

AIKBase 向量数据库管理系统生态建设方案

（试行）

一、背景与目标

KBase 数据库管理系统是知网自主研发，管理海量非结构化、半结构化数据，具备智能信息处理能力，以中文信息处理为特色的国产多模数据库管理系统。拥有国际领先的全文检索性能，支持并行计算机和集群部署，具有高可用性和高扩展性，具备相似检索、文本指纹、图片检索等丰富的特色功能。内核自主可控、全面适配国产 CPU、全面适配国产操作系统。

2023 年，面向 AI、大模型，知网基于 KBase 打造原生向量数据库系统 AIKBase，支持千亿级别的向量存储管理，实现了 KBase 对向量数据、标量数据、全文数据的混合查询，实现单机上，高维向量数据规模为十亿量级时，响应时间达到毫秒级别。

2025 年，知网转型为同方数科公司，以“服务科技创新，赋能数智转型，促进学术传播，承担社会责任”为使命，以成为人工智能和数据要素双轮驱动的世界一流数智科技公司为目标，支撑集团数字化转型，打造集团公司数字信息产业发展平台。AIKBase 向量数据库管理系统作为大模型的“智慧外脑”将推动知网大模型升级，助力知识服务智能化升级，为数字政务、数智企业、智慧教育、智慧医疗、智慧法治等场景提供解决方案。

因此，我们将打造 AIKBase 产业生态体系，开展广泛合作：联合高校、科研院所协同技术创新，促进科研成果转化，联合数据厂商、大模型厂商、整体解决方案集成商等合作开发产品、集成方案、共享市场，精准定位垂直行业应用场景，共同开发行业解决方案，协同提升产品竞争力与盈利能力。

二、产品价值

（一）赋能大模型

AIKBase 向量数据库弥补了大模型在知识实时性、长期记忆、事实准确性、隐私保护、成本效率和多模态处理等方面的关键短板。通过 RAG 等模式，使得 AI 大模型成为一个能够持续学习、精准可靠、理解深入且安全可控的智能系统，为检索、推荐、问答、对话、内容生产等应用场景带来质的飞跃。当前，RAG+向量数据库已成为当前解锁大模型在企业级应用中最实用、最主流的技术范式之一。

（二）规模用户市场基础

二十余年来，KBase 数据库管理系统作为知网总库的核心数据管理与检索平台面向全国 2000 余家高校、2000 余家图书馆、8000 余家企事业单位提供服务，为升级为 AIKBase 向量数据库管理系统奠定了坚实的市场基础，潜在用户规模可观，经济效益可期。

（三）知网原生生态优势

为保障知网中文资源总库的数据安全，同时提供最大用户价值，KBase 以总库平台建设的自身需求驱动，作为知网自主研发的搜索型数据库管理系统应运而生。知网近 3 亿篇

数据资源、数百个内容产品的生产、传播、安装、应用依托于 KBase 且不可分离,同时赋能数据应用场景完全契合科研、教育、技术革新、管理创新等知识服务与知识管理需求,并且通过 AIKBase 向量数据库管理系统支撑外部模型应用。

（四）系统构建向量数据库能力

AIKbase 向量数据库管理系统内置文本/图像向量化模型,提供文本、图像存储与检索时的内置向量生成功能特性,集成主流嵌入模型,实现即用的多模态数据检索功能,简化数据处理流程,解决了向量外部生成依赖,降低应用开发门槛,大幅度提升了业务对接能力。

三、产品特征

（一）自主可控，国产化适配

系统自主可控,全面适配主流国产化环境,支持鲲鹏、飞腾等国产 CPU,统信、麒麟等国产操作系统,系统运行稳定,性能表现优异,符合国家信创产业标准和要求。

（二）千亿规模数据存储与检索能力

AIKBase 支持亿级向量数据存储和毫秒级查询,向量查询性能接近甚至超越主流开源数据库。3000 万知网文献的篇名、摘要、小标题、全文分段等数据均应用 AIKBase 向量数据库构建数据集提供索引检索和文本生成能力。

（三）分布式集群能力

拥有超大规模数据的存储扩展能力和分布式查询能力;同时具备集群节点扩展管理与查询负载均衡机制,保障系统稳定高效运行。

（四）特色增强检索

混合增强检索能力：支持向量与文本、标量的混合检索，满足复杂业务场景下的多样化搜索需求，提升搜索的准确性和实用性。

多向量字段存储和检索能力：单字段可同时容纳多个不同向量数据，支持这些向量的高效检索操作，通过整合多向量数据简化数据架构、提升查询效率与精度，为复杂业务场景下高效数据处理与决策提供关键支撑。

（五）多元场景适配

多重索引机制：当前系统提供高性能内存向量索引与低内存向量索引的双重解决方案，用户可根据实际业务需求灵活选择，实现性能、成本与精度的最佳平衡。高性能内存索引，适用对查询延迟敏感、要求极致检索精度的关键业务场景；低内存索引，适用资源受限环境或对成本敏感的业务场景。

高性能存储引擎：支持高并发写入与数据实时更新，满足数据高频写入场景，避免高频写入时的稳定性问题，且采用可插拔式存储引擎设计，灵活适配不同应用场景。

四、合作模式

（一）大模型增强检索可信溯源应用

以“开源\闭源大模型+向量数据库”为基础方案，面向集成商、大模型厂商、数据库厂商广泛开展合作，集成双方能力，共享销售渠道与用户，增强竞争力，拓展市场。

1. 合作价值

一方面在优势市场整合知网品牌优势；另一方面在整体解决方案中优势互补，增强多模态数据库能力，并实现国产化替代，降低研发、销售成本。

2. 合作路径

合作伙伴以渠道价格从我司“进货”，将 AIKBase 整合进销售方案或直接代理产品项目投标。

（二）向量数据库赋能垂直领域应用

面向出版社、行业协会、行业数据中心等行业数据、原创资源版权方提供向量数据生成、存储、检索、应用服务。双方共享下手渠道与用户，共同运营产品。

1. 合作价值

通过将特色数据产品化或提供数据服务在垂直领域的专业应用场景下实现数据变现，提升数据附加值以及品牌效应。

2. 合作路径

双方形成“数据+平台”的向量数据库产品或数据管理、分析服务，面向终端用户销售推广。具体可能包括：License 授权、OEM、定制开发等。产品和服务进入双方渠道共同销售。

（三）知网集成，发展供应链

面向终端具有数据采集、加工、存储、应用、分析等数据需求的机构用户，知网提供一站式解决方案，包括软硬件集成方案、加工方案、数据平台建设方案等，提供售前、售中、售后服务，保障用户得到最优解决方案。在这个过程中，由知网集成多家产品与服务，即该类合作的对象为以 KBase

数据库管理系统为核心的上中下游供应商。

1. 合作价值

在知网优势市场，例如高校、图书馆等通过方案集成，拓展市场以及提升产品粘性和竞争力，同时提升品牌影响力。

2. 合作路径

知网与供应商逐一签订框架协议进行产品与服务采购，按约定在协议范围内进行项目投标，签单。