

关于“Google 学术检索”的相关问题解答说明

一、说明

本文主要解释几个 Google 学术检索现象，以及为什么会出现“有些文章能检索出来，有些文章检索不出来”的情况。

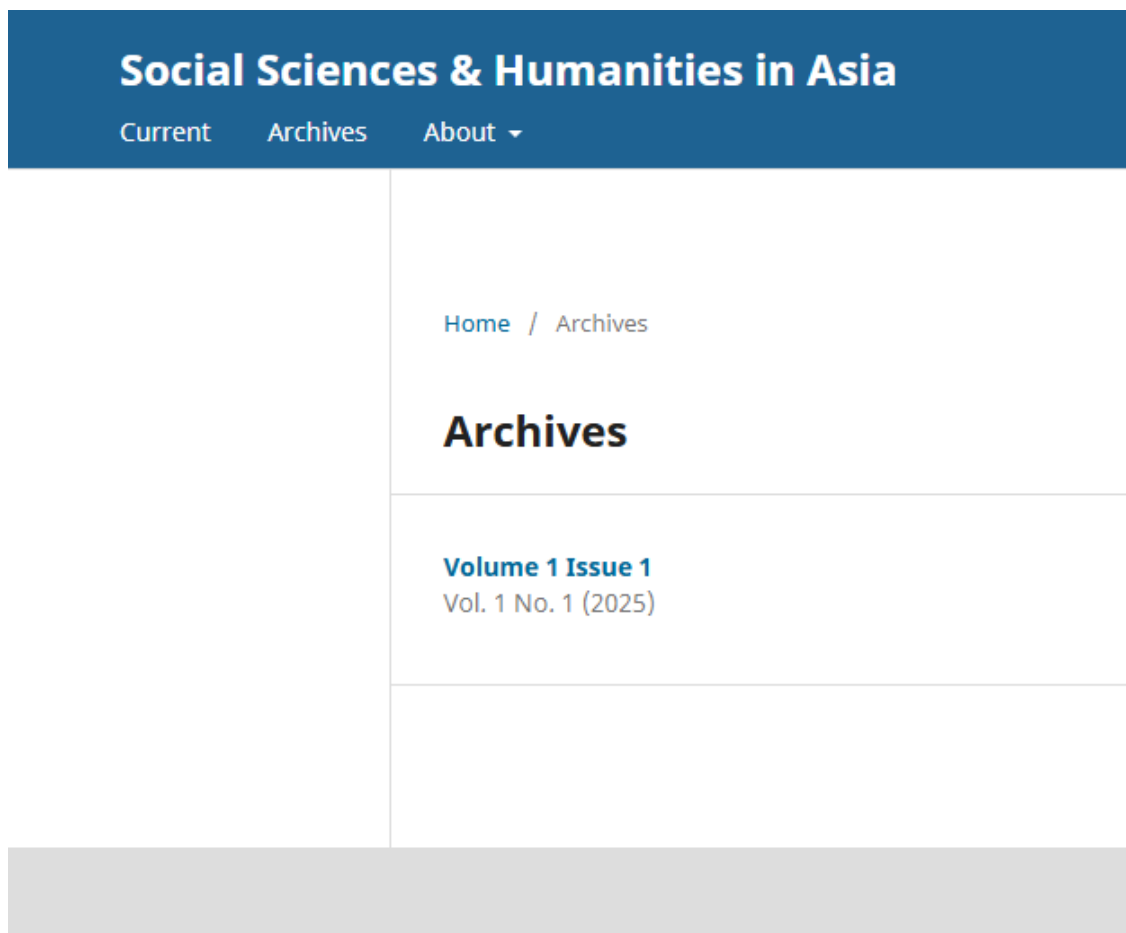
需要先说明一点：Volksonpress 官网本身特殊架构情况，不参与下文 Google 学术检索答疑，后文会单独说明其对 Google 学术抓取和 SEO 的影响。

二、几个具体问题的答疑

1. 关于“SSHA 3 期文章只有 2025 年第 1 卷第 1 期（5 篇）已被 Google 学术检索，2025 年第 1 卷第 2 期、2026 年第 2 卷第 1 期（共 10 篇）没有被检索”

这个情况并不是 Google 学术只检索了其中一部分文章，而是因为在当前 OJS 系统中 (<https://ojs.volksonpress.com/index.php/SSHA/issue/archive>)，SSHA 期刊实际只存在一个卷期：

