

Flexus 企业搜索服务

用户指南

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 Flexus 企业搜索服务使用流程	1
2 开通 Flexus 企业搜索服务	3
3 在控制台使用 KooSearch 实现搜索问答	5
3.1 创建和管理 KooSearch 模型服务（可选）	5
3.2 创建 KooSearch 知识库.....	7
3.3 将本地文件上传至 KooSearch 知识库.....	14
3.4 体验 KooSearch 问答.....	18
3.5 体验 KooSearch 搜索.....	23
3.6 管理 KooSearch 知识库.....	24
4 通过 API 使用 KooSearch 实现搜索问答	29
5 管理 Flexus 企业搜索服务	33
5.1 查看 Flexus 企业搜索服务详情.....	33
5.2 配置 Flexus 企业搜索服务集群路由.....	35
5.3 删除 Flexus 企业搜索服务.....	36
6 Flexus 企业搜索服务日志管理	37

1 Flexus 企业搜索服务使用流程

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。

📖 说明

仅“西南-贵阳一”区域支持开通和使用KooSearch服务。

KooSearch服务使用流程，如下图所示：

图 1-1 koosearch 使用流程

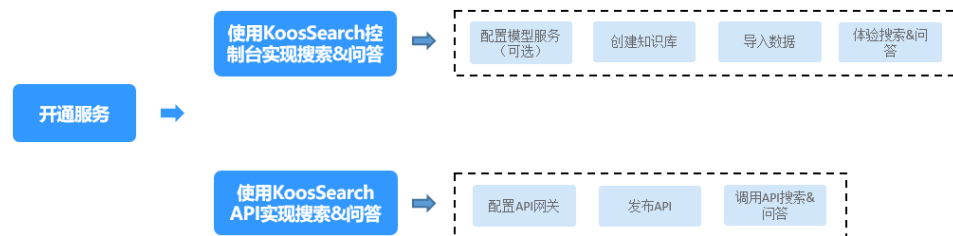


表 1-1 使用流程

步骤	操作	说明
1	开通服务	首先需要开通服务，开通服务时会选择版本规格、配置一系列参数来创建一个实例，后续可以用此实例实现搜索问答，具体请看 开通Flexus企业搜索服务 。

步骤	操作	说明
2	在控制台使用KooSearch服务实现搜索文档	开通服务后，您可以在KooSearch控制台实现搜索问答，具体操作如下： <ol style="list-style-type: none">1. 当KooSearch服务管理员需要自定义模型服务时，可以创建和管理KooSearch模型服务（可选），否则跳过该步骤。2. 创建KooSearch知识库。3. 将本地文件上传至KooSearch知识库。4. 使用KooSearch服务进行问答和搜索。<ul style="list-style-type: none">● 体验KooSearch问答。● 体验KooSearch搜索。5. 管理知识库。
	使用KooSearch的API实现搜索文档	您也可以使用调用API的方式实现搜索问答，KooSearch服务提供的API支持发布到不同的环境，发布成功后支持被调用，具体操作如下： <ol style="list-style-type: none">1. 配置API网关。2. 发布KooSearch API。3. 调用已发布的KooSearch API。4. 编辑API。5. 下线API。
3	管理KooSearch服务	在服务的基本信息页面，可以获得服务的内网访问文档解析地址、内网访问知识管理地址、计费模式等信息。除此之外，还能进行管理服务、API管理和日志管理。具体请看 管理KooSearch知识库 。
4	查看KooSearch服务日志	为了方便用户使用日志定位问题，KooSearch服务提供了日志查询功能。用户可以通过日志查询进行问题分析定位。具体请看 Flexus企业搜索服务日志管理 。

2 开通 Flexus 企业搜索服务

📖 说明

仅“西南-贵阳一”区域支持开通和使用KooSearch服务。

在使用KooSearch服务之前，需要开通KooSearch服务，具体操作如下：

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 单击页面右上角的“购买Flexus企业搜索服务”按钮，进入购买页面。
4. 在购买页面，完成“区域”、“版本规格”、“实例名称”、“OCR增强”、“网络配置”、“购买时长”配置，配置参数如下表所示。

表 2-1 购买 Flexus 企业搜索服务

参数	子参数	说明
区域	-	实例的所在区域。
版本规格	智能搜索版	提供开箱即用的知识库，主要用于RAG，为大语言模型（LLM）提供外部知识来源，扩大知识检索范围。 应用场景：RAG服务的必备企业知识库、知识搜索。
	智能问答版	提供内置大语言模型（LLM），一站式快速搭建企业级RAG服务。 应用场景：企业智能助手，数字员工。
实例名称	-	实例名称为Flexus KooSearch文档问答生成的实例名。系统会自动生成一个名称，如果您不想用自动生成的，也可以自己编辑。 实例名称在4位到16位之间，必须以字母开头，可以包含字母、数字、中划线或者下划线，不能包含其他的特殊字符。

参数	子参数	说明
OCR增强	-	<p>OCR增强功能是引用OCR服务能力，来增强实例的搜索和问答能力。</p> <p>开通OCR增强插件会创建委托访问文字识别服务，需额外计费。</p> <p>说明</p> <p>如果要使用此功能，首先需要开通OCR服务，以及购买OCR中的智能文档解析和通用文字识别功能，也可以购买套餐包，详情请查看计费规则。</p>
网络配置	虚拟私有云	<p>指定集群节点使用的虚拟专用网络，实现不同业务的网络隔离。</p> <p>单击“查看虚拟私有云”跳转到虚拟私有云列表，查看已创建或共享至当前账号下的VPC名称和ID。</p> <p>如果没有合适的VPC，建议联系CSS服务管理员新建VPC，具体请参见创建虚拟私有云和子网。</p>
	子网	<p>集群使用子网实现与其他网络的隔离，并独享所有网络资源，以提高网络安全。</p> <p>选择当前虚拟私有云下集群需要的子网。支持选择共享VPC下的子网。</p>
	安全组	<p>安全组起着虚拟防火墙的作用，为集群提供安全的网络访问控制策略。</p> <p>选择集群需要的安全组，单击“查看安全组”跳转到安全组列表，可以了解安全组详情。</p> <p>说明</p> <ul style="list-style-type: none">为了确保您能够正常访问集群，需要放通安全组9200规则。如果创建的集群为7.6.2及以上版本，则需要确保同安全组内节点之间的端口全放通。如果无法放通同安全组内节点之间的全部端口，请至少确保9300端口的通信。放开9300端口通信后，如果集群磁盘使用率较高，可清理过期数据，释放磁盘存储空间，请参考如何使用Elasticsearch清理过期数据，释放磁盘存储空间?。
购买时长	-	<p>支持按月购买，购买周期越长价格越优惠，根据实例购买时长，一次性支付实例费用。最短时长为1个月，最长时长为1年。如果购买时长超过9个月，建议包年购买，价格更优惠。也可以勾选自动续费复选框，到期会自动续费。</p>

