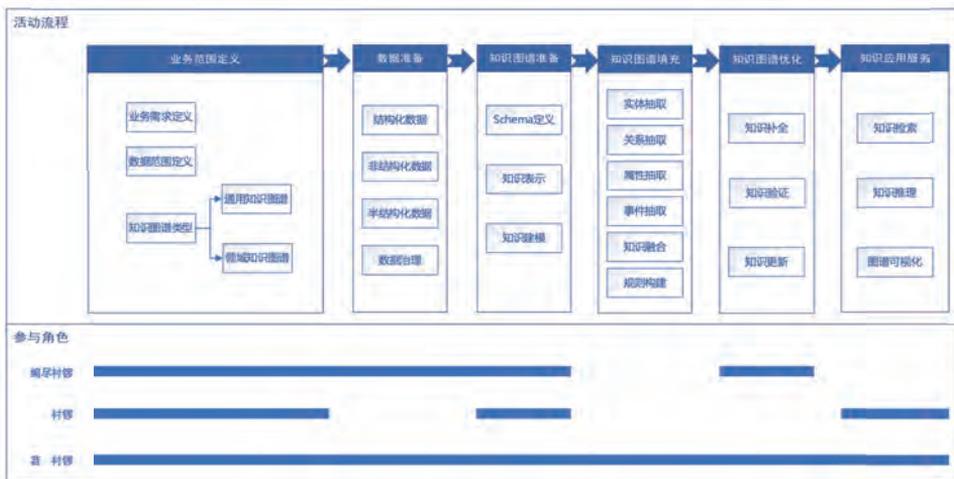


图7-4 知识图谱构建阶段主要活动及其参与角色

### (1) 数据准备

根据确定的业务数据范围，梳理领域内的重要业务术语，为知识图谱的构建提供数据资源，并进行数据治理，保证数据质量。主要数据形式包括结构化数据、非结构化数据、半结构化数据等。

通用知识图谱的数据范围较为广泛，一般以互联网开放数据为基础，数据规模可以逐步扩大，对数据的质量要求有一定的容忍度。

领域知识图谱的数据则以领域知识、领域业务数据等具体数据为主，内容包括领域内的知识库、术语集、规范资料等文本数据，以及业务实际生产数据，对数据质量的要求较高。

数据治理是通过一些处理方法将采集的原始数据进行数据预处理，提高数据质量，从而提升知识图谱构建的效率、准确性等。治理过程包括多源异构数据的融合，数据纠错、删除冗余、规格统一、补缺空值等一系列数据清洗操作。

### (2) 知识表示与知识建模

本活动是知识图谱数据构建的前提活动，包括定义知识表示的方式，进行知识建模，确定知识图谱构建的框架和工具，应用本体思想的建模方法设计知识图谱的数据结构，用来表达领域业务的知识体系，内容包括定义业务相关的概念、关系/属性以及数据类型、约束等，为知识图谱的构建实施做好准备。

### (3) 知识获取与知识融合

本活动是根据设计的知识图谱的数据结构，应用人工众包、算法模型等方式将经过处理后的业务数据进行知识获取与知识融合。知识获取可包括实体抽取、关系抽取、属性抽取、事件抽取等。知识融合是已有知识图谱与构建中的知识图谱进行知识体系映射和知识数据去除冗余、实现知识统一结构存储的过程。知识融合包括本体概念层的匹配与融合和实例层的匹配与融合。

### (4) 知识图谱优化

该活动中是对初步构建完成的知识图谱需进行知识补全和验证，同时在知识图谱维护时对知识进行更新。上述活动通过质量评估和持续维护以保证知识图谱提供知识的相对完备和准确性。

知识验证可分为两部分内容：一是对知识图谱中的实体、属性、关系的验证；二是对规则的验证。实体和关系的验证是研究知识图谱中知识单元集的可信度、一致性、准确性等，简化冗余的知识，并修正不正确的知识等。规则的验证是对构建于知识图谱中的规则或基于知识图谱建立的规则进行验证，如验证规则执行的正确性等。

知识补全是利用已有知识预测未知的隐含知识，用于完善现有知识图谱,常用的实现方法包括三元组分类和链接预测等。知识更新是保证知识图谱能够持续提供正确知识服务的实现手段之一。知识更新的内容为知识图谱全部知识单元，包括新的实体、关系、属性、规则等。从更新内容上，知识图谱的更新可包括本体层的更新和实例层的更新；从更新比例上，可分为增量更新和全量更新；从实现方式上，可分为人工维护更新和程序自动更新等方式。

### (5) 知识应用服务

该活动是根据功能应用的场景进行知识图谱的部署，通过知识图谱数据库的查询语言和查询语法提供知识检索、知识推理、知识可视化等接口和服务。其中，知识可视化是通过图形接口将知识图谱中的知识单元以可视化的形式提供应用服务，满足数据检索（子图）的可视化表达。