- 2025 年第 1 卷第 1 期



也就是说，在 OJS 系统的归档页面中，目前只能看到这一个卷期。因此 Google 学术能够检索到的，也只能是这个卷期下已经存在的文章。

所以，“2025 年第 1 卷第 2 期、2026 年第 2 卷第 1 期（共 10 篇）没有被检索”这个问题，本质原因不是 Google 学术没有抓取，而是：

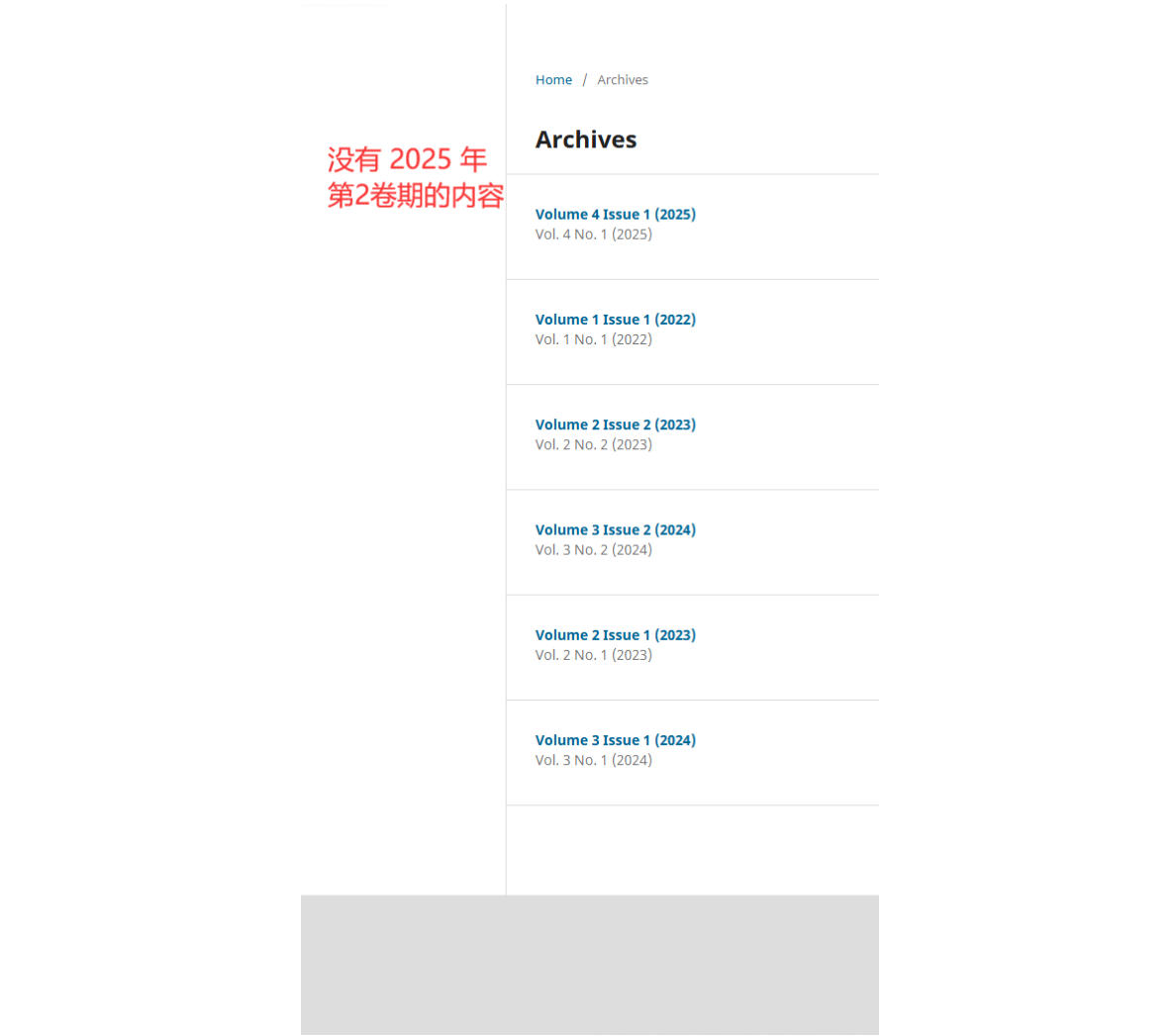
当前 OJS 系统中并不存在这些卷期和文章页面，因此 Google 学术没有可抓取、可索引的目标页面。

简单来说：
Google 学术只能检索已经真实存在于网站上的文章页面，不能检索系统中不存在的卷期或文章。

2. 关于“JRA 期刊 2025 年第 2 期无法检索，2022-2024 年检索文章正常跳转至 JRA 的 OJS 投稿系统”

