

知识图谱

快速入门

文档版本 01
发布日期 2024-04-24



版权所有 © 华为技术有限公司 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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 知识图谱使用简介.....	1
2 快速创建知识图谱.....	2
3 使用 API 调用知识图谱服务.....	18
4 入门实践.....	23

1 知识图谱使用简介

知识图谱服务（Knowledge Graph，简称KG）提供一站式知识图谱全生命周期管理服务，包括本体可视化构建、自动化图谱流水线构建，以及图谱问答、搜索、推理等图谱应用能力，企业可以灵活掌控图谱配置，适合复杂多变的业务场景。

2 快速创建知识图谱

数据是知识图谱的基础，在知识图谱服务创建知识图谱时，需要从对象存储服务（Object Storage Service，简称OBS）导入用于构建知识图谱的基础数据。知识图谱服务支持的数据包括XLSX、CSV、JSON、多行单句文本格式，详情请见[数据格式要求](#)，其中XLSX、CSV、JSON格式的数据属于结构化数据。

本章节提供一个与电影有关的结构化数据样例，帮助您快速熟悉知识图谱的创建过程。此样例通过在控制台创建本体，完成信息抽取、信息映射、信息融合、图谱质检的配置，创建一个与电影有关的知识图谱。创建完成后，您可以搜索预览相关实体的知识图谱，也可以在创建的图谱基础上全量更新或增量更新图谱。

开始使用样例前，请仔细阅读[准备工作](#)罗列的要求，提前完成准备工作。在控制台上创建知识图谱的步骤如下所示：

- [步骤1：准备数据](#)
- [步骤2：创建本体](#)
- [步骤3：选择图谱规格](#)
- [步骤4：配置数据源](#)
- [步骤5：配置图谱本体](#)
- [步骤6：配置信息抽取](#)
- [步骤7：配置知识映射](#)
- [步骤8：配置知识融合](#)
- [步骤9：配置图谱质检](#)
- [步骤10：生成图谱](#)
- [步骤11：图谱质检](#)

准备工作

已[注册](#)华为账号，并完成实名认证，具体申请信息请见[账号注册](#)和[实名认证](#)。且在使用知识图谱服务前检查账号状态，账号不能处于欠费或冻结状态。

步骤 1：准备数据

在创建图谱之前，您需要将用于创建图谱的基础数据上传至OBS桶及文件夹。

1. 登录OBS服务控制台，创建桶和文件夹，用于存放样例数据集，创建名称为“kg-test”的OBS桶，名称为“demo-data”的文件夹。

创建OBS桶和文件夹的操作指导请参见[创建桶](#)和[新建文件夹](#)。为保证数据能正常访问，请务必保证创建的OBS桶与KG服务在同一区域，桶的存储类别为“标准存储”。

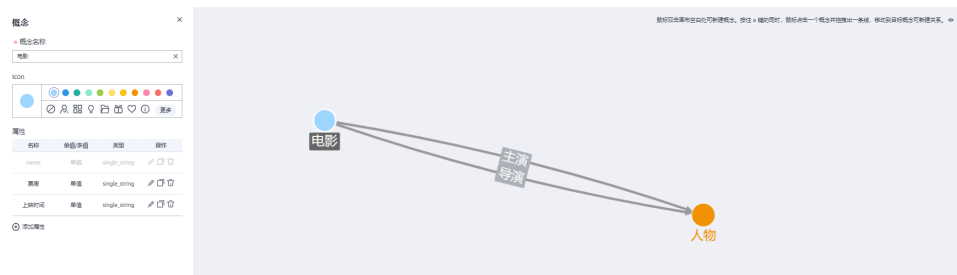
当前KG服务仅支持在“北京四”区域，建议创建OBS桶时，选择“北京四”区域。

2. 单击[数据集下载链接](#)，将数据集下载至本地，并命名为“kg_demo.xlsx”。
3. 参考[上传文件](#)，使用单个文件上传方式将“kg_demo.xlsx”文件上传至OBS路径“kg-test/demo-data”下。

