

DDE岩浆岩数据库

**参考文献补全工具**（依据百度学术）

DDE-OnePetrology Reference Tool

（V 0.1）

DDE岩浆岩团队

中国地质科学院地质研究所

2024.06

目录

[1. 引言 3](#_Toc170236960)

[1.1. 编写目的 3](#_Toc170236961)

[1.2. 项目背景 3](#_Toc170236962)

[1.3. 术语定义 3](#_Toc170236963)

[1.4. 所有权说明 3](#_Toc170236964)

[1.5. 参考资料 3](#_Toc170236965)

[2. 软件概述 4](#_Toc170236966)

[2.1. 目标 4](#_Toc170236967)

[2.2. 功能 4](#_Toc170236968)

[2.3. 性能 4](#_Toc170236969)

[3. 运行环境 5](#_Toc170236970)

[3.1. 硬件要求 5](#_Toc170236971)

[3.2. 支持软件 5](#_Toc170236972)

[4. 使用说明 6](#_Toc170236973)

[4.1. 安装和初始化 6](#_Toc170236974)

[4.2. 软件启动 6](#_Toc170236975)

[4.3. 界面布局及操作简介 8](#_Toc170236976)

[4.4. 关键词检索 8](#_Toc170236977)

[4.5. 百度学术搜索链接 11](#_Toc170236978)

[4.6. 组合检索 13](#_Toc170236979)

[5. 下一版本规划功能 14](#_Toc170236980)

[6. FAQ 15](#_Toc170236981)

[6.1. 一直在灰色，转圈转个不停，无法操作 15](#_Toc170236982)

[6.2. 没有能搜索全部的关键词结果，就弹框报错了 15](#_Toc170236983)

[6.3. 下载结果一闪找不到了 15](#_Toc170236984)

[6.4. 下载结果是一个zip文件，该怎么用？ 15](#_Toc170236985)

[6.5. 我只想导入到note express中部分结果怎么办 15](#_Toc170236986)

# 引言

## 编写目的

本文档用于介绍DDE岩浆岩数据库参考文献补全工具的功能，使得阅读者能够使用本软件开展工作，以便共同建设DDE岩浆岩数据库。

## 项目背景

DDE岩浆岩数据库隶属于DDE大科学计划，由中国地质科学院地质研究所牵头，联合中国地质调查局发展中心等单位共同建设，

## 术语定义

DDE：深时地球，国际大科学计划，详见 <http://www.ddeworld.org/>

OnePetrology: 岩浆岩数据库，隶属于DDE，包括后台数据库、网站、研究工作平台桌面软件，工具软件集（本说明书所属）等组成。

Python：编程语言，用于科学计算和AI等。本软件采用python结合playwright，flet等众多开源软件组件构建而成。

百度学术: 由百度公司建设的学术搜索引擎。

## 所有权说明

数据来源是百度学术，归百度所有。本软件为岩浆岩工作组研发，归岩浆岩工作组所有。

## 参考资料

暂略.

# 软件概述

## 目标

辅助科学家自动补全参考文献，并提供导出文件到Note Express等文献管理工具中，提高科研效率。

## 功能

实现自动从百度学术搜索的结果中提取所需要的信息，节约科学家的时间。

## 性能

本工具软件理想的运行环境是桌面版，以便利用分布式算力。为了方便科学家们使用（无需安装），特地部署了网络版，但网络版会严重消耗服务器端的算力，导致用的人多了会压垮服务器，因此请各位科学家们错峰使用，并尽量不要往团队外扩散，以免因使用人数过多导致服务器崩溃。