这个情况与 SSHA 的情况类似。

目前在 JRA 的 OJS 系统中，并没有对应的“2025 年第 2 期”卷期页面 (<https://ojs.volksonpress.com/index.php/JRA/issue/archive>)，因此 Google 学术自然无法检索到该期文章。



而 2022-2024 年的卷期和文章页面在 OJS 系统中是实际存在的，所以这些文章可以被 Google 学术正常检索，并且点击后能够跳转到 JRA 的 OJS 投稿系统文章页面。

因此，这个问题也可以归纳为：

Google 学术不是按照我们口头或业务上认为“应该存在”的卷期来检索，而是按照网站上实际存在、可访问、可抓取的文章页面来检索。

如果某个卷期没有在 OJS 系统中发布，Google 学术就没有办法检索到该卷期下的文章。

3. 关于“2025 年第 1 期只检索到 RAR 期刊（应该是JRA吧？）的摘要”

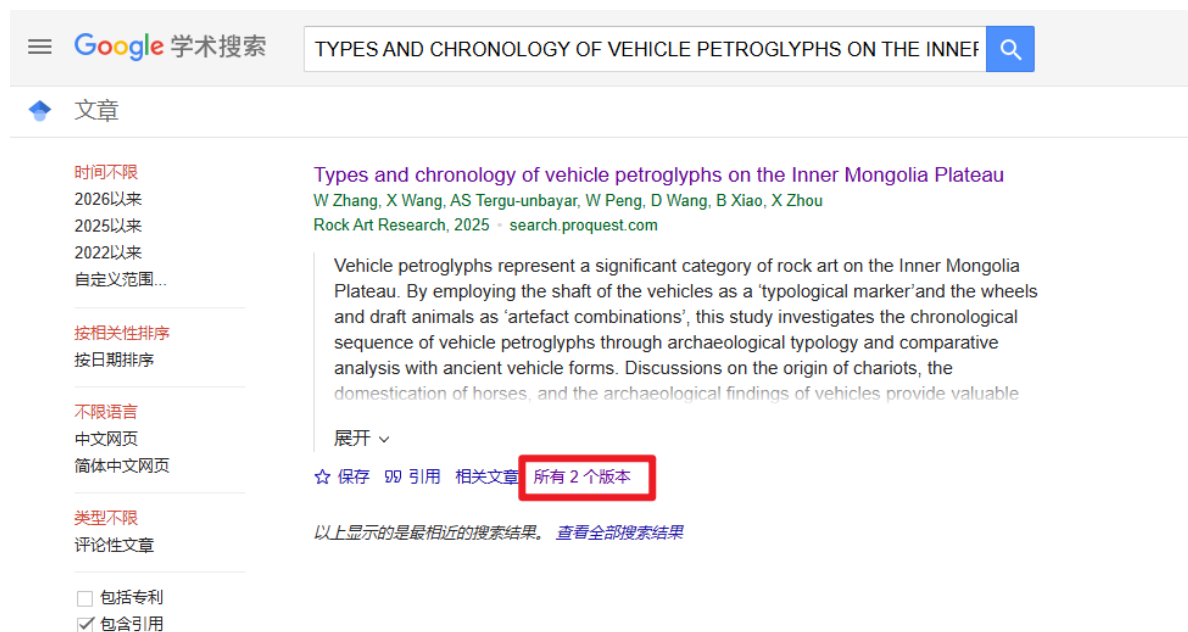
这个情况需要单独说明。

这并不是说 Google 学术只能检索到 RAR 期刊（如果是JRA）的摘要，也不是说我们的 OJS 文章完全没有被检索到。

从检索结果来看，OJS 中对应的文章实际上是可以被 Google 学术识别到的，只是当前搜索结果中，ProQuest 平台的页面显示权重更高，因此排在了更靠前的位置。