步骤 2：创建本体

在创建图谱之前，您还需要提前创建本体，根据样例数据，此处本体创建2个概念和2种关系，2个概念为“电影”和“人物”，2种关系为“导演”和“主演”。

图 2-1 创建本体



1. 登录KG服务管理控制台，在左侧菜单栏中单击“我的图谱资产库 > 我的本体”，进入“我的本体”页面。

图 2-2 本体管理



2. 单击左上角“创建本体”，弹出“创建本体”对话框。
3. 在对话框中，“本体名称”文本框中填写本体名称“person_film”，在“本体描述”文本框中填写本体描述“包括电影和人物两个概念”，单击“确定”。

图 2-3 创建本体

创建本体

* 本体名称

本体描述

12/255

说明

在创建本体时，所有概念都默认有一个属性为“name”。

4. 创建两个概念节点“电影”和“人物”。
具体操作如下：
 - a. 双击界面空白处，新建概念节点。
 - b. 右键单击概念图标，选择“编辑”或双击概念图标，左侧弹出概念编辑框。
 - c. 在概念编辑框中，按表2-1编辑“概念名称”、“Icon”和“属性”，单击“保存”。

图 2-4 编辑概念

概念
×

* 概念名称

电影
×

Icon

🚫
👤
📁
💡
📂
🎁
❤️
ℹ️
更多

属性

名称	单值/多值	类型	操作
name	单值	single_string	✎ 📄 🗑️
上映时间	单值	single_string	✎ 📄 🗑️
票房	单值	single_string	✎ 📄 🗑️




⊕ 添加属性

保存

取消

表 2-1 编辑概念

参数	说明	推荐填写
概念名称	概念名称只能由大小写字母、数字、中文、下划线组成，长度为1-63位。	本样例需要创建两个概念，分别可命名为“电影”和“人物”。

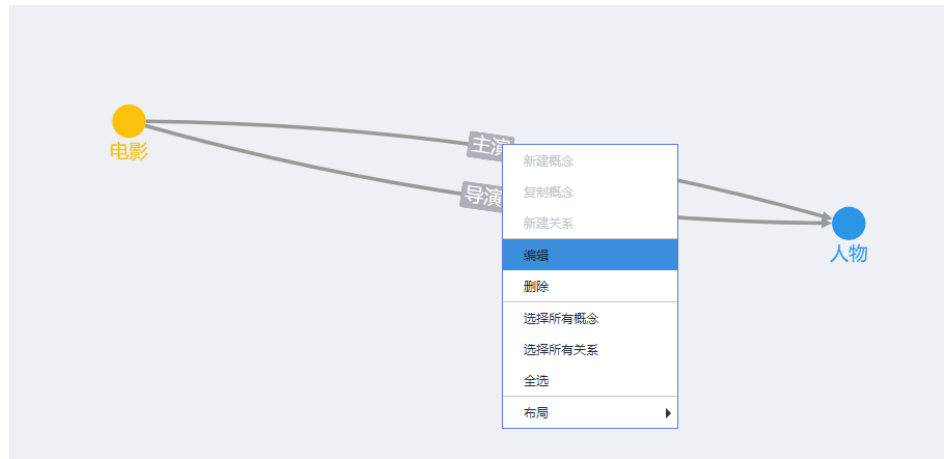
参数	说明	推荐填写
Icon	单击选择概念图标的颜色和样式，单击右侧的“更多”，可选择更多图标样式。	-
属性	<p>编辑当前概念的属性。所有概念都默认有一个属性为“name”，且默认属性类型为“single_string”。</p> <ul style="list-style-type: none"> ● 添加属性 单击“添加属性”可添加属性文本框，在文本框中填写属性的“名称”，选择“单值/多值”、“类型”、“操作”。 <ul style="list-style-type: none"> - 名称：属性名称只能由大小写字母、数字、中文、下划线组成，长度为1-63位。 - 单值/多值：选择当前属性是单值还是多值。 - 类型：选择属性的类型。 单值的可选类型有“single_string”、“single_int”、“single_double”、“single_bool”。 多值的可选类型有“set_string”、“set_int”、“set_double”。 <p>填写完概念的属性后，单击“操作”列的 ，保存信息。</p> <ul style="list-style-type: none"> ● 复制属性 单击已有属性“操作”列的 ，复制当前属性。 ● 删除属性 单击已有属性“操作”列的 ，删除当前属性。 	<ul style="list-style-type: none"> ● 概念“电影”新增属性“票房”和“上映时间”。 ● 概念“人物”新增属性“国籍”、“职业”和“出生日期”。