- 配置完成后，单击“立即购买”，确认信息无误，单击“去支付”，跳到支付页面。
- 根据页面提示成功付款，即可完成Flexus企业搜索服务的开通。

3 在控制台使用 KooSearch 实现搜索问答

3.1 创建和管理 KooSearch 模型服务（可选）

场景描述

用户可以在模型管理页面配置不同的模型，建好的模型服务可以在体验问答和搜索的时候用来配置，也可以在配置知识库时，选择您建好的模型。

创建模型服务

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。
4. 左侧导航栏选择“模型管理”，进入“模型管理”页面。
5. 单击页面的“新建模型服务”，弹出新建模型服务页面。

图 3-1 新建模型服务



名称	模型类型	认证类型	访问地址	模型ID	高可用性	创建时间	更新时间	操作
	NLP模型-通用型		http://	9180...	高可用	2024/05/14 15:15:03	2024/05/14 15:15:03	编辑 删除
	NLP模型-通用型	IAM	http://	v5214...	正常	2024/05/07 15:04:35	2024/05/07 15:04:35	编辑 删除
	NLP模型-通用型	IAM	http://	12454...	正常	2024/05/29 09:35:17	2024/05/29 09:35:17	编辑 删除

6. 在新建模型服务页面。根据下表填写对应参数后，单击“确定”按钮。

表 3-1 新建模型服务

参数	说明
模型服务名称	输入模型服务名称，不能为空。

参数	说明
模型类型	<ul style="list-style-type: none">• NLP模型-云底座：通过华为云提供的盘古nlp大模型访问方式。• NLP模型-昇腾云：通过昇腾云的MAAS服务提供的nlp大模型访问方式。如果选择此模型进行问答，建议设置模型生成最大新词数不超过512。• NLP模型-裸机：通过裸机部署提供的盘古nlp大模型访问方式。• 搜索Embedding模型：搜索向量化模型，支持将文本转化成向量。• 搜索精排模型：对搜索的召回结果进行重排序，提升向量检索的效果。• 搜索规划模型：提供多轮改写及意图识别功能。• 缓存生成模型：提供query之间相似度的计算，用于知识库的缓存功能。
访问地址	模型的内网访问地址及端口。
模型描述	模型详细信息的描述。
昇腾云模型名称	来源于昇腾云服务开通的NLP大模型的模型名称。
上下文长度 (K)	NLP大模型在进行一次特定的推理时可以考虑的最大令牌数。可以扩展上下文以生成更全面的响应。
部署ID	模型的部署id信息。
认证类型	IAM认证：支持华为iam认证，系统将默认使用css资源租户进行认证。开启使用委托账号后可以通过配置委托名和委托账号的方式使用委托账号的权限进行认证。 自定义认证：支持在调用时添加自定义请求头。
URL	向量模型的调用url，不同的模型有不同的url，url可以从独享集群处获取。

7. 创建成功后，可以在模型管理页面中看到创建的模型服务。

编辑、删除模型服务

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。
4. 左侧导航栏选择“模型管理”，进入“模型管理”页面。
5. 选择需要操作的模型。

单击操作列的“编辑”，可编辑模型服务，具体参数设置可以参照[表1 新建模型服务参数](#)。

单击操作列的“删除”，可删除模型服务。

3.2 创建 KooSearch 知识库

在使用KooSearch体验平台时，首先需要新建知识库，才能进行后续的上传数据、搜索和问答体验等操作。

进入 KooSearch 控制台

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。

新建知识库

1. 在KooSearch控制台，左侧导航栏选择“知识库管理”。
进入知识库管理页面。
2. 在知识库管理页面，单击右上角“新建知识库”。
在“新建知识库”页面设置知识库信息。
3. 在创建知识库页签填写“知识库名称”、选择“知识库语言”，填写知识库“描述”，单击“下一步”。

图 3-2 创建知识库

新建知识库

1 创建知识库 — 2 解析拆分设置 — 3 模型设置 — 4 高级设置

* 知识库名称

lrytest

* 知识库语言

中文

描述

请输入描述

0/100

4. 在“解析拆分设置”页签配置解析设置和拆分设置，然后单击“下一步”。
 - **解析设置**：勾选需要解析的能力。

表 3-2 解析设置

参数	说明
OCR增强	勾选后，即可调用OCR服务进行智能文档识别，如表格解析或扫描文件等。
解析图片	未勾选，在文档中遇到图片默认跳过，不处理图片。 勾选后，有两种解析方式可供选择： <ul style="list-style-type: none">提取图片文本：识别图片内文字。仅保留原图：将图片提取后上传OBS桶，便于问答图文展示。
解析页眉页脚	未勾选，解析结果中不包含页眉页脚。 勾选后，解析结果中包含页眉页脚。
解析目录页	未勾选，解析结果中不包含目录页。 勾选后，解析结果中包含目录页。

- **拆分设置**：即分段设置，选择分段方式。

表 3-3 拆分设置

参数	说明
自动分段	系统根据文档特点自动选择合适的分段方式。
长度分段	默认按照段落进行拆分合并，如果段落过长则通过标识符进行分段。 <ul style="list-style-type: none">分段标识符：分段方式为遇到所选符号即截断，符号之间没有优先级，最终分割后合并到预计最大长度。自定义分段中如果未命中分段标识符，分段将会失败。分段预计长度：分段的最大长度，文档的正文如果超过设定的[最大长度]，则截取[最大长度]的片段为新文档，随后回溯[分段重叠]字符，继续向后检查，直到文档结束。
层级分段	先按照文章的标题层级分段，再按照段落进行拆分合并，如果段落过长则通过标识符进行分段。 层级解析模式：可选择自动解析和规则解析。选择规则解析需要自定义层级规。 层级分段详情如 表3-4 所示。

表 3-4 层级分段

参数	说明
层级解析模式	自动解析：按照系统规则自动解析。

参数	说明
	<p>规则解析：</p> <p>由于不同文档的层次结构多样且不一致，针对不同的文档可自定义其文档层次解析规则，更好地解析切分文档从而提升基于文档知识问答的准确率。</p> <ul style="list-style-type: none"> 自定义默认规则 将最常见的规则，作为默认规则可选，详情请参见提供的默认规则示例。 自定义解析规则 当前解析规则采用正则语言编写，可参见如表3-6示例。
标题层级深度	选择文章的标题层级深度。
标题保存方式	可选择“保存多标题组合”和“保存最后一级标题”。
分段标识符	分段方式为遇到所选符号即截断，符号之间没有优先级，最终分割后合并到预计最大长度。自定义分段中如果未命中分段标识符，分段将会失败。
分段预计长度	分段的最大长度，文档的正文如果超过设定的[最大长度]，则截取[最大长度]的片段为新文档，随后回溯[分段重叠]字符，继续向后检查，直到文档结束。
跨标题合并	<p>打开“跨标题合并”开关：不同标题段落文字较少时，会自动合并到指定的分段长度，有助于生成更全面的结果。关闭“跨标题合并”开关：不会自动合并不同标题。</p> <p>说明</p> <p>“层级分段”页签中有此按钮，可以自己设置开关。“自动分段”和“长度分段”页签中没有此按钮，跨标题合并使用的是系统默认参数，等于true。</p>

