

迭代一数据集说明

迭代一数据集概览：

共包含 4 个.csv 格式数据表：

- 1、**project.csv**: 从 Maven Central Repository、Pypi、Rubygems 和 Packagist 四个库文件管理平台中抽取的项目数据。
- 2、**repository.csv**: 从 GitHub、GitLab 和 Bitbucket 三个代码托管平台抽取的各项目对应的代码仓库数据
- 3、**project_dependencies.csv**: 每个 project 不同版本所依赖的第三方库数据。
- 4、**repository_dependencies.csv**: 从 manifest 文件或库文件管理平台声明的依赖信息中抽取的项目所依赖的第三方库数据。

各数据表具体字段含义：

1、project.csv

- (1) project_id: 项目的唯一编号；
- (2) project_name: 项目的名称
- (3) platform: 项目所属的库文件管理平台
- (4) language: 项目采用的编程语言
- (5) description: 项目描述信息
- (6) homepage_url: 项目地址
- (7) created_timestamp: 初次收集该项目数据的时间戳
- (8) updated_timestamp: 更新该项目数据的时间戳
- (9) latest_release_publish_timestamp: 最新版本发布的时间戳
- (10) latest_release_number: 最新版本对应的版本号
- (11) repository_id: 项目所属代码仓库的编号
- (12) repository_url: 项目所属代码仓库的地址
- (13) licenses: 许可证

2、repository.csv

- (1) repository_id: 仓库唯一编号；
- (2) host_type: 仓库所属托管平台名称；
- (3) repository_name: 仓库名称；
- (4) repository_owner: 仓库所有者；
- (5) language: 采用的编程语言；
- (6) description: 仓库描述信息；

- (7) homepage_url: 仓库地址;
- (8) created_timestamp: 初次收集该仓库数据的时间戳;
- (9) updated_timestamp: 更新该仓库数据的时间戳;
- (10) last_push_timestamp: 最后一次进行 push 操作的时间戳;
- (11) fork: 是否可以进行 fork 操作;
- (12) forks_count: forks 数量;
- (13) watchers_count: watchers 数量;
- (14) stars_count: stars 数量;
- (15) contributors_count: 贡献者数量;
- (16) open_issues_count: 开放的 issue 数量;
- (17) default_branch: 默认分支;
- (18) licenses: 许可证。

3、project_dependencies.csv

- (1) dependency_id: 依赖的唯一编号;
- (2) project_id: 项目的唯一编号;
- (3) project_name: 项目的名称;
- (4) platform: 项目所属的库文件管理平台;
- (5) project_version: 项目版本号;
- (6) dependency_project_id: 依赖的项目编号;
- (7) dependency_project_name: 依赖的项目名称;
- (8) dependency_project_platform: 依赖的项目所属库文件管理平台;
- (9) dependency_requirements: 依赖的项目版本要求;
- (10) dependency_type: 依赖类型 (根据使用阶段划分, 如 runtime、test、development 和 build)。

4、repository_dependencies.csv

- (1) dependency_id: 依赖唯一编号;
- (2) repository_id: 仓库唯一编号;
- (3) repository_name: 仓库名称;
- (4) repository_owner: 仓库所有者;
- (5) host_type: 仓库所属托管平台名称;
- (6) dependency_project_id: 依赖的项目编号;
- (7) dependency_project_name: 依赖的项目名称;
- (8) dependency_requirements: 依赖的项目版本要求;
- (9) dependency_type: 依赖类型 (根据使用阶段划分, 如 runtime、test、development 和 build)。