5. 创建“电影”与“人物”之间的关系。

具体操作如下：

- a. 按住键盘a键，同时鼠标单击概念节点“电影”，拖拽出一条线，移动至目标概念“人物”。
或鼠标右键单击一个概念节点“电影”，选择“新建关系”，然后拖拽出一条线，移动至目标概念“人物”。
- b. 右键单击关系曲线，选择“编辑”。
右侧弹出关系编辑框。

图 2-5 编辑关系



- c. 在关系编辑框中，填写“关系名称”，只能由大小写字母、数字、中文、下划线组成，长度为1-63位。
本样例需创建两次“电影”与“人物”之间的关系，关系名称分别为“主演”和“导演”。

图 2-6 编辑关系

The screenshot shows a dialog box titled '关系' (Relationship) with a close button (X) in the top right corner. Below the title, there is a red asterisk and the text '* 关系名称'. A text input field contains the text '主演'.

- d. 单击“保存”。

步骤 3：选择图谱规格

1. 登录KG服务管理控制台，默认进入“我的图谱”页面。
2. 单击+，然后在创建图谱页面填写信息。

图 2-7 服务选型

The screenshot shows the '创建图谱' (Create Knowledge Graph) page. At the top, there is a progress bar with three steps: 1. 服务选型 (Service Selection), 2. 版本确认 (Version Confirmation), and 3. 完成 (Complete). The '服务选型' step is active. Below the progress bar, there are input fields for '图谱名称' (Knowledge Graph Name) and '图谱规格' (Knowledge Graph Specification). The '图谱规格' field has four options: '体验版 一万边', '标准版 百万边', '标准版 千万边', and '高级版 千万边'. There is also a toggle switch for '是否购买套餐包' (Whether to purchase the package).

表 2-2 服务选型参数说明

参数	说明
“图谱名称”	待创建图谱的名称，名称只能由大小写字母、数字、中文、下划线组成，长度为1-63位。
“图谱规格”	待创建图谱的规格大小。当前仅支持在控制台选择“体验版 一万边”，如果根据业务需要，选择标准版 百万边、标准版 千万边、高级版 千万边，请通过官网产品页“服务咨询”联系我们进行咨询购买。
“是否购买套餐包”	体验版图谱默认不能开通套餐包。如果需要选择业务版图谱，请通过官网产品页“服务咨询”联系我们进行咨询购买。

- 单击右下角“下一步”，在“版本确认”页签确认版本信息。
- 单击右下角“确认创建”。

页面提示“知识图谱创建任务提交成功”。

单击“返回”，返回至“我的图谱”页面，您会看到新创建的图谱卡片，“运行状态”为“创建中”。等待十几分钟后，运行状态变为“初始化”状态。

图 2-8 创建图谱



步骤 4：配置数据源

- 在“我的图谱”页面鼠标移至新创建的图谱卡片，单击“图谱构建 Workspace”。
进入图谱流水线构建页面。

图 2-9 图谱构建 Workspace




2. 在流水线构建页面，单击“数据源”，页面下方弹出数据源配置对话框，单击右侧  按钮可以放大对话框。

图 2-10 配置数据源




3. 在数据源配置对话框，填写相关信息。
 - “选择数据格式”：此样例选择“XLSX”。
 - “选择数据源文件”：单击 ，弹出“选择数据源文件”对话框，选择数据源存放在OBS的路径：
 - i. “OBS桶”：选择数据源文件存放的OBS桶“kg-test”。确保您使用的OBS桶与KG服务在同一区域，桶的存储类别为“标准存储”。
 - ii. “存储路径”：选择数据源文件存放在OBS桶的文件路径“demo-data/kg_demo.xlsx”。
 选择的OBS路径不能是加密路径，否则可能会访问失败。
 - iii. “OBS桶授权”：如果OBS桶未授权，请勾选“确认授权”。
 选择完成后，单击“确定”。