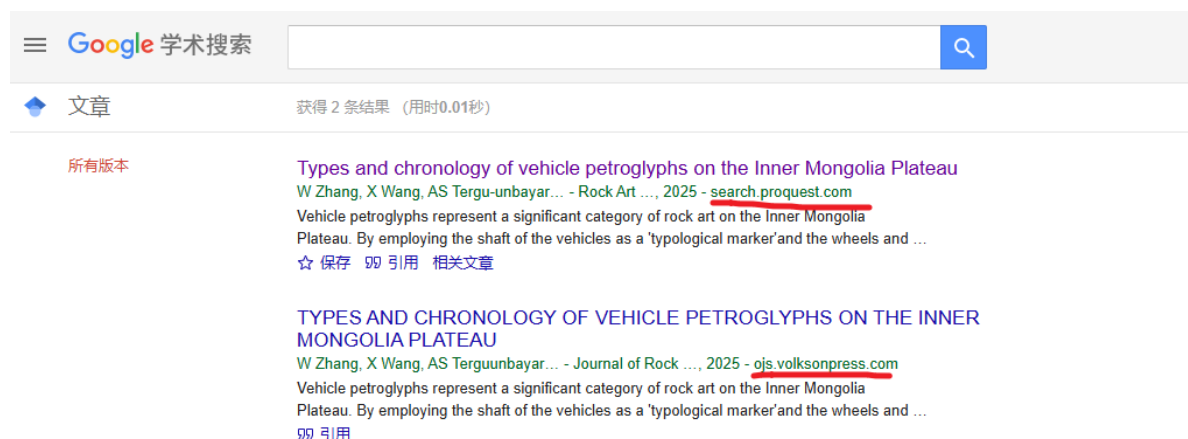
判断依据是：

在 Google 学术检索结果的右下角，出现了“所有 2 个版本”的入口。



这说明 Google 学术认为当前这篇文章至少存在两个版本，例如：

- ProQuest 平台上的版本
- OJS 系统中的原始期刊版本



Google 学术会将它们识别为同一篇学术文章的不同来源版本，并进行合并展示。最终哪个版本排在主结果位置，取决于 Google 学术自己的排序规则，包括页面权威性、可访问性、结构化程度、抓取稳定性、历史收录情况等因素。

因此，这里的准确解释应该是：

RAR (JRA) 的 OJS 文章并不是没有被 Google 学术检索到，而是被 ProQuest 这个第三方学术平台以更高的展示权重排在了主结果位置。OJS 页面作为另一个版本，被收纳在“所有 2 个版本”中。

三、为什么会出现“我的电脑又能检索出来了”或者“我的电脑检索不到，其他人都能检索到”的情况？

如果是在同一本期刊、同一篇文章、同一个卷期已经真实存在的情况下，不同用户之间的检索结果通常应该是接近的。那么我结合以上问题答疑我推测的结果是：

两位同事打开的期刊卷期不是同一个

如果我上述的推测成立，建议两位同事都使用同一个卷期的文章进行检索，结果应该比较相近或相等。

四、关于 Volksonpress 网站的特殊情况

Volksonpress 的情况需要单独说明，因为它和传统 OJS 系统的页面结构不同（SEO能力弱，也就是无法被 Google 学术 检索到的主要原因）。

目前 Volksonpress 网站采用的是前端 SPA 单页面应用架构。SPA 的特点是：

- 页面初始 HTML 内容较少；
- 文章内容通常依赖 JavaScript 动态渲染；
- 路由由前端控制；
- 搜索引擎首次抓取页面时，可能拿不到完整的文章标题、作者、摘要、PDF 链接、DOI 等信息。

对于普通 Google 搜索来说，Google 对 JavaScript 渲染有一定支持；但对于 Google 学术来说，它更依赖稳定、清晰、机器可读的学术文章页面结构。

Google 学术更希望文章页面具备以下特征：

- 每篇文章有独立、稳定、可直接访问的 URL；
- 页面初始 HTML 中直接包含文章标题、作者、摘要、期刊名、出版时间、DOI 等元数据；
- 页面提供标准的学术元标签，例如：
 - `citation_title`
 - `citation_author`
 - `citation_publication_date`
 - `citation_journal_title`
 - `citation_volume`
 - `citation_issue`
 - `citation_firstpage`
 - `citation_lastpage`
 - `citation_doi`
- PDF 文件可被搜索引擎直接访问；
- 网站没有通过 robots.txt、登录权限、接口限制等方式阻止爬虫访问；
- 文章页面不是纯前端渲染后才出现内容。

因此，如果 Volksonpress 的文章详情页主要依靠 Vue3 SPA 动态渲染，那么 Google 学术在抓取时可能会遇到以下问题：

1. 页面初始 HTML 内容为空或内容过少

- 爬虫访问文章 URL 时，看到的可能只是一个空的 `#app` 容器。

- 真正的文章标题、作者、摘要、DOI 等内容需要等 JavaScript 执行后才出现。
- 这会降低 Google 学术识别文章的成功率。

2. 缺少标准学术元数据

- 如果页面中没有 `citation_*` 相关 meta 标签, Google 学术就很难准确判断这是一个学术文章页面。
- 即使页面能被普通 Google 搜索收录, 也不代表一定能被 Google 学术收录。

