

fuscan

极速通用文件扫描器

用户手册 · 版本 0.2.2

适用对象：初级用户 · GUI 图形界面

目录

1. 软件简介
2. 安装与启动
3. 主界面总览
4. 工作区与任务管理
5. 第一次扫描
6. 查看扫描结果
7. 详情区与命中操作
8. 任务级配置覆盖
9. 规则管理
10. 设置
11. 统计信息
12. 关于页
13. 快捷键
14. 常见问题

1. 软件简介

fuscan 能做什么：

- 多格式扫描：PDF、Word (.docx)、Excel (.xlsx/.xls/.wps)、PowerPoint (.pptx)、ODF (.odt/.ods)、纯文本，以及 ZIP/RAR/7Z 压缩包内文件
- 规则灵活：按文件名、文件内容、路径三类目标匹配，支持 contains/equals/regex 等模式，可 AND/OR/NOT 逻辑组合
- 正则验证：规则编辑器内嵌正则速查手册与验证面板，支持捕获组、命名组、大小写敏感性切换
- 内置规则：随包分发通用安全规则，开箱即用
- 工作区管理：每个扫描任务独立工作区，支持任务级配置覆盖与规则集绑定
- 实时反馈：扫描进度分阶段显示（收集/解析/压缩包），当前文件、命中数、错误数实时刷新
- 结果过滤：按文件路径模糊搜索、严重度过滤、四种排序字段，快速定位关键命中
- 批量替换：对命中内容按规则 replace_with 批量替换，支持从 .bak 备份撤销
- 导出归档：扫描结果可导出为 CSV 或 JSON
- 暗色模式：一键切换浅色/深色主题，所有界面元素自适应

2. 安装与启动

安装

```
pip install fuscan
```

或使用 uv：

```
uv add fuscant
```

注意：GUI 功能需要 PySide2，仅在 Python 3.8~3.10 可用。

若使用 Python 3.11+，会自动使用 PySide6。

启动 GUI

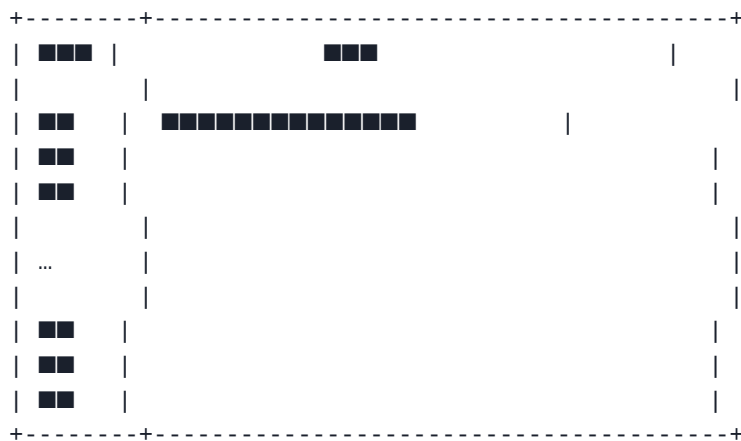
命令行输入：

```
fuscant gui
```

即可打开图形界面。

3. 主界面总览

主界面采用 WSL Dashboard 风格的左右分栏布局：



侧边栏

侧边栏位于左侧，从上到下依次为：

- Logo 区：fuscant 图标与名称
- 首页：工作区列表，管理所有扫描任务
- 添加任务：新建扫描任务表单
- （弹性间距）
- 设置：全局配置
- 关于：版本信息、用户手册与配置目录入口
- 暗色模式开关：一键切换浅色/深色主题

选中某项后，右侧内容区会切换到对应页面，侧边栏同时高亮当前选中项。

内容区

内容区根据侧边栏选中项切换以下页面：

- 首页：工作区卡片列表（默认页面）
- 添加任务：新建扫描任务表单

- 规则：从工作区卡片「定义规则」按钮进入，编辑该工作区的规则集
- 结果：从工作区卡片「查看结果」按钮进入，浏览命中
- 统计：从工作区卡片「统计」按钮进入，查看扫描指标
- 设置：全局配置面板
- 关于：版本与依赖信息

