

Flexus 企业搜索服务

产品介绍

文档版本 01
发布日期 2024-08-27



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为云计算技术有限公司

地址：贵州省贵安新区黔中大道交兴功路华为云数据中心 邮编：550029

网址：<https://www.huaweicloud.com/>

目录

[1 什么是 Flexus 企业搜索.....1](#)

1 什么是 Flexus 企业搜索

什么是 KooSearch

华为云企业搜索KooSearch是基于华为云的云搜索服务搭建的一站式智能搜索解决方案，帮助企业聚焦业务场景和应用开发，场景服务化、技术简单化、低门槛化，满足开发者基于业务场景的二次开发。在RAG（Retrieval-Augmented Generation）及搜索场景提供效果和性能突出的组件化服务，架构理想、灵活编排的机制，帮助企业客户快速构建RAG以及搜索服务。

什么是 Flexus 企业搜索

基于业界效果突出的搜索模型、云搜索服务的高性能GaussDB向量数据库以及LLM模型组成行业RAG解决方案，导入非结构化或者结构化业务数据，帮助企业客户快速一站式搭建企业智能问答应用。搜索增强大模型，数据来源于搜索，解决大模型幻觉问题，生成的问答结果更可靠、安全。

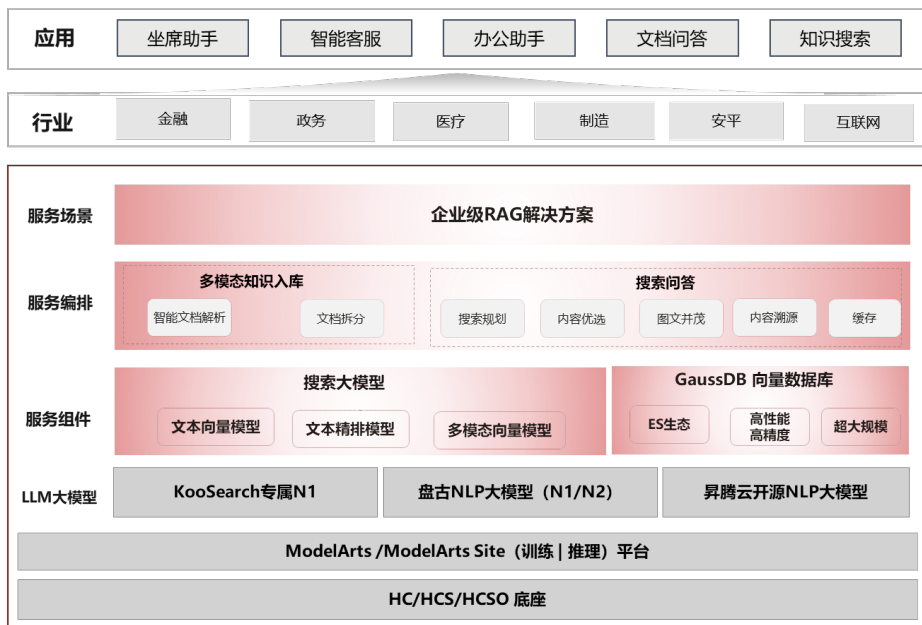
产品功能

- 开箱即用
导入非结构化或者结构化业务数据，帮助企业客户快速一站式搭建企业智能问答应用。
- 效果突出的搜索模型
文本向量模型在中文语义向量评测基准C-MTEB效果突出；精排模型对召回结果重排序进一步提升TopK准确率；搜索规划模型实现Query分类、多轮改写、复杂查询分解。
- 丰富的LLM（Large Language Model）模型
内置基于RAG场景下SFT（Supervised Fine-tuning）优化后的检索增强大模型，支持对接盘古NLP（Natural Language Processing）大模型，支持对接晟腾云开源大模型。
- 智能解析切分
通过OCR（Optical Character Recognition）增强准确识别图片/表格等各式版面布局、配置开放、拆分优化等措施提升文档解析效果。
- 多模态输入和输出
基于业务数据生成问答结果，可溯源，输出结果图文并茂。
- 物理多租、租户隔离、全托管

支持物理多租、租户隔离、全托管服务，独享资源更稳定、性能更高。

产品架构

图 1-1 KooSearch 架构



KooSearch服务开箱即用，帮助企业客户快速构建企业级RAG解决方案。

- **多模态知识入库**
 - 智能文档解析
用于文档内容解析，支持PDF、JPG、DOC、TXT、XLS、PPT等文档格式，通过OCR增强功能准确识别图片、表格等各式版面布局。
 - 文档切分
支持多种拆分方式：自动拆分、层次拆分、长度拆分和自定义规则拆分。
- **搜索问答**
 - 搜索规划
利用LLM技术的Query分类、多轮改写、复杂查询分解。
 - 图文并茂&内容溯源
支持输出参考链接，和上下文语义相关的图片。
 - 内容优选
通过query和片段的语义关系，精简出和query相关的web片段，达到减少下游大模型推理成本，以及提升大模型的推理效果。
 - 缓存
利用缓存技术，预先存储频繁请求的信息，系统从缓存中快速提供，节省时间和资源。
- **搜索大模型**
 - 文本向量模型
支持将文本转化为用数值表示的向量形式用于语义检索。

- 文本精排模型
可更加精准找到相关文本，不同于向量模型输出向量，直接计算query和doc的相关性更加准确，可用于对向量召回结果的重新排序。
- **GuassDB向量数据库**
基于Elasticsearch的华为云自研向量检索引擎。

访问方式

公有云提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）管理方式。

- **API方式**
如果用户需要将公有云平台上的集成到第三方系统，用于二次开发，请使用API方式访问，具体操作请参见《[Flexus企业搜索服务API参考](#)》。
- **控制台方式**
其他相关操作，请使用管理控制台方式访问。如果用户已注册公有云，可直接登录管理控制台，在服务列表搜索“云搜索服务”。如果未注册，请单击右上方“注册”，根据界面提示填写用户基本信息，然后单击“同意协议并注册”完成管理控制台用户注册。