图 2-11 选择数据源文件



4. 填写完信息后，单击“保存”，完成数据源的配置。

步骤 5：配置图谱本体

1. 在流水线构建页面，单击“图谱本体”，页面下方弹出图谱本体配置对话框，单击右侧 按钮可以放大对话框。
2. 选择左侧“我的库”页签，在“图谱本体组件”中拖拽 **步骤2：创建本体** 中创建的 **本体“person_film”** 到虚线框中，单击“保存”，完成本体选择。

说明

“我的库”页签下“图谱本体组件”呈现的是“我的图谱资产库”中“我的本体”页面创建或OBS导入的所有本体。

可单击页面右下角的“编辑”，进入本体修改页面，修改后的本体将保存至“我的库”。

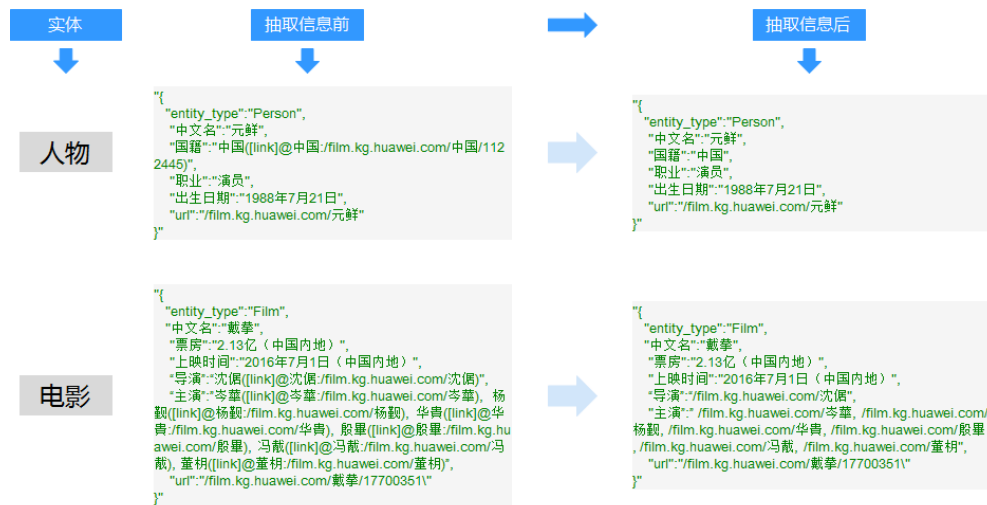
图 2-12 图谱本体



步骤 6：配置信息抽取

以两条数据为例，抽取信息前后的实体信息如图2-13所示，展示如何配置信息抽取。

图 2-13 信息抽取



1. 在流水线构建页面，单击“信息抽取”，页面下方弹出“信息抽取”对话框，单击右侧 按钮可以放大对话框。
2. 在“信息抽取”对话框中，“抽取方式”默认为“结构化抽取”，“编辑方式”选择“交互界面”，关闭“默认抽取”开关。
3. 在信息抽取页面填写信息抽取项，按表2-3填写。

表 2-3 信息抽取项示例

实体类型	数据字段	字段函数
Person	identifier	`\${url}`
	中文名	<code>regexp_replace(`\${中文名}`, '\\([link\\]@.*?:/film.*?)\\)', '')</code>
	国籍	<code>regexp_replace(`\${国籍}`, '\\([link\\]@.*?:/film.*?)\\)', '')</code>
	职业	`\${职业}`
	出生日期	`\${出生日期}`
	url	`\${url}`
Film	identifier	`\${url}`
	中文名	`\${中文名}`
	票房	`\${票房}`
	上映时间	`\${上映时间}`
	导演	<code>regexp_extract_all(`\${导演}`, '\\([link\\]@.*?:/film.*?)\\)', 1)</code>
	主演	<code>regexp_extract_all(`\${主演}`, '\\([link\\]@.*?:/film.*?)\\)', 1)</code>