页面切换使用淡入动画，避免视觉跳变。

4. 工作区与任务管理

fuscan 采用工作区（Workspace）作为扫描任务的基本单元。每个工作区独立保存扫描目标、规则集、任务级配置覆盖与上次扫描结果状态，重启后自动恢复。

工作区卡片

首页工作区列表中，每个工作区显示为一张卡片，包含以下信息：

- 第一行：任务图标 + 任务名 + 状态徽标
- 状态颜色：就绪=蓝色、扫描中=黄色、已暂停=灰色、已完成=红色（有命中）或绿色（无命中）
- 第二行：模式（全盘/盘符/文件夹）、目标路径、规则标签
- 规则标签：内置规则=灰色，用户规则=绿色
- 第三行：最近摘要（命中数、跳过数等）
- 第四行：分类计数（符合 / 通过 / 命中 / 跳过 / 错误）
- 第五行：操作按钮
- 左侧主操作：「定义规则」、「启动扫描」/「暂停」、「更新扫描」、「查看结果」
- 右侧辅助操作：「统计」、「展开」

工作区操作按钮

按钮	启用条件	功能
定义规则	任意状态	进入规则页编辑该工作区的规则集
启动扫描	未在扫描中	开始扫描，按钮变为主色高亮
暂停	扫描中	暂停当前扫描
更新扫描	已完成（含用户取消）	对该工作区重新扫描
查看结果	已完成（含用户取消）	进入结果页，按钮变为主色高亮
统计	任意状态	进入统计页查看扫描指标
展开	任意状态	展开更多操作（切换目标、导出、设置、删除）

展开后操作

点击「展开」按钮后，卡片底部显示更多操作：

- 切换目标：修改该工作区的扫描模式与目标路径（扫描中/暂停中禁用）

- CSV：导出扫描结果为 CSV（仅当有命中时启用）
- JSON：导出扫描结果为 JSON（仅当有命中时启用）
- 设置：打开任务级配置覆盖对话框（详见第 8 节）
- 删除：删除该工作区及其所有数据

清空所有工作区

首页右上角「清空」按钮（仅当有任务且无扫描进行时显示）可一键清空所有工作区。

点击后弹出二次确认对话框，避免误操作。若扫描正在进行，会提示无法清空。

5. 第一次扫描

步骤 1：新建任务

1. 点击侧边栏「添加任务」进入新建任务表单
2. 输入任务名称（留空将自动生成，如「任务 1」）
3. 选择扫描模式：
 - 全盘扫描：扫描所有盘符（耗时较长）
 - 盘符扫描：选择具体盘符（如 C:、D:），从盘符按钮列表选中点选
 - 文件夹扫描：选择具体文件夹（最常用），可点击「选择」按钮浏览目录，

或在「最近扫描目录」列表中点击历史路径快速填入

4. 配置规则：
 - 勾选「启用内置通用规则」使用随包分发的安全规则（默认启用）
 - 点击「添加规则文件」加载 YAML 规则文件（.yaml 或 .yml），可加载多个
 - 点击规则文件右侧的关闭图标可移除该规则文件
5. 点击「创建任务」返回首页

步骤 2：开始扫描

在工作区卡片上点击「启动扫描」按钮。

扫描期间，首页会切换为扫描进度面板，显示：

- 任务名称、扫描模式、目标路径
- 双进度条：
 - 收集进度（walk 阶段）：已发现文件数、按白名单跳过数、用户标记跳过数
 - 解析进度（scan 阶段）：已扫描/总文件数、命中数、错误数、速度
- 当前正在扫描的文件名
- 已用时

扫描过程中可点击「暂停」临时暂停，点击「继续」恢复；

也可在「暂停」状态下点击「启动扫描」恢复。

取消加速：扫描过程中侧边栏「首页」会显示扫描进度面板。

取消扫描会立即取消所有未启动的扫描任务，不再等待它们进入扫描队列。

已开始读取的少量文件（最多等于并发线程数）会快速完成，

配合大文件跳过通常在百毫秒内退出扫描循环。

步骤 3：查看结果

扫描完成后，工作区卡片状态徽标变为「已完成」（含命中数颜色提示）。

点击「查看结果」按钮进入结果页。详见第 6 节。

6. 查看扫描结果

结果页布局

结果页采用左右分栏：

- 左侧：命中结果清单（带过滤工具栏）
- 右侧：命中详情面板（文件信息 + 命中详情 + 底部操作栏）

顶部显示标题、当前任务状态、命中总数，以及「← 返回」按钮回到首页。

过滤与排序工具栏

结果清单上方提供过滤+排序工具栏：

- 路径搜索框：按文件路径关键词模糊匹配（不区分大小写），输入后 300ms 防抖触发
- 严重度过滤：下拉选择「全部/严重/警告/信息」
- 排序字段：下拉选择「默认顺序/文件路径/命中数/严重度」
- 排序方向：升序或降序
- 清除：一键重置所有过滤与排序条件
- 计数：显示「过滤后数 / 总数」，如 12 / 156