3. 文章 URL 对爬虫不够稳定

- 如果使用 hash 路由, 例如 `/#/article/detail?id=xxx`, 对 Google 学术并不友好。
- Google 学术更适合抓取传统路径形式的 URL, 例如 `/article/xxx` 或 `/journal/vol/issue/article-id`。

4. PDF 与文章详情页关联不清晰

- Google 学术通常需要通过页面元数据识别 PDF 链接。
- 如果 PDF 链接是接口动态返回, 或者需要点击后才生成, 可能导致 Google 学术无法正确关联文章 HTML 页面和 PDF 文件。

五、Google 学术的检索和收录机制

Google 学术和普通 Google 搜索不同, 它不是简单地收录所有网页, 而是重点识别学术内容。

它通常会经历以下几个过程:

1. 发现页面

Google 学术首先需要发现文章页面。

常见发现方式包括:

- 网站内部链接;
- 卷期归档页面;
- 文章列表页;
- sitemap;
- PDF 链接;
- 外部网站引用;
- DOI 或第三方平台页面。

如果一篇文章没有被任何页面链接到, 或者只存在于前端接口数据中, Google 学术就很难发现它。

2. 抓取页面

发现 URL 后, Google 学术会访问该页面。

如果页面可以正常访问, 并且返回的是有效内容, 才有机会进入后续识别流程。

常见抓取失败原因包括:

- 页面不存在;
- 返回 404;
- 需要登录;
- 被 robots.txt 屏蔽;
- 页面加载过慢;
- 页面依赖 JavaScript 渲染;
- PDF 无法直接访问;

- URL 不稳定;
 - 服务器对爬虫访问不友好。
-

3. 识别是否为学术文章

Google 学术不会把所有网页都当成学术文章。
它需要判断页面是否包含学术论文特征，例如：

- 标题;
- 作者;
- 摘要;
- 期刊名;
- 出版日期;
- 卷、期、页码;
- DOI;
- PDF 全文;
- 参考文献;
- 标准学术 meta 标签。

如果页面结构不清晰，或者缺少这些信息，即使页面被普通 Google 搜索收录，也不一定会进入 Google 学术。

4. 提取元数据

Google 学术需要从页面中提取文章元数据，包括：

- 文章标题;
- 作者;
- 出版年份;
- 期刊名称;
- DOI;
- 摘要;
- PDF;
- 引用信息。

如果页面中有标准的 `citation_*` meta 标签，提取准确率会更高。

如果没有这些标签，Google 学术只能从页面正文中推断，准确率会下降。

5. 合并不同版本

同一篇文章可能出现在多个平台上，例如：

- OJS 官网;
- ProQuest;
- 作者主页;
- 学校机构库;
- ResearchGate;
- 预印本平台;
- 其他数据库平台。

Google 学术可能会把它们识别为同一篇文章的不同版本，并在搜索结果中显示：

所有 2 个版本
所有 3 个版本

这就是为什么有时搜索结果主链接不是 OJS，而是 ProQuest 或其他平台。

这不一定代表 OJS 没有被收录，而可能只是 OJS 页面被折叠到了“所有版本”里面。

6. 排序展示

Google 学术最终会选择一个版本作为主结果展示。

排序可能受到以下因素影响：

- 页面稳定性；
- 网站权威性；
- 元数据完整度；
- PDF 可访问性；
- 外部引用情况；
- 页面历史收录情况；
- 是否为更容易访问的全文版本；
- 是否有更清晰的学术结构；
- 是否被更多学术平台引用。

因此，同一篇文章即使 OJS 是原始来源，也不一定永远排在第一位。

7. 定期更新

Google 学术不是实时系统。

文章发布后，可能需要一段时间才会被发现、抓取、识别和展示。

尤其是新卷期、新文章、新站点，通常不会马上出现在 Google 学术中。

如果网站结构不够友好，例如 SPA 页面、缺少 meta 标签、缺少 sitemap、PDF 不稳定，那么收录周期可能会更长，甚至出现部分文章长期无法被识别的情况。

六、结论

针对目前反馈的问题，可以总结为以下几点：

1. **SSHA 和 JRA 的部分卷期无法检索，主要原因是 OJS 系统中没有对应卷期和文章页面。**
 - 卷期不存在，Google 学术无法检索。
2. **RAR (JRA) 期刊不是只能检索到摘要，也不是 OJS 没有被检索。**
 - 实际情况是 ProQuest 的展示权重更高，OJS 页面可能被折叠进“所有版本”中。
3. **Volksonpress 如果需要被检索则需要单独进行SEO优化。**
 - 由于当前网站是 Vue3 SPA 单页面架构，文章内容依赖前端动态渲染，Google 学术识别弱。