图 2-15 实体电影的抽取项



表 2-4 实体电影的抽取项

属性/关系名称	抽取项
name	中文名
票房	票房
上映时间	上映时间
电影:主演:人物	主演
电影:导演:人物	导演


- 单击实体“人物”对应的图标。
- 填写实体“人物”映射前的抽取项：“Person”。
- 填写映射到实体“人物”抽取项，如[图2-16](#)和[表2-5](#)所示。

图 2-16 实体人物的抽取项



表 2-5 实体人物的抽取项

属性	抽取项
name	中文名
国籍	国籍
职业	职业
出生日期	出生日期



3. 单击“保存”，完成知识映射配置。

步骤 8：配置知识融合

以如下融合条件为例：

对电影的属性“name”相似的数据进行知识融合判断，当属性“name”的相似度大于等于0.90，且属性“上映时间”的相似度大于等于0.90时，数据进行融合。对人物的属性“name”相似的数据进行知识融合判断，当属性“name”的相似度大于等于0.90，且属性“职业”的相似度大于等于0.90时，数据进行融合。相似度函数均为编辑距离函数“EditSim”。

展示如何配置知识融合。

1. 在流水线构建页面，单击“知识融合”，页面下方弹出“知识融合”对话框，单击右侧  按钮可以放大对话框。
2. 在“知识融合”对话框，打开“知识融合”开关。
3. 在左侧本体预览图中单击实体“电影”图标。
4. 在实体“电影”对应的“融合标识符”栏单击 ，选择标识知识融合的本体属性“name”。
5. 填写实体“电影”的属性配置项信息。

可添加多组属性配置项，每组属性配置项可添加多个判断融合的属性配置。至少存在一组属性配置项，组内所有判断融合的属性相似度均大于相似度阈值时，即可融合这两条数据。

本样例只需添加一个属性配置分组，一组里填写两个属性配置项信息，如[图2-17](#)所示。

单击“添加分组”，即可添加一组新的属性配置项。

单击“增加属性配置项”，即可在组内添加新的属性配置项。

实体配置完成后，单击“确定”。

图 2-17 实体电影的知识融合属性



6. 在左侧本体预览图中单击实体“人物”图标。
7. 在实体“人物”对应的“融合标识符”栏单击 ▾，选择标识知识融合的本体属性“name”。
8. 填写实体“人物”的属性配置项信息。

可添加多组属性配置项，每组属性配置项可添加多个判断融合的属性配置。至少存在一组属性配置项，组内所有判断融合的属性相似度均大于相似度阈值时，即可融合这两条数据。

本样例只需添加一个属性配置分组，一组里填写两个属性配置项信息，如图2-17所示。

单击“添加分组”，即可添加一组新的属性配置项。

单击“增加属性配置项”，即可在组内添加新的属性配置项。

实体配置完成后，单击“确定”。

图 2-18 实体人物的知识融合属性



9. 单击“保存”。

步骤 9：配置图谱质检

1. 在流水线构建页面，单击“图谱质检”，页面下方弹出“图谱质检”对话框，单击右侧 🔍 按钮可以放大对话框。

图 2-19 图谱质检



2. 在“图谱质检”对话框，打开“图谱质检”开关，单击“保存”。

说明

图谱质检为非必选流程，关闭和开启状态均不影响图谱构建。开启质检任务后，后台会对当前版本的知识图谱的知识质量和流水线构建进行评估，并生成质检报告，详见[查看质检报告](#)。