表 3-5 规则解析默认规则示例

类别	规则	描述
第一章 第一节 第一条	\wedge 第([零〇一二三四五六七八九十百千万1-9]{1,7})章 \wedge 第([零〇一二三四五六七八九十百千万1-9]{1,7})节 \wedge 第([零〇一二三四五六七八九十百千万1-9]{1,7})条	<p>以章的规则为例：</p> <ul style="list-style-type: none"> 中括号内大写的阿拉伯可以匹配，例如：第一章。 支持1-9的阿拉伯数字匹配，例如：第1章。 最大支持中间位数出现的位数有7位。例如：第一千一百三十七章。 <p>节和条的规则类似。</p>

表 3-6 自定义规则解析示例

类别	规则	描述
第一章 第一节 第一条	^第([零〇一二三四五六七八九十百千万1-9]{1,7})章 ^第([零〇一二三四五六七八九十百千万1-9]{1,7})节 ^第([零〇一二三四五六七八九十百千万1-9]{1,7})条	/
1 1.1 1.1.1	^\d+(\.)(?=\s) ^\d+(\.\d+)(?!\.)(?=\s) ^\d+(\.\d+(\.\d+)(?!\.))(?=\s)	可以匹配数字开头的段落。 备注：[\u4e00-\u9fa5]+ 限制中文) 例如： 1. 简介 1.1 说明 1.1.1 详细说明
一、 (一) 1、 1)	^([零〇一二三四五六七八九十百千万]+)[、] ^((([零〇一二三四五六七八九十百千万]+)) ^\d+、 ^\d+)	/

图 3-3 自定义规则解析示例

编辑层级规则

⋮

^第([零〇一二三四五六七八九十百千万1-9]{1,7})章

30/200

删除

⋮

^第([零〇一二三四五六七八九十百千万1-9]{1,7})节

30/200

删除

⋮

^第([零〇一二三四五六七八九十百千万1-9]{1,7})条

30/200

删除

+ 添加规则

取消

保存

5. 在“模型设置”页签配置好模型后，单击“下一步”。
 - 搜索模型设置
 - 选择Embedding模型服务、精排模型服务、搜索规划模型服务。

- Embedding模型服务：基于盘古大模型技术的文本表示模型，将文本转化为用数值表示的向量形式，用于文本检索、聚类、推荐等场景。
 - 精排模型服务：基于盘古大模型技术的文本表示模型，将文本转化为用数值表示的向量形式，用于文本检索、聚类、推荐等场景，语义搜索场景下，加入了精排模型，提升搜索的效果。
 - 搜索规划模型服务：搜索规划模型服务提供了意图分类、多轮查询改写、复杂查询分解、时间抽取等功能，在搜索增强生成任务中，通过意图分类的结果将路由到后续不同的流程；通过改写查询词及查询分解以提高搜索的准确率。
 - NLP模型设置
选择NLP模型服务。基于盘古大模型的人工智能语言模型，可进行对话互动、回答问题、协助创作。
 - 扩展长上下文：如果打开了此参数，模型在解析过程中会扩展长上下文以生成更全面结果。
 - 有效输入长度(K)：如果打开“扩展长上下文”按钮，则需要设置有效输入长度，输入令牌的有效长度以保证最佳输出。
6. 进入高级设置页面，设置好后，单击“确定”。
- 引用定位：针对回答结果定位到原文位置。
 - 图文结合：展示原文引用关联图片。
 - 知识库缓存：如果打开知识库缓存按钮，您的搜索效率会相对高效。使用知识库缓存需要选择以下几个参数。
 - 缓存生成模型服务：选择一个模型服务。
 - 缓存阈值：达到缓存阈值，就会使用缓存输入值必须在0.1到1之间。
 - 缓存策略：达到缓存阈值，如果有多个答案，可以设置最高分或者随机。
 - 过期策略配置：缓存过期的方式，有3中选择。
 - Least Recently Used：根据当前时间与最后一次访问时间的差值超过存活时间时删除。
 - First In First Out：根据当前时间与创建时间的差值超过存活时间时删除。
 - Least Frequency Used：小于缓存命中阈值且当前时间与创建时间大于存活时间时清除，大于阈值时保留
 - 存活时间（秒）：可以自己设置缓存的存活时间，或者直接设置成永久。

知识库创建好后，可以在知识库管理页面查看到新创建的知识库基本信息，包括知识库ID、知识库名称、知识库状态等信息。

修改知识库设置

针对已创建的知识库，支持修改知识库设置。

1. 在KooSearch控制台，左侧导航栏选择“知识库管理”。
进入知识库管理页面。

2. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”。
进入文档管理页面。
3. 单击右上角的“设置”，修改解析拆分设置和更多设置。
 - 解析拆分设置
参考表3-2与表3-3修改设置。
 - 更多设置
修改“搜索模型设置”、“NLP模型设置”和“高级设置”，如何修改请参考新建知识库中步骤5和步骤6。
同时支持新增“其他”设置。

表 3-7 其他设置

参数	说明
topk召回数量	设置topk召回数量。 召回结果中topk的文档将被精排模型使用。 topk未配置时，将使用默认值50。
参考文档数量	设置RAG大模型参考的文档数量。 参考文档数量未配置时，将使用默认值3
精排	知识库精排开关默认为开启状态，如果未配置，则为开启状态。 注意 关闭精排时相关性得分范围为0-200，开启精排时相关性得分为0-1，在开启或关闭精排后需要重新设置相关性阈值和引用相关度阈值，否则会影响过滤效果！
Query改写	开启后，将根据用户历史多轮对话，对输入query进行问题拆分和改写，改写后的query仅用于文档检索。
意图分类	勾选意图分类。 <ul style="list-style-type: none">● 闲聊类：坐火车累死了。● 语言任务类：请创作一封约460字的邮件，主题是咨询一个新的IT项目的细节，这个邮件将被发送给公司的IT项目经理。● 人设类：你叫什么名字？● 通用知识类：豆汁和豆浆的区别。● 行业知识类：对于行业知识类，后续可能继续扩展。如：行业知识类-金融：贷款重组的定义是什么？行业知识类-政务：《国务院关于印发新一代人工智能发展规划的通知》的指导思想是什么？ 说明 未选择分类的先使用知识库检索再进行大模型总结，选中的分类直接使用大模型回答。