结果清单

每个命中结果显示为清单项：

- 左侧严重度色条（红=严重、橙=警告、蓝=信息）
- 文件路径（过长时省略中间部分）
- 规则名 + 命中数
- 右侧严重度标签

单击清单项，右侧详情面板会显示该文件的详细信息。

大规模命中（1000+）时，清单会预渲染屏幕外条目，避免滚动时频繁重建。

严重等级

等级	颜色	含义
严重 (CRITICAL)	红色	高危，需立即处理
警告 (WARNING)	橙色	中等风险，建议关注
信息 (INFO)	蓝色	提示性信息

7. 详情区与命中操作

详情面板位于结果页右侧，选中清单项后显示该命中的详细信息与操作按钮。

文件信息卡片

显示：

- 文件信息标题 + 压缩包条目标签（若为压缩包内部条目）
- 文件路径（等宽字体显示）+ 「定位」按钮（在文件管理器中打开并选中该文件）
- 元信息网格：文件大小、命中规则数
- 压缩包内部条目会显示「压缩包路径 / 内部条目路径」双字段，定位按钮禁用

命中详情

列出该文件命中的所有规则，每条包含：

- 严重度色条 + 规则名
- 严重度标签
- 匹配目标（filename/content/path）与匹配次数
- 规则描述（如有）
- 匹配文本（红色等宽字体高亮）
- 上下文片段

右上角「展开/收起」按钮可切换详情列表的折叠状态：

折叠时仅显示第一条摘要，展开时显示全部命中详情。

底部操作栏

详情面板底部提供三行操作按钮：

第一行 - 命中导航（左对齐）：

- 上一条：跳转到上一条命中结果
- 下一条：跳转到下一条命中结果

第二行 - 单文件操作（右对齐）：

- 移至暂存：复制到暂存区隔离目录并标记为跳过（压缩包条目禁用）
- 替换内容：备份源文件为 .bak，按规则 replace_with 替换命中内容

第三行 - 批量操作（右对齐）：

- 全部替换：对当前过滤后的所有命中结果执行批量替换
- 撤销批量：撤销最近一次批量替换，从 .bak 备份恢复所有文件
- 撤销当前：撤销当前选中结果的最近一次替换（从 .bak 恢复）

操作消息会显示在操作栏顶部，成功为绿色，失败为红色，提示为灰色。

替换操作会先备份源文件到 `~/.fuscan/backup/` 目录，

然后按规则定义的 `replace_with` 字段替换命中文本。

压缩包内部条目不支持替换（无法直接写回压缩包）。

8. 任务级配置覆盖

每个工作区可独立覆盖以下全局配置项，不影响其他工作区与全局设置：

- 扫描压缩包：是否扫描 ZIP/RAR/7Z 内文件
- 并发线程数：1-16，0 表示单线程
- 大文件跳过阈值：1B - 500MB
- 扫描深度：限制目录递归深度，空表示不限
- 忽略目录：不扫描的目录名列表（每行一个）