步骤 10：生成图谱

在图谱流水线页面，单击右上角“生成图谱”，完成图谱的创建。

图 2-20 生成图谱

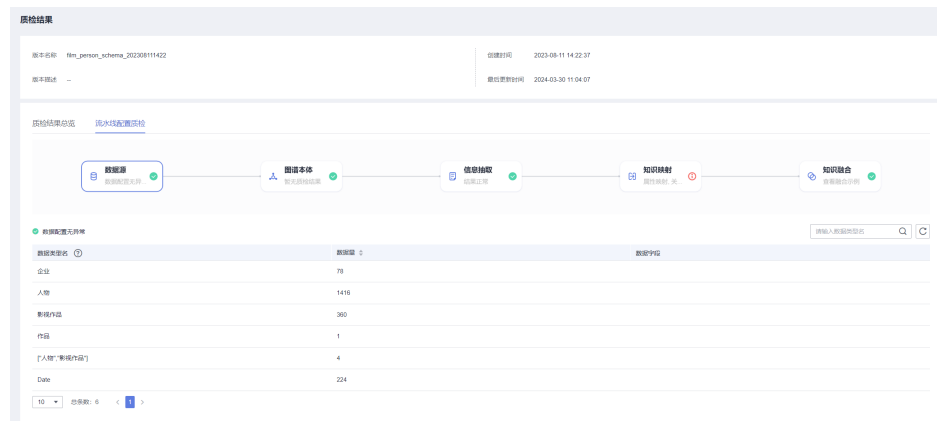


跳至新建图谱“test”的详情页，查看图谱的状态，当图谱的“状态”由“运行中”变为“可用”后，即可单击“图谱预览”，预览图谱。

步骤 11：图谱质检

1. 登录KG服务管理控制台，单击左侧导航栏“我的图谱”，在“我的图谱”页面单击图谱“test”，进入图谱详情页。
2. 在“图谱版本”页签，等待图谱运行完成，单击“更多”列的“查看质检报告”，进入“质检结果”页面。

图 2-21 图谱质检



3 使用 API 调用知识图谱服务

本章节提供了通过Postman调用知识图谱服务API的样例，帮助您快速熟悉零编码使用KG服务的方法。

调用KG服务API的详细步骤如下所示：

步骤1：配置环境

步骤2：Token认证鉴权

步骤3：调用服务

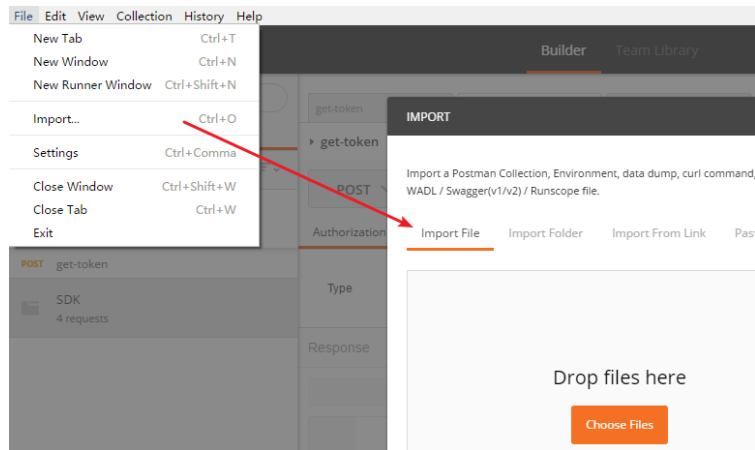
准备工作

- 已注册华为账号，并完成实名认证。具体申请信息请见[账号注册](#)和[实名认证](#)。账号不能处于欠费或冻结状态。
- 已在知识图谱服务上创建知识图谱。

步骤 1：配置环境

1. 下载并安装Postman。
Postman建议使用7.24.0版本。
2. 下载知识图谱的Postman配置文件。
下载地址：[KG_OpenApi.postman_collection.json](#)。
3. 导入配置文件。
 - a. 打开并登录Postman。
 - b. 单击左上角“File > Import > Import File > Choose Files”导入配置文件。