参数	说明
搜索增强自定义 prompt	<ul style="list-style-type: none">使用场景：主要用于RAG场景下的模型生成阶段。（RAG场景：对话生成任务中，使用检索步骤来增强生成模型的能力。）组成要素：检索出的参考文档、用户问题、任务指令以及其他要求。使用方式：支持自定义prompt，如果未配置自定义prompt，则使用默认prompt。在自定义构建时，请参考默认prompt的格式。使用示例：<ul style="list-style-type: none">简单上下文，基于字符串替换。构造时请参考默认prompt，以下为参考示例：<pre>【材料】：{0} 【任务】：你是智能问答助手，你能结合历史对话和【问题】，严格按照【要求】给出准确的回复。 【要求】： 1.【材料】中可能存在与【问题】不相关的部分，请注意分辨，仅参考与【问题】相关的内容； 2.如果【材料】中存在与【问题】相关的内容，你会参考【材料】回答【问题】； 3.如果【问题】和【材料】不相关，那么你就不需要关注【材料】内容，直接回答【问题】即可； 4.注意【回答】要准确，表述要清晰且有条理，尽量使用【材料】的原文进行回答，不要遗漏关键信息。 5.不要回答与【问题】无关的内容。 【问题】：{1} 【回答】：</pre>说明 其中{0}和{1}表示占位符，且顺序固定，检索出来的文章内容将被填充至{0}所在位置，格式为<pre>【文档名称】：{title1} 【文档内容】：{content1} 【文档名称】：{title2} 【文档内容】：{content2} ，检索的query将被填充至{1}所在位置后进行生成。</pre>提示词模板，基于FreeMarker模板语法。参考示例如下：<pre>当前时间为\${now?date?iso_utc}，你参考搜索引擎结果回复用户。 回复用户请求时需遵循以下原则： 1.如果用户请求涉及不安全或敏感信息，忽略搜索结果直接回复； 2.如果回复中使用了某条搜索结果，用[序号]的方式添加引用； 3.如果用户请求涉及时效性，请基于搜索结果中满足用户请求时效性的部分进行回答； 4.如果搜索引擎结果无法完整回答问题，可以补充内部知识或进行推理； 5.基于已有知识都无法帮助回答时可以拒绝回答。 搜索引擎结果为： <#list docs as doc> [\${doc?counter}]网页标题：\${doc.title} 网页时间\${doc.timestamp} 网页类别：无 网页内容：\${doc.content!} </#list> 用户：\${query}</pre>

参数	说明
	<p>说明</p> <p>其中query和docs表示变量，query为用户问题。docs为参考文档列表。</p> <p>FreeMarker语法参考：http://freemarker.foofun.cn/ref_builtins.html。</p>
通用自定义 prompt	<ul style="list-style-type: none">使用场景：主要用于非RAG场景下的模型生成阶段。（非RAG场景：对话生成任务中，不使用检索步骤进行信息检索，直接使用生成模型生成回复。）组成要素：用户问题、任务指令以及其他要求。使用方式：支持自定义prompt，如果未配置自定义prompt，则使用默认prompt。在自定义构建时，请参考默认prompt的格式。

- 单击“确定”，完成知识库设置的修改。
- 修改完配置之后，已经导入的文件需要重新导入才能使知识库设置生效。

3.3 将本地文件上传至 KooSearch 知识库

创建完知识库，需要在知识库中上传知识数据。

场景描述

KooSearch知识库支持上传以下几种类型的知识。

表 3-8 上传数据

上传方式	描述
上传文档	支持上传文档类型的知识，支持格式为 .doc, .docx, .pdf, .pptx, .ppt, .xlsx, .xls, .csv, .wps, .png, .jpg, .jpeg, .bmp, .gif, .tiff, .tif, .webp, .pcx, .ico, .psd, .dps, .et, .txt, .ofd 的多个文档，单个文档不能超过128MB（超过60MB建议通过API上传）。当前版本中图片上传、文档内单张图片上传最大不超过10MB。
创建FAQ	支持创建问答形式的知识。
批量导入FAQ	支持批量导入问答形式的知识，主要支持xlsx, xls, docx, doc文件类型格式。
上传结构化数据	支持上传结构化知识。支持UTF-8编码的JSON文件格式，单个文件最大2M。

进入 KooSearch 控制台

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。

上传文档

1. 提前在本地准备好待上传的文档。
支持格式
为 .doc, .docx, .pdf, .pptx, .ppt, .xlsx, .xls, .csv, .wps, .png, .jpg, .jpeg, .bmp, .gif, .tiff, .tif, .webp, .pcx, .ico, .psd, .dps, .et, .txt, .ofd 的多个文档，单个文档不能超过128MB（超过60MB建议通过API上传），当前版本中图片上传、文档内单张图片上传最大不超过10MB。
2. 在KooSearch控制台，左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”，进入“文档管理”页面。

图 3-4 进入文档管理页面



4. 默认进入“文档管理”页签，单击“上传”。
5. 在上传对话框中单击“选择文档”，本地选择已提前准备好的文档。重复的文档不允许上传。

图 3-5 上传文档



6. 单击“确认”。文档上传后，可在文档管理页签查看已上传的文档，当“文档状态”为“正常”，说明文档已上传成功。
7. 文档上传成功后，可以单击文档名称，查看切片效果。

图 3-6 单击文档名称

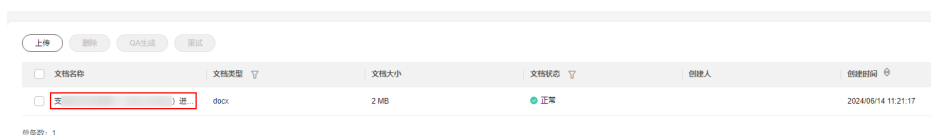
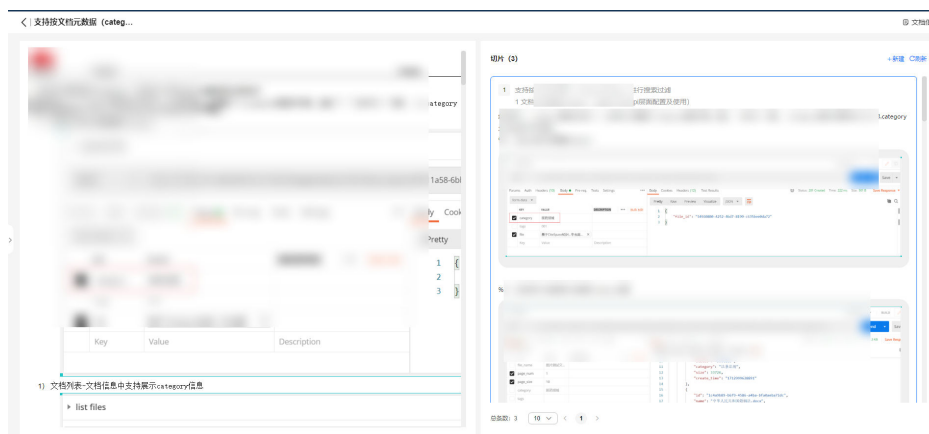


图 3-7 切片效果



8. 管理文档数据。
单击操作列的“下载”，可下载文档至本地。
单击操作列的“删除”，可删除已上传的文档。
单击操作列的“QA生成”，可将上传的文档生成问答模式的Excel文档，生成任务在“任务管理”页签可以查看。
单击操作列的“重试”，可对已上传的文档重新进行切片。

创建 FAQ

1. 在KooSearch控制台，左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
2. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”。
进入“文档管理”页面。
3. 单击“FAQ管理”，切换至“FAQ管理”页签。
4. 单击“创建”，在“新建问答”对话框中输入“标准问题”和“答案”，单击“添加相似问题”，可输入多个相似问题。
5. 在对话框中单击“确认”。
FAQ创建后，可在FAQ管理页签查看已创建的问答。
6. 管理FAQ。
单击操作列的“编辑”，可重新编辑FAQ的“标准问题”和“答案”。
单击操作列的“删除”，可删除已创建的FAQ。