配置入口

工作区卡片展开后，点击「设置」按钮打开任务级设置对话框。

对话框初始值显示当前覆盖值（未覆盖时显示全局值），修改后点击确定生效。

持久化

任务级配置覆盖会持久化到 `~/.fuscan/workspaces.json`，重启后自动恢复。

扫描时优先使用任务级覆盖值，未覆盖的字段回退到全局配置。

9. 规则管理

从工作区卡片点击「定义规则」按钮进入规则页。规则页支持工作区绑定模式：

进入时自动绑定当前工作区，规则编辑仅作用于该工作区的规则集，不影响其他工作区或全局规则。返回首页时自动解除绑定。

规则列表

左侧显示已加载的规则文件列表，右侧显示当前合并后的规则集：

- 规则名称、严重等级、适用扩展名

加载规则文件

点击「加载规则」按钮选择 YAML 规则文件（.yaml 或 .yml）。

若选择的文件已在列表中（如在外部编辑器修改后想刷新），会弹出确认对话框。

调整规则优先级

多个规则文件按顺序链式合并，后加载的规则文件覆盖先加载的同名规则。

通过右键菜单或快捷键调整顺序：

- 上移：右键 → 上移
- 下移：右键 → 下移
- 移除：右键 → 移除（快捷键 Delete）

编辑规则

点击「编辑规则」按钮打开规则编辑器，可修改当前加载的规则文件内容。

正则表达式验证面板

规则编辑器底部集成正则验证面板，帮助编写和调试 mode: regex 规则：

- 速查手册：右侧显示 Python re 模块常用语法速查（字符类、量词、锚点、分组与捕获、零宽断言、内联修饰符），以及邮箱、IPv4、中国手机号等常用示例
- 测试输入：在「正则表达式」输入框填入待测 pattern，在「测试文本」输入框填入待匹配文本
- 大小写敏感性：勾选「区分大小写」后按字面大小写匹配；默认不区分大小写
- 触发测试：点击「测试」按钮或按回车键触发匹配
- 结果展示：列出每个命中的序号、位置（起止索引）、匹配文本；

若 pattern 含捕获组（`(...)`）或命名组（`(?P...)`），同时列出捕获结果

注意：验证面板使用与扫描引擎一致的 `re.compile(...).finditer(text)`

行为，匹配不重叠（从上次结束位置继续），因此测试结果与实际扫描结果一致。

10. 设置

点击侧边栏「设置」进入设置页。设置页采用 TabBar 分组，避免单页滚动过长。

Tab 1：通用

字体设置：

- 字体族：下拉选择系统字体
- 字号：基础字号（默认 14）
- 最小字号：高 DPI 屏幕下字号下限（默认 12，范围 8-16）
- 加粗：是否启用粗体
- 实时预览：修改时即时显示效果

字号会根据基础字号动态计算其他字号档位：

caption=base-2、small=base-1、body=base、heading=base+2、title=base+4、page_title=base+8，

并受最小字号约束避免过小。

Tab 2：扫描

- 缓存：启用扫描缓存可大幅提升重复扫描速度（推荐开启）
- 扫描压缩包：是否扫描 ZIP/RAR/7Z 内文件
- 压缩包密码：扫描加密压缩包时使用的密码
- 大文件跳过阈值：超过此大小的文件不读取内容，避免卡死；

默认 50MB，0 表示不限制

- 并发线程数：多线程扫描，建议 4-8（默认 5）
- 扫描深度：限制目录递归深度，空表示不限
- 显示网络驱动器：盘符列表是否显示网络驱动器
- 性能日志：启用扫描性能日志记录

Tab 3：忽略目录

TextArea 多行编辑忽略目录列表（每行一个目录名），

常见如 .git、node_modules、__pycache__、.venv 等。

文件类型

文件类型列表按类别分组（文档/表格/演示文稿/压缩包/其他），

每个类别支持三态勾选：

- 全选：勾选该类别下所有文件类型
- 部分选中：显示半选状态
- 全不选：取消该类别下所有文件类型

每个文件类型显示：

- 格式标签（蓝色，显示扩展名或格式名）
- 解析速度档次（T1-T5 五格条形图，悬停显示「解析速度：T2 快速」等提示）
- 复选框（勾选启用该类型扫描）

文件类型未全部勾选时，`enabled_extensions()` 返回勾选列表；

全部勾选时返回 `None`，表示扫描所有支持的类型。

路径历史

记录最近扫描过的文件夹路径，点击可快速填入添加任务页的扫描目录。

支持「清除」一键清空全部历史。

重置

设置页右上角「重置」按钮可一键重置扫描相关配置为默认值

（保留路径历史与已禁用的文件类型）。

11. 统计信息

从工作区卡片点击「统计」按钮进入统计页，展示该工作区的扫描统计：

- 状态摘要：当前状态、扫描摘要文本
- 分类计数：符合文件类型数 / 通过 / 命中 / 跳过 / 错误
- 阶段进度：walk 阶段发现数、scan 阶段扫描数
- 耗时统计：总耗时、平均速度

点击「← 返回」按钮回到首页。

12. 关于页

点击侧边栏「关于」进入关于页，显示：

- fuscan Logo 与版本号
- 应用描述
- 作者与许可证
- 快捷入口：
- 用户手册：打开随包分发的 PDF 用户手册（系统默认 PDF 阅读器）
- 配置目录：在文件管理器中打开 `~/.fuscan` 目录
- 第三方依赖列表

若 PDF 手册缺失会弹出提示，不影响主流程。

配置目录包含 `workspaces.json`（工作区持久化）、`cache.db`（扫描缓存）、`backup/`（替换备份）等文件。

13. 快捷键

快捷键	功能
F1	打开用户手册
Ctrl+O	加载规则文件
Ctrl+S	开始扫描
Ctrl+,	打开设置
F3	跳转到下一个命中位置
Shift+F3	跳转到上一个命中位置
Delete	移除选中的规则文件