图 3-1 导入配置文件



配置文件导入成功后，将显示在Postman左侧导航栏中。

图 3-2 配置文件



说明

本配置文件中服务所属区域为“华北-北京四”。

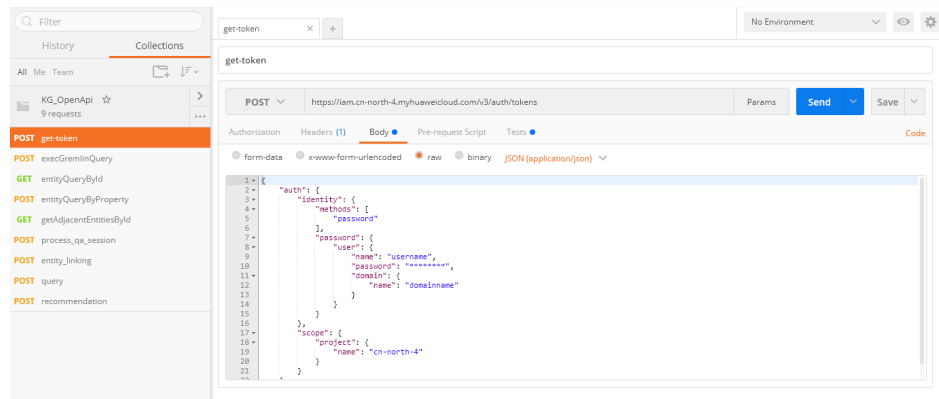
步骤 2: Token 认证鉴权

Token是用户的访问令牌，承载了用户的身份、权限等信息，用户调用API接口时，需要使用Token进行鉴权。

1. 在Postman左侧导航栏中单击“get-token”配置文件。
2. 单击“Body”配置项。填写用户名、密码、账号名。

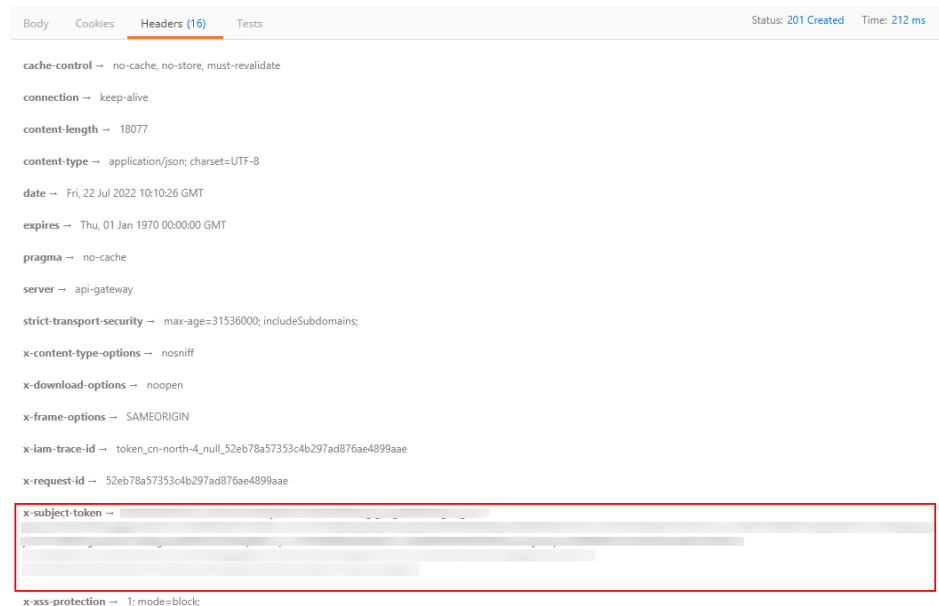
请登录“[我的凭证](#)”页面，复制IAM用户名、账号名至username、domainname参数处，并填写密码。

图 3-3 Token 认证鉴权



3. 单击“Send”，发送请求，获取并复制Token。
获取返回结果“Headers”中的“X-Subject-Token”即为Token。Token有效期为24小时。

图 3-4 获取 Token



步骤 3: 调用服务

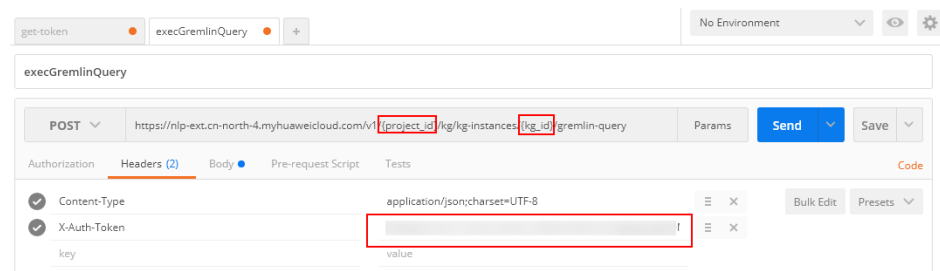
1. 在Postman左侧导航栏中单击“execGremlinQuery”配置文件。
2. 登录[我的凭证](#)，获取“华北-北京四”区域的项目ID，替换配置文件URL中的{project_id}。
3. 登录[知识图谱服务管理控制台](#)，在图谱管理页面，单击图谱卡片，进入图谱详情页，复制图谱ID替换配置文件URL中的{kg_id}。