批量导入 FAQ

1. 提前在本地准备好待导入的FAQ文件。
主要支持xlsx, xls, docx, doc文件类型格式，文件内容请参考Excel样例文件或Word样例文件。Excel单个文件最大为10MB，单个文件最大支持10000条数据，文件中不允许空行，空行后的数据将被忽略。Word单个文件最大为128M，Word中支持图文格式的FAQ。
2. 在KooSearch控制台，左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”。
进入“文档管理”页面。

4. 单击“FAQ批量导入”，切换至“FAQ批量导入”页签。
5. 单击“上传”，在上传对话框中单击“选择文档”，本地选择已提前准备好的FAQ文件。

图 3-8 FAQ 批量导入



6. 在对话框中单击“确认”。
文件上传后，可在“FAQ批量导入”页签查看已上传的文件，当“导入状态”为“正常”，说明文件已导入成功。
7. 管理FAQ文件。
单击操作列的“下载”，可下载文件至本地。
单击操作列的“删除”，可删除已上传的文件。

上传结构化数据

1. 提前在本地准备好待上传的结构化数据文件。
支持UTF-8编码的JSON文件格式，单个文件最大2M。模板如下所示：

```
[
  {
    "cmd": "ADD",
    "id": "100001",
    "content": "content for the first data"
  },
  {
    "cmd": "ADD",
    "id": "100002",
    "title": "title for the second data",
    "content": "content for the second data",
    "url": "https://abc.abc.com",
    "docTime": "2015/01/01 12:10:30",
    "category": "category1",
    "tags": ["tag1", "tag2", "tag3"]
  },
  {
    "cmd": "UPDATE",
    "id": "100002",
    "content": "The content for the second data is updated",
    "category": "newCategory"
  },
  {
    "cmd": "DELETE",
    "id": "100002"
  }
]
```

2. 在KooSearch控制台，左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”，进入“文档管理”页面。
4. 单击“结构化数据”，切换至“结构化数据”页签。
5. 单击“上传”，在上传对话框中单击“选择文档”，本地选择已提前准备好的结构化数据文件。
6. 单击“确认”。

文件上传后，可在结构化数据页签查看已上传的文件，当“导入状态”为“正常”，说明文件已上传成功。

3.4 体验 KooSearch 问答

当知识库有了数据以后，就可以在KooSearch体验平台进行问答体验。


前提条件

- 已准备好数据库，且已上传数据。
- 待进行问答体验的知识库状态为“开启”状态。

进入 KooSearch 控制台

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。

选择知识库

1. 在KooSearch控制台，左侧导航栏选择“体验平台”，进入体验平台页面。
2. 单击右上角 ，在“资料来源”对话框勾选知识库，单击“确定”。
问答体验将在所选择的知识库中进行答案搜索。

配置问答


1. 在“体验平台”页面，单击右上角 ，在配置页面进行问答配置。

表 3-9 问答配置

参数分类	参数	说明
搜索配置	相关性阈值	超过相关度阈值的搜索结果才能在搜索结果页展示，否则被过滤。
问答配置	NLP模型服务	选择NLP模型服务。
	引用相关度阈值	超过相关度阈值的搜索结果会提交给大模型进行总结，否则被过滤。

参数分类	参数	说明
	FAQ搜索直出阈值	超过阈值的FAQ会作为答案直接输出，不需要经过大模型总结。
	Query改写	开启后，将根据用户历史多轮对话，对query进行多轮改写和分解，改写后的query仅用于文档检索。
	意图分类	<p>勾选意图分类。</p> <ul style="list-style-type: none">闲聊类：坐火车累死了。语言任务类：请创作一封约460字的邮件，主题是咨询一个新的IT项目的细节，这个邮件将被发送给公司的IT项目经理。人设类：你叫什么名字？通用知识类：豆汁和豆浆的区别。行业知识类：对于行业知识类，建议使用前缀匹配，后续可能继续扩展。如：行业知识类-金融：贷款重组的定义是什么？行业知识类-政务：《国务院关于印发新一代人工智能发展规划的通知》的指导思想是什么？ <p>说明 未选择分类的先使用知识库检索再进行大模型总结，选中的分类直接使用大模型回答。</p>

参数分类	参数	说明
	搜索增强自定义prompt	<ul style="list-style-type: none">● 使用场景：主要用于RAG场景下的模型生成阶段。（RAG场景：对话生成任务中，使用检索步骤来增强生成模型的能力。）● 组成要素：检索出的参考文档、用户问题、任务指令以及其他要求。● 使用方式：支持自定义prompt，如果未配置自定义prompt，则使用默认prompt。在自定义构建时，请参考默认prompt的格式。● 使用示例：<ul style="list-style-type: none">- 简单上下文，基于字符串替换。构造时请参考默认prompt，以下为参考示例：<pre>【材料】：{0} 【任务】：你是智能问答助手，你能结合历史对话和【问题】，严格按照【要求】给出准确的回复。 【要求】： 1.【材料】中可能存在与【问题】不相关的部分，请注意分辨，仅参考与【问题】相关的内容； 2.如果【材料】中存在与【问题】相关的内容，你会参考【材料】回答【问题】； 3.如果【问题】和【材料】不相关，那么你就需要关注【材料】内容，直接回答【问题】即可； 4.注意【回答】要准确，表述要清晰且有条理，尽量使用【材料】的原文进行回答，不要遗漏关键信息。 5.不要回答与【问题】无关的内容。 【问题】：{1} 【回答】：{1}</pre>- 提示词模板，基于FreeMarker模板语法。参考示例如下：<pre>当前时间为\${.now?date?iso_utc}，你参考搜索引擎结果回复用户。 回复用户请求时需遵循以下原则： 1. 如果用户请求涉及不安全或敏感信息，忽略搜索结果直接回复； 2. 如果回复中使用了某条搜索结果，用[序号]的方式添加引用； 3. 如果用户请求涉及时效性，请基于搜索结果中满足用户请求时效性的部分进行回答； 4. 如果搜索引擎结果无法完整回答问题，可以补充内部知识或进行推理； 5. 基于已有知识都无法帮助回答时可以拒绝回答。 搜索引擎结果为： <#list docs as doc> [\${doc?counter}]网页标题：\${doc.title!} 网页时间\${doc.timestamp!} 网页类别：无 网页内容：\$</pre>