14. 常见问题

Q: 扫描很慢怎么办？

- 启用缓存（设置 → 扫描 → 启用扫描缓存）

- 缩小扫描范围（用文件夹扫描代替全盘扫描）
- 适当增加并发线程数（设置 → 扫描 → 并发线程数：4-8）
- 忽略无关目录（设置 → 忽略目录：如 .git、node_modules、__pycache__、.venv）
- 调整「大文件跳过阈值」（默认 50MB，超过此大小的文件不读取内容）
- 检查文件类型列表是否勾选了不必要的格式

Q: 扫描时界面卡顿？

- 进度回调已做节流与批处理，正常情况不应卡顿
- 若仍卡顿，可能是命中文件列表过长，可在结果页用过滤工具栏缩小显示范围
- 单个超大文件读取会导致短暂卡顿，可降低「大文件跳过阈值」
- 并发扫描时 PyO3 提取器（pdf_oxide/calamine）会释放 GIL，UI 通常较流畅

Q: 取消扫描后还要等很久才退出？

- fuscant 已实现取消加速：取消后立即取消未启动任务，不再等待它们进入扫描队列
- 仅有少量已开始读取的文件（最多等于并发线程数）会快速完成
- 若仍感觉卡顿，可能是某个 worker 正在读取超大文件，降低「大文件跳过阈值」即可

Q: 找不到某些文件？

- 检查文件扩展名是否在文件类型列表中被取消勾选
- 检查所在目录是否在「忽略目录」列表中
- 检查扫描深度是否受限
- 若在压缩包内，确认「扫描压缩包」已开启
- 检查文件大小是否超过「大文件跳过阈值」（内容规则不会命中，但文件名规则仍生效）

Q: 加密压缩包扫描不出内容？

- 在设置 → 扫描中填写「压缩包密码」
- 或在工作区卡片的任务级设置中填写该任务专属密码
- ZIP 使用标准加密格式，RAR 与 7Z 支持各自加密格式
- 密码错误时该条目会被跳过并记录到错误数，不会中断扫描

Q: 规则不生效？

- 从工作区卡片点击「定义规则」检查该工作区的规则集是否正确加载
- 检查文件类型列表是否勾选了目标文件扩展名
- 尝试菜单：文件 → 编辑规则，确认规则语法正确
- 对 mode: regex 规则，可用规则编辑器底部的「正则验证面板」先测试 pattern
- 注意：规则页采用工作区绑定模式，编辑仅作用于当前工作区

Q: 替换内容后如何撤销？

- 详情面板底部「撤销当前」按钮：撤销当前选中文件的最近一次替换

- 详情面板底部「撤销批量」按钮：撤销最近一次「全部替换」操作
- 撤销会从 `~/.fuscan/backup/` 目录的 `.bak` 备份恢复源文件
- 若备份文件不存在或已损坏，撤销会失败并提示

Q: 任务级设置与全局设置的关系？

- 任务级设置仅对该工作区生效，不影响其他工作区与全局设置
- 未覆盖的字段自动回退到全局设置
- 在工作区卡片展开后点击「设置」按钮可修改任务级覆盖
- 全局设置修改不影响已有任务级覆盖（覆盖值优先级更高）

Q: 重启后工作区还在吗？

- 工作区持久化到 `~/.fuscan/workspaces.json`，重启后自动恢复
- 恢复内容包括：任务名、扫描模式、目标路径、规则文件、任务级覆盖、上次扫描状态
- 运行时状态（如扫描中、已暂停）会重置为「就绪」
- 扫描结果不会持久化，需重新扫描或导出归档

Q: 如何查看历史扫描结果？

fuscan 目前不保存历史扫描结果，每个工作区仅保留最近一次扫描的状态摘要。

若需归档，请扫描完成后点击工作区卡片展开区的「CSV」或「JSON」按钮导出。

路径历史可在添加任务页的「最近扫描目录」列表中查看与清除。

Q: 内置规则有哪些？

随包分发通用安全规则，覆盖：

- 敏感文件名检测（password、secret、key 等）
- AWS 密钥泄露检测
- 数据库连接字符串
- 私钥文件检测
- 其他常见敏感信息

可在规则页查看，或通过任务级设置关闭「启用内置通用规则」仅用自定义规则。

如需更多信息，请参阅项目 README

或通过菜单：帮助 → 关于 → 用户手册 查看本手册 PDF 版本。