对该功能比较满意打算长期使用的，可联系作者索取python脚本或者桌面安装程序。由于涉及到与百度学术是否产生争议暂未可知，故暂不在网站上公布该信息。

# 运行环境

## 硬件要求

网页版本对硬件运行环境无要求，普通PC即可。

桌面版要求内存大于16GB，CPU i7及以上。

本软件运行需要联网，并对网速有一定要求，网速过慢容易导致卡顿。

## 支持软件

经测试，网页版本可运行在Edge、Chrome、firefox等主流浏览器上。

桌面版在Win10 64位可流畅运行本软件。后期可单独打包linux及国产操作系统的发行版。

# 使用说明

## 安装和初始化

桌面版需要安装，请单独索取安装程序或者python脚本。

Web版本，访问如下网址：

<https://dde.igeodata.org/ref/> 注意：最后一个/不可缺少

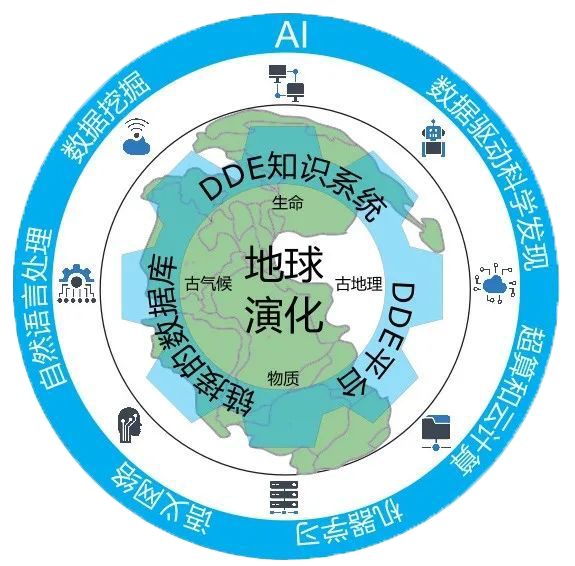
访问网址后，部分浏览器如Edge和Chrome的新版本，可以把网址安装为应用，进而在桌面和任务栏创建快捷方式，方法即观察地址栏的右侧，是否出现install的图标。安装为应用以后，再打开浏览器输入地址，地址栏会显示一个在应用内打开的图标。

## 软件启动

桌面软件和python脚本启动，分别双击exe和运行python即可。

Web启动方式，打开浏览器，输入网址 <https://dde.igeodata.org/ref/> 后期在浏览器中安装应用后，可直接双击创建的快捷方式。

首先出现一个Splash画面



然后系统初步完成加载后，即显示主功能窗口：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

中间的空白区域是输出检索结果的。操作以后可以看到滚动显示一条条参考文献信息。

## 界面布局及操作简介

图形用户界面, 文本, 应用程序

描述已自动生成

如图所示，参照图中说明。工具共支持五种格式，请勾选自己需要的，默认选择ris（据说该格式被广大软件广泛支持）。如果需要其他的格式，请在搜索之前提前勾选。

## 关键词检索

关键词检索可以按照关键词最大化的检索出百度学术中的资料，以A型花岗岩为例，在检索框中输入“ A-type granite” ，点击检索按钮，即可看到一条条的参考文献出现在下方滚动框中：

文本

描述已自动生成

等待检索全部完成或者出错了的时候，会出现 download result 按钮

在检索过程中，可以点击 停止按钮 停止操作，在点击停止按钮以后，需要等待一段时间才可以看到下载按钮，因为它会坚持把当前百度学术的分页内容采集完才结束。

“a-type granite“ 在100结果以后点击停止，它会自动下载完当前页内容到110（百度默认分页一页10个）个以后停止，并显示downlaod result按钮

图形用户界面, 文本

描述已自动生成

点击“download result “,会触发浏览器的下载动作，

图形用户界面, 文本, 应用程序

描述已自动生成

下载文件名为搜索关键词+时间戳.zip

双击打开该文件：

图形用户界面

描述已自动生成

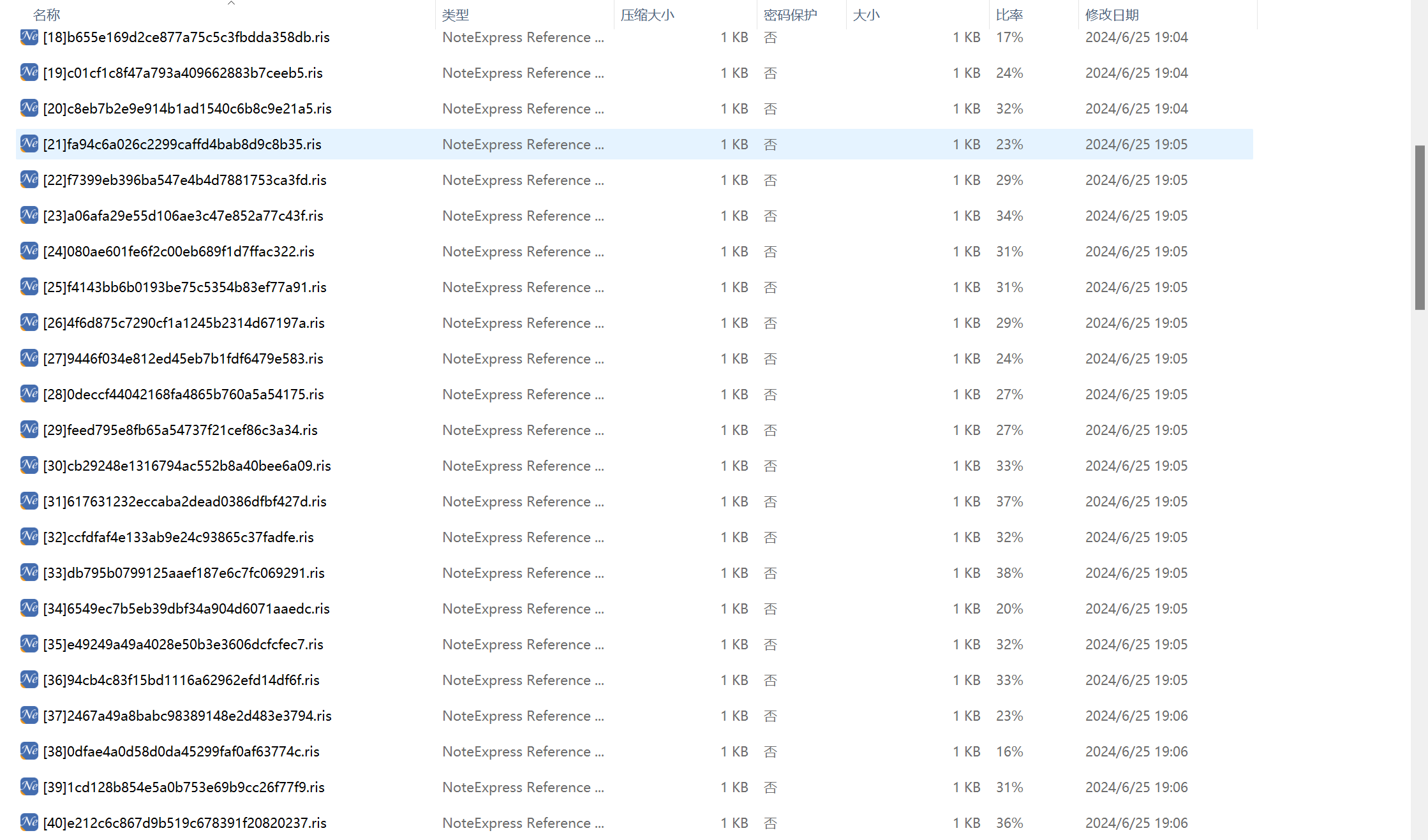
其中，all.ris，为检索到的110个参考文献的ris格式汇总，双击可以把110篇结果直接导入note express。