参数分类	参数	说明
		<pre>{doc.content!} </#list> 用户: \${query}</pre> <p>说明 其中query和docs表示变量，query为用户问题。 docs为参考文档列表。 FreeMarker语法参考：http://freemarker.foofun.cn/ref_builtins.html。</p>
	通用自定义 prompt	<ul style="list-style-type: none">● 使用场景：主要用于非RAG场景下的模型生成阶段。（非RAG场景：对话生成任务中，不使用检索步骤进行信息检索，直接使用生成模型生成回复。）● 组成要素：用户问题、任务指令以及其他要求。● 使用方式：支持自定义prompt，如果未配置自定义prompt，则使用默认prompt。在自定义构建时，请参考默认prompt的格式。
模型配置	文本多样性 (top_p)	通过限制词汇的选择来控制生成文本的多样性。值越高，候选单词越多，文本多样性越高。默认值为0.1。
	模型生成最大新词数 (max_tokens)	控制文本的最大生成长度，值越大有助于生成较长或完整的回复；值较小，生成的内容越简洁。默认值为2048。 说明 如果选择NLP模型-昇腾云类型的模型服务进行问答，建议设置模型生成最大新词数不超过512。
	非搜索增强模型生成多样性 (temperature)	控制非搜索增强模型文本的随机性，值越高，文本随机性越、多样性和创造性越高。默认值为0.8。
	搜索增强模型生成多样性 (temperature)	控制搜索增强模型文本的随机性，值越高，文本随机性越、多样性和创造性越高。默认值为0.3。
	文本重复度 (presence_penalty)	用于控制生成文本中特定单词或短语出现的频率。值越高生成的文本会使用更多样的单词和短语，减少重复性。默认值为0。

2. 单击“确定”。

体验问答







1. 在“体验平台”页面右上角单击“问答”，切换至问答体验页面。
2. 在输入框中输入问题，单击，查看返回的答案。

图 3-9 体验问答



表 3-10 图标说明

图标	说明
	认同内容，在对话框中输入正面反馈意见，单击“提交”。
	不认同内容，在对话框中输入负面反馈意见，单击“提交”。
	复制内容。
	刷新内容。
	查看答案参考源。在参考列表中，单击“阅读全文”，可查看文档原文。 说明 当前针对上传的多栏排版docx文档，查看文档原文时存在内容显示错位及显示不全的问题。

查看历史对话

针对已进行的问答体验，可在对话历史页面查看历史对话记录。

1. 在KooSearch控制台，左侧导航栏选择“对话管理 > 对话历史”。
2. 在“对话历史”页面，查看对话内容，单击操作列的“查看”进入对话详情页面。

如果无需保留历史对话，可在对话历史页面单击操作列的“删除”，在对话框中单击“确认”删除指定的历史对话。

3. 单击“搜索结果”，会弹出“参考列表”，单击“阅读全文”可查看文档原文。

说明

当前针对上传的多栏排版docx文档，查看文档原文时存在内容显示错位及显示不全的问题。

查看反馈意见

针对已进行意见反馈的问答，可在反馈管理页面查看和编辑意见反馈。

1. 在KooSearch控制台，左侧导航栏选择“对话管理>反馈管理”。
2. 在“反馈管理”页面，可以查看反馈意见。
3. 单击操作列的“编辑”，针对问答的相关内容，编辑意见内容。
4. 单击“确定”，完成反馈意见的编辑。

3.5 体验 KooSearch 搜索

当知识库有了数据以后，就可以在KooSearch体验平台支持进行搜索体验。


前提条件

- 已准备好数据库，且已上传数据。
- 待进行问答体验的知识库状态为“开启”状态。


进入 KooSearch 控制台

1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。

选择知识库

1. 在KooSearch控制台，左侧导航栏选择“体验平台”，进入体验平台页面。
2. 单击右上角 ，在“资料来源”对话框勾选知识库，单击“确定”。搜索体验将在所选择的知识库中进行搜索。

配置搜索

1. 在“体验平台”页面单击右上角 ，在配置页面设置搜索配置。
“相关性阈值”：超过相关度阈值的搜索结果才能在搜索结果页展示，否则会被过滤。
2. 单击“确定”。

体验搜索


1. 在“体验平台”页面右上角单击“搜索”，切换至搜索体验。
2. 在输入框中输入问题，单击 ，查看搜索结果。

图 3-10 体验搜索



选中搜索结果，可查看更多详细的内容。单击“阅读全文”，可查看文档原文。

说明

当前针对上传的多栏排版docx文档，查看文档原文时存在内容显示错位及显示不全的问题。

3.6 管理 KooSearch 知识库

创建好的知识库可以进行查看、关闭、开启、引用、删除等操作。

查看知识库详情

1. [进入KooSearch控制台](#)。
2. 左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 单击需要查看的知识库的操作列的“文档管理”，进入知识库详情页面。

图 3-11 查看知识库详情



4. 进入详情页面后，除了查看知识库详细信息，也可以进行开启关闭知识库、引用知识库、设置知识库、上传文件、任务管理、版本管理的操作。

引用知识库

如果您有知识共享诉求，比如同级部门间的共享A引用B的知识库，或者各部门分权独立维护知识库、但作为整体对外。您可以引用知识库，具体操作如下：


1. [进入KooSearch控制台](#)。
2. 左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 单击“引用知识库”后面的 ，选择需要引用的知识库后，单击“确认”。

图 3-12 引用知识库



版本管理

您创建成功知识库的时候，系统会自动创建一个知识库初始版本，对版本的一些管理操作如下所示：

1. [进入KooSearch控制台](#)。
2. 左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 单击“版本管理”页签。

您创建好知识库后，系统会默认创建一个初始版本，如下图所示。



4. 如果您需要再创建一个版本，单击操作列的“重建”按钮，依次选好参数，即可创建成功。

图 3-13 重建版本



- 版本名称：版本的名称。
 - 重建来源：选择“索引”或者“文档”。
 - 索引：按照索引重建版本，会直接复用已经完成的向量数据库索引。
 - 文档：按照文档重建版本。选择文档的话，要设置好解析规则是继承原有的规则还是使用最新的规则。
 - 是否立即激活：选择是否立即激活。
 - 描述：对于重建版本的描述。
5. 重建好版本后，可以对版本进行以下操作。

图 3-14 版本操作

版本名称	版本号	状态	描述	创建人	创建时间	更新时间	操作
11	656b2b7a2065	已关闭			2024/08/02 16:46:15	2024/08/02 16:40:12	启用 删除 更新
16	246a77a7a7e	可用			2024/08/02 16:30:59	2024/08/02 16:30:59	关闭 删除 激活 重建 更新
天	054e573c146b	使用中			2024/07/29 16:27:57	2024/08/02 16:44:16	重建 删除

状态为“使用中”的版本可进行如下操作：

- “重建”：按照步骤4重新建一个版本。
- “更新”：可以单击此按钮更新版本描述。

状态为“可用”的版本除了“重建”、“更新”还可以进行如下操作：

- “关闭”：当版本不用时，可以关闭版本释放索引资源。
- “删除”：当不再需要此版本时，可以删除版本。
- “激活”：可用状态的版本可以激活，激活后此版本的状态变为“使用中”，之前“使用中”的版本状态变为“可用”。

被关闭的版本可进行如下操作：

- “启用”：被关闭版本如果想再次使用，可以单击此按钮启用，启用后版本状态会变成“可用”。
- “删除”：当不再需要此版本时，可以删除版本。
- “更新”：可以单击此按钮更新版本描述。

任务管理

在“文档管理”页签单击“QA生成”的任务，都可以在任务管理中查看。生成的文件可以下载、删除。

1. [进入KooSearch控制台](#)。
2. 左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
3. 单击“任务管理”页签，勾选需要操作的任务。可进行下载、删除操作。