## (四) 角色与分工

本阶段的参与人员角色及其任务简介如表7-7所示：

表7-7 图谱构建阶段参与人员角色及其任务简介

序号	角色	承担任务简介	交付物名称（如有）
1	业务专家	提供专家知识，对设计的软件需求、知识图谱的模型和结果进行评审和审核，并参与部分知识图谱内容验证。	专业术语词典 本体模型 业务需求变更文档 知识图谱设计文档
2	专业业务人员	参与知识图谱的业务范围的定义，为知识图谱的知识建模设计提供业务支持，提供用于支持知识图谱构建的业务数据资源等内容。同时，参与知识图谱中知识更新和验证，对其他角色人员特别是标注人员进行业务培训。	专业术语词典 业务需求变更文档 知识图谱设计文档 知识图谱质量评测报告
3	产品经理	负责知识图谱产品的调研与需求分析，合理规划产品发展与功能规划。	业务需求变更文档 图谱使用说明文档
4	数据工程师	负责数据范围制定、数据治理等工作。	数据预处理工具 本体模型
5	标注人员	制定标注规范来标注数据。	标注数据
6	算法工程师	负责知识图谱构建阶段全流程的活动，完成知识图谱的设计和构建、算法模型的研发，及知识图谱基础应用接口和服务的开发，研发知识图谱相关常用算法，可包括知识获取、知识融合等图谱构建模型和知识图谱应用算法模型，并参与知识图谱运维管理工具的开发。	知识图谱设计 图谱算法模型设计文档 知识图谱接口设计文档 知识图谱 知识图谱运维管理工具
7	开发工程师	负责知识图谱应用系统开发，并参与知识图谱运维管理工具的开发。	知识图谱接口设计文档 知识图谱运维管理工具
8	质量工程师	对知识图谱应用系统研发过程进行质量管控，分析过程质量情况并对知识图谱及相关工具进行质量评估。	知识图谱质量评测报告
9	测试工程师	负责知识图谱构建过程中知识图谱及相关工具的测试工作。	测试报告

### （五）需用户方参与的内容

在图谱构建阶段，用户方需配合参与的工作可包括：

#### （1）知识图谱应用场景的确定和业务指导

知识图谱的需求源于应用场景中的需求描述，用户提供清晰的应用场景有助于需求人员确定需求范围，也方便算法人员从知识图谱的角度为这些场景中的需求进行可行性分析，同时可以界定知识图谱的构建所需的数据范围。另外，用户方一般具有专业的业务知识，特别是在垂直领域知识图谱构建过程中，专业知识的指导尤为重要。用户为知识图谱构建提供专业的业务指导，有助于提高知识图谱的构建效率和专业性。

#### （2）知识图谱数据资源的提供和评估

知识图谱的构建因涉及业务数据，用户需针对知识图谱服务方获取的数据资源进行评估，判断数据资源是否准确可用。另外，用户提供的业务数据资源更具有精准性和专业性，也是知识图谱构建的重要数据基础。

#### （3）对知识图谱新知识内容提供和运维支持

用户在知识图谱构建完成后，需对知识图谱和知识图谱应用系统进行运维和管理，并在有些领域和场景对知识图谱内容提出更新需求。此外，如果应用需求中的数据内容发生变化，用户可提供新的知识内容，辅助知识图谱服务方来完成知识图谱的知识更新和应用系统的迭代升级。

#### （4）审核构建阶段的输出

在图谱构建的各个阶段，用户需审核部分输出内容，主要包括原型系统、演示系统以及构建的知识图谱等。用户结合自身的需求，对输出内容审阅后，提出修改、优化意见，使得图谱构建人员可逐步优化图谱构建的细节，提高并保证最终的知识图谱质量。在图谱构建完成，需对交付内容进行验证测试，确保知识图谱构建整个流程及形成的知识图谱内容符合自身的预期。

## 四、应用开发与集成部署阶段

### （一）阶段建设目标

应用开发与集成部署阶段是在构建的知识图谱基础上，根据用户部署环境及现有系统的集成要求，开发满足业务需要的应用算法模型和应用功能模块，并将各个模块进行集成形成满足用户要求的知识图谱应用系统的过程。

#### 应用开发与集成部署阶段的建设目标主要包括：

**目标1：**从构建的知识图谱中以接口等方式获取知识单元或基础服务，并开发满足智能应用需求的应用算法模型，如：基于自然语言交互的业务知识问答、在线监测数据的风险预警、异常事件处置方案、相似事件推荐等。

**目标2：**根据部署环境和现有系统集成要求，遵循设计方案要求，开发系统功能模块，将业务系统的数据与知识图谱的知识相结合，为用户提供具有智能应用体验的完整统一系统。

**目标3：**对开发完成的知识图谱应用系统进行集成测试，并对照系统指标要求，利用真实业务数据，验证系统是否达到预期的应用要求。

**目标4：**根据系统环境要求，以敏捷开发方式和快速上线机制，提供系统自动部署服务。

**目标5：**为知识图谱应用系统建立运行监控机制，为系统运维管理提供必要服务请求的接口，以满足应用系统智能运维的需要。

### （二）阶段输入输出

#### （1）应用开发与集成部署阶段输入

本阶段的输入可包括但不限于：

- 系统设计方案；
- 系统业务模型；
- 本体模型；
- 系统集成质量要求；
- 构建的知识图谱。

### (2) 应用开发与集成部署阶段输出

本阶段形成的交付物及内容简介如表7-8所示：

表7-8 应用开发与集成部署阶段交付物清单

序号	交付物名称	简介	重要性 (1-5级)	是否可选
1	知识图谱应用系统	开发完成并集成知识图谱的系统	5	必须交付
2	系统使用相关的文档	可包括系统使用说明书、系统规格说明书等。	5	必须交付
3	系统开发相关文档	可包括需求变更报告、设计文档、API 文档、系统测试报告、缺陷记录跟踪表、验收报告等。	5	必须交付
4	系统部署相关文档	可包括系统部署方案、硬件设备清单等。	5	必须交付
5	系统监测相关文档	可包括系统监测报告、排错指南等。	1	可选交付

### (三) 主要活动

知识图谱应用系统开发的过程既要遵循软件开发的规范和流程，也要考虑如何在集成中充分利用已构建的知识图谱，满足业务智能应用中的人