Apa.txt为APA格式的110篇参考文献的汇总

Gbt.txxt为“GB/T 7714-2015“格式的110篇参考文献的汇总。

Mla.txt为“MLA“格式的110篇参考文献格式的汇总。

Refman为110篇单个文献的ris格式（如果选择了其他导出格式，也会出现对应的子目录）子目录，其内容如下：



选择其他格式同理。

\*\*\*\*以上为单次关键词检索的操作步骤。如果需要一次性搜索多组关键词进行叠加，则可以在停止后重新输入新的关键词再次搜索，结果会追加在后面。（如果两组关键词搜索结果产生了交集，并不会被滤重，在导入Note Express等软件时自行滤重。

\*\*\*\*如果希望每次关键词的结果不叠加，请关闭当前窗口并重新打开即可。

## 百度学术搜索链接

可以直接粘贴百度学术配置好的搜索的链接来进行搜索。还是以“A-type granite“关键词为例，在百度学术的页面输入这两个关键词，即可看到如下界面：

**图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成**

比如我们选择“2024年以来的 地质学领域的文献“，即可得到如下结果：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

可以看到经过设置条件后，准确定位到了9个结果，把该链接地址从浏览器地址栏中复制并粘贴到搜索框中，即可准确提取这9个结果。如下图所示：

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

\*\*\*\*\*该功能适用需要进一步缩小范围进行精准检索，同样，也适合网络原因等导致分页下载的时候失败以后进行继续操作时，例如下载到100的时候失败了，下次可以跳转到第11页，然后手工复制链接进行继续检索。

\*\*\*\*\*\*所有来自百度学术链接的搜索结果，下载时会统一命名为：自定义搜索+时间戳.zip 。这就可能产生一种极端情况：有两个或者多个科学家在同一秒中点击了搜索按钮，大家的结果就可能会混在一起。。。希望不要出现这种中奖的情况：），使用桌面版可以规避该问题。

## 组合检索

即混合以上两种检索方式，一直检索到自己满意为止，再统一下载。

如果不想保留当前的检索结果，请关闭窗口后再打开。如果只是刷新链接，当前结果会被保留。下一个版本考虑增加手工清除功能。

# 下一版本规划功能

看应用需求，暂不考虑增加更多功能。可能会增加一个别人都在搜啥的通知选项。。。

# FAQ

## 一直在灰色，转圈转个不停，无法操作

这可能是网络断开了，或者用的人太多了把服务器搞宕机了。请联系索取单机版或者python脚本。

## 没有能搜索全部的关键词结果，就弹框报错了

服务器模拟人工去访问百度学术（爬虫技术会被封），由于服务器的网络问题导致无法继续（也有可能访问太密集被百度封锁了）。请稍后再试，或者手工从百度学术复制分页链接地址再点击继续。如果频繁报错不能使用，请联系索取单机版或者python脚本。

## 下载结果一闪找不到了

每个浏览器对下载的响应动作不同，可以去浏览器的下载菜单中去查看。一般都是默认在当前用户中的下载目录下。

## 下载结果是一个zip文件，该怎么用？

双击或者解压缩就可以看到内部结果。

## 我只想导入到note express中部分结果怎么办

去子目录中，按照编号查找对应的单个文件，双击导入到对应软件中。