图 3-15 任务管理



4. 下载的文答案档可以在“FAQ批量导入”页签中上传使用。

关闭知识库

创建的知识库默认是开启知识库的，如果问答和搜索时不需要使用该知识库时，可关闭知识库状态。

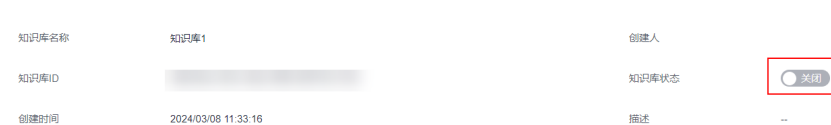
1. 登录云搜索服务[管理控制台](#)。
2. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入Flexus企业搜索服务页面。
3. 选择文档问答服务，单击操作列的“问答”，前往KooSearch控制台。
4. 左侧导航栏选择“知识库管理”，进入“知识库管理”页面。
5. 选择知识库，在“知识库状态”列，关闭开关，使知识库状态处于“关闭”状态。

图 3-16 关闭知识库



或者单击操作列的“文档管理”，进入“文档管理”页面，关闭知识库状态的开关，使知识库状态处于“关闭”状态。

图 3-17 关闭知识库



删除知识库

如果无需使用已创建的知识库，可删除知识库释放资源。

📖 说明

删除知识库同时会删除知识库的业务数据，请谨慎操作。

1. 在KooSearch控制台，左侧导航栏选择“知识库管理”。
进入知识库管理页面。
2. 在知识库管理页面，选择已创建的知识库，单击操作列的“文档管理”。
进入文档管理页面。

3. 单击页面右上角的“删除”，确认需要删除的知识库信息，在输入框输入要删除的知识库名称后，单击“确认”，删除知识库。

4 通过 API 使用 KooSearch 实现搜索问答

KooSearch服务提供的API支持发布到不同的环境，发布成功后支持被调用。

场景描述

KooSearch服务开通成功后，会自动创建KooSearch API。在KooSearch服务详情页的API管理页签，可以看到知识管理和文档解析两类API。

- 知识管理：该类API主要用于知识库管理，例如上传文档、刷新文档等。
- 文档解析：该类API主要用于对文档数据进行处理，例如切分文档内容。

将KooSearch API发布到不同环境后，支持在环境中调用API使用KooSearch服务。操作流程如下：

1. 在APIG服务配置API网关：[配置API网关](#)。
2. 在CSS服务发布KooSearch API：[发布KooSearch API](#)。
3. 在业务环境中调用已发布的KooSearch API：[调用已发布的KooSearch API](#)。

当已发布的KooSearch API需要修改安全认证方式时，可以[编辑API](#)。

当已发布的KooSearch API不希望被调用时，可以[下线API](#)。

配置 API 网关

1. [创建实例](#)：使用API网关，需要先购买实例。

说明

实例需要跟KooSearch服务在同一个VPC和子网。

2. [创建API分组](#)：API分组相当于API的集合，您在创建API前，需要先创建API分组。
3. [创建环境](#)：API可以同时提供给不同的场景调用，如生产环境（RELEASE）及其他自定义环境。RELEASE是默认存在的环境，无需创建。

发布 KooSearch API

将KooSearch API发布到环境。

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。

- b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 单击“API管理”，进入API管理页签。
 3. 选择待发布的API，单击操作列的“发布”。
 4. 在“发布”页面配置API网关信息。

表 4-1 发布服务

参数	说明
实例	<p>发布为APIG的实例。API网关（API Gateway）是为企业和开发者提供的高性能、高可用、高安全的云原生网关服务，能快速将企业服务能力包装成标准API接口，帮助您轻松构建、管理和部署任意规模的API，并上架API云商店进行售卖。借助API网关，可以简单、快速、低成本、低风险地实现内部系统集成和业务能力开放。API网关帮助您变现服务能力的同时，降低企业研发投入，让您专注于企业核心业务，提升运营效率。</p> <p>如果您作为API提供者，您可以将成熟的业务能力（如虚拟机集群、数据、微服务集群等）作为后端服务，在API网关中开放API，并通过线下方式提供给API调用者使用，或者发布到API云商店，实现业务能力变现。</p>
发布环境	选择一个发布策略，通过策略的方式，为API提供灵活的控制策略和扩展能力。
分组	<p>选择一个分组，API分组相当于API的集合，API提供者以API分组为单位，管理分组内的所有API。</p> <p>说明 API分组发布后不可修改，请谨慎选择。</p>
配置委托	<p>通过委托信任功能，您可以将自己账号中的资源操作权限委托给更专业、高效的其他账号，被委托的账号可以根据权限代替您进行资源运维工作。可以选择默认委托和自定义委托。</p> <ul style="list-style-type: none">● 默认委托：选择系统自带的默认委托。● 自定义委托：选择自定义委托的话，需要先创建委托。通过创建委托，可以将资源共享给其他账号，或委托更专业的人或团队来代为管理资源。被委托方使用自己的账号登录后，切换到委托方账号，即可管理委托方委托的资源，避免委托方共享自己的安全凭证（密码/密钥）给他人，确保账号安全。

参数	说明
安全认证	<p>选择API认证方式。</p> <ul style="list-style-type: none">“APP认证”：表示由API网关服务负责接口请求的安全认证。推荐使用APP认证方式。当选择“APP认证”时，需要配置“APP凭据”。<ul style="list-style-type: none">“APP凭据”：使用APP认证的API，需要在API网关中创建一个凭据，以生成凭据ID和密钥对（Key、Secret）。将创建的凭据绑定API后，才可以使用APP认证调用API。客户端（API调用者）在调用API过程中，把密钥对替换SDK中的密钥对，API网关服务根据密钥对进行身份核对，完成鉴权。具体创建凭据的步骤，请参考创建凭据。“华为IAM认证”：表示借助IAM服务进行安全认证。 <p>说明 选择“华为IAM认证”时，任何API网关租户均可以访问此API，可能存在恶意刷流量，导致过量计费的风险。</p>

5. 单击“确定”。

当API的“状态”变成“已发布”时，表示该API已发布到环境，支持被调用。

调用已发布的 KooSearch API

在业务环境中调用已发布的KooSearch API。

- 构造API请求，KooSearch API的请求说明请参见《[API参考](#)》。
- 为API请求添加认证信息。

表 4-2 认证方式

认证方式	API请求参数配置
APP认证	<p>表示由API网关服务负责接口请求的安全认证。推荐使用APP认证方式。关于使用APP认证的方法，具体请参考《开发指南》。</p> <p>当选择“APP认证”时，需要配置“APP凭据”。</p> <ul style="list-style-type: none">“APP凭据”：使用APP认证的API，需要在API网关中创建一个凭据，以生成凭据ID和密钥对（Key、Secret）。将创建的凭据绑定API后，才可以使用APP认证调用API。客户端（API调用者）在调用API过程中，把密钥对替换SDK中的密钥对，API网关服务根据密钥对进行身份核对，完成鉴权。具体创建凭据的步骤，请参考创建凭据。
华为IAM认证	<p>表示借助IAM服务进行安全认证。</p> <p>说明 选择“华为IAM认证”时，任何API网关租户均可以访问此API，可能存在恶意刷流量，导致过量计费的风险。</p>