图 3-5 图谱 ID



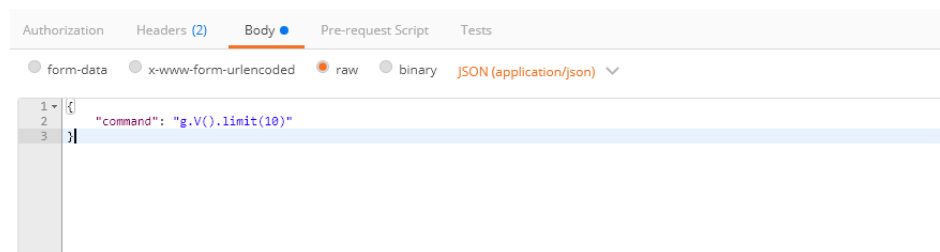
4. 单击“Headers”配置项。将获取的Token复制到“X-Auth-Token”的参数值中。

图 3-6 填写 Token



5. 单击“Body”配置项，修改配置文件内容，也可以使用下图中提供的语句。其中，command参数为必填项，需要输入查询语句。

图 3-7 修改配置文件



6. 单击“Send”，发送请求，获取调用结果。

图 3-8 响应结果

```
1- [
2-   {
3-     "data": {
4-       "vertices": [
5-         {
6-           "id": "8040650499211750405",
7-           "label": "专辑",
8-           "properties": {
9-             "音乐风格": [
10-              "流行"
11-            ],
12-             "发行地区": [
13-              "中国"
14-            ],
15-             "name": [
16-              "忘不了的"
17-            ],
18-             "专辑语言": [
19-              "普通话, 粤语"
20-            ],
21-             "曲目数量": [
22-              "31"
23-            ],
24-             "uri": [
25-              "忘不了的"
26-            ],
27-             "发行时间": [
28-              "2010年10月13日"
29-            ]
30-           }
31-         },
32-         {
33-           "id": "8983060685448381666",
34-           "label": "人物",
35-           "properties": {
36-             "职业": [
37-              "音乐人"
38-            ],
39-             "出生日期": [
40-              "1964年4月18日"
41-            ],
42-             "别名": [
43-              "苏芮; Susan"
44-            ]
45-           }
46-         }
47-       ]
48-     }
49-   }
50- ]
```

4 入门实践

当购买知识图谱并选择图谱规格之后，可以根据自身的业务需求使用KG提供的一系列常用实践。

表 4-1 常用最佳实践

实践	描述
结构化数据创建图谱	本实践指导使用结构化数据（XLSX、CSV、JSON格式的数据）创建图谱。以与电影有关的结构化数据为例，通过在控制台创建本体、完成信息抽取、知识映射、知识融合、图谱质检的配置，运行对应流水线任务，创建一个与电影有关的知识图谱。创建完成后，可以搜索预览相关实体的知识图谱，也可以在创建的图谱基础上全量更新或增量更新图谱。
非结构化数据创建图谱	本实践指导使用非结构化数据（多行单句文本文件）创建图谱。以与人物、电影有关的非结构化数据为例，通过在控制台进行数据标注、模型训练，构建一个人物、电影有关的信息抽取模型，在自定义抽取模型的基础上去创建图谱。
使用公有资产库中的图谱流水线套件构建图谱	本实践提供一个与电影有关的非结构化数据样例，以帮助用户快速熟悉使用流水线套件构建知识图谱的过程，以及如何使用公有资产库中的图谱本体组件、模型组件。创建完成后，可以搜索预览相关实体的知识图谱，也可以在创建的图谱基础上全量更新或增量更新图谱。