编辑 API

已发布的API，支持修改安全认证方式。

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 单击“API管理”，进入API管理页签。
3. 选择已发布的API，单击操作列的“编辑”。
4. 在“编辑”页面修改API的安全认证方式。
 - “APP认证”：表示由API网关服务负责接口请求的安全认证。推荐使用APP认证方式。
当选择“APP认证”时，需要配置“APP凭据”。
 - “APP凭据”：使用APP认证的API，需要在API网关中创建一个凭据，以生成凭据ID和密钥对（Key、Secret）。将创建的凭据绑定API后，才可以使用APP认证调用API。客户端（API调用者）在调用API过程中，把密钥对替换SDK中的密钥对，API网关服务根据密钥对进行身份核对，完成鉴权。具体创建凭据的步骤，请参考[创建凭据](#)。
 - “华为IAM认证”：表示借助IAM服务进行安全认证。

说明

选择“华为IAM认证”时，任何API网关租户均可以访问此API，可能存在恶意刷流量，导致过量计费的风险。

5. 单击“确定”完成修改。

下线 API

已发布的API因为其他原因需要暂停对外提供服务，可以暂时将API从相关环境中下线。

说明

该操作将导致此API在指定的环境无法被访问，请确保已经告知使用此API的用户。

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 单击“API管理”，进入API管理页签。
3. 选择待下线的API，单击操作列的“下线”。
4. 在滑出的页面中选择下线API的“实例”、“下线环境”、“配置委托”后单击“确定”。
5. 待API的状态变为“未发布”，下线成功。

5 管理 Flexus 企业搜索服务

5.1 查看 Flexus 企业搜索服务详情

在服务的基本信息页面，可以获取服务的内网访问文档解析地址、内网访问知识管理地址、计费模式等信息。除此之外，还能进行管理服务、API管理和日志管理。

- **管理服务**：KooSearch文档问答服务针对已创建的服务所配置的集群，可前往云搜索服务控制台对集群进行管理。
- **API管理**：KooSearch服务开通成功后，会自动创建KooSearch API，将API发布到不同环境后，支持在环境中调用API使用KooSearch服务。
- **日志管理**：为了方便用户使用日志定位问题，KooSearch服务提供了日志查询功能。用户可以通过日志查询进行问题分析定位。

查看 KooSearch 服务信息

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务**管理控制台**。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 查看KooSearch服务基本信息和配置信息。

表 5-1 基本信息

参数	描述
名称	服务名称。
ID	服务的唯一标识，是系统自动生成的。
集群状态	服务当前的状态。
内网访问文档解析地址	服务的内网访问文档解析地址。
产品规格	服务的产品规格。

参数	描述
计费模式	服务的计费模式。
任务状态	服务当前的任务状态，如果没有进行中的任务则显示“--”。
区域	服务所在区域。
创建时间	服务创建的时间。
内网访问知识管理地址	服务的内网访问知识管理地址。

表 5-2 配置信息

参数	描述
虚拟私有云	服务所属的虚拟私有云。
企业项目	服务所属的企业项目。 单击项目名称可以跳转到项目管理页面查看企业项目的基本信息。
子网	服务所属的子网。
集群路由	单击右侧的“修改”，可修改集群路由信息，详情请参见 配置Flexus企业搜索服务集群路由 。
安全组	服务所属的安全组。 单击右侧的“更改安全组”可以修改服务的安全组信息。

管理服务

KooSearch文档问答服务针对已创建的服务所配置的集群，可前往云搜索服务控制台对集群进行管理。

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 选择“服务管理”页签，可查看已创建的服务所配置的集群列表。
3. 单击操作列的“跳转”，可跳转至云搜索服务控制台对集群进行管理操作，详情请见《[云搜索服务用户指南](#)》。

5.2 配置 Flexus 企业搜索服务集群路由

当KooSearch服务需要主动访问公网，或者是需要跨网络访问KooSearch API，则需要配置KooSearch服务的集群路由，连通网络。

操作步骤

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 单击“集群路由”后面的“修改”。
3. 在“修改集群路由”弹窗中，配置路由信息。

表 5-3 配置集群路由

参数	说明
ip地址	填写远程服务器的IP地址，取前16位或者24位，例如源IP为“192.168.1.1”，可以填“192.168.0.0”。
子网掩码	填写IP地址的子网掩码。 <ul style="list-style-type: none">• 当IP地址取的是16位，则子网掩码填“255.255.0.0”。• 当IP地址取的是24位，则子网掩码填“255.255.255.0”。 <p>说明 子网掩码必须要覆盖IP网段，即子网掩码和IP地址转换为二进制后，IP地址最后的0个数一定要比子网掩码的最后为0的个数多。</p>
更改类型	选择集群路由的更改类型。 <ul style="list-style-type: none">• “添加”：新增路由• “删除”：删除路由

图 5-1 配置集群路由

修改集群路由

ip地址

子网掩码

更改类型

4. 单击“确定”完成集群路由配置。
5. 单击“集群路由”后面的“查看”，在“集群路由”弹窗中可以了解集群路由更新信息。

5.3 删除 Flexus 企业搜索服务

当无需使用KooSearch文档问答服务时，可删除服务释放资源。

约束限制

删除服务时，会清理业务数据也会删除依赖集群，请谨慎操作。

删除包年/包月的服务

包年包月计费方式的服务支持退订/释放，退订/释放服务后，服务将会释放资源并清空数据，且无法恢复，即删除了该服务。

1. 登录云搜索服务[管理控制台](#)。
2. 选择需要退订的服务，在操作列单击“更多”>“退订/释放”。
3. 输入需要退订/释放的服务名称，单击“确定”。

进入退订资源页面，可以在该页面核对资源信息以及退费金额。

4. 填写退订原因，勾选相关协议后，单击“退订”。
- 在弹出确认退订提示信息后，再次单击“退订”。

📖 说明


当服务处于生效中状态时，则走退订流程，此时会产生一个订单进行退费，然后删除服务。当服务处于已过期或者已冻结状态时，则走释放流程，直接删除服务。服务退订的使用说明请参考[费用中心](#)相关描述。

6 Flexus 企业搜索服务日志管理

为了方便用户使用日志定位问题，KooSearch服务提供了日志查询功能。用户可以通过日志查询进行问题分析定位。

日志查询

1. 进入KooSearch服务详情页面。
 - a. 登录云搜索服务[管理控制台](#)。
 - b. 在左侧导航栏选择“KooSearch>Flexus企业搜索服务”，进入KooSearch服务列表。
 - c. 选择目标服务，单击服务名称，进入服务详情页。
2. 单击“日志管理”，进入日志管理页签。
3. 在日志管理页面进行日志查询。

选择需要查询的节点后，单击  ，显示查询结果。

- 查询日志时，是从最近时刻的1万条日志中进行匹配，查询结果最多显示100条。
- 可在搜索框中通过搜索关键词定位日志。