【每日一库】ripgrep - grep的替代者

原创 Mike Tang [Rust语言中文社区](javascript:void(0);) 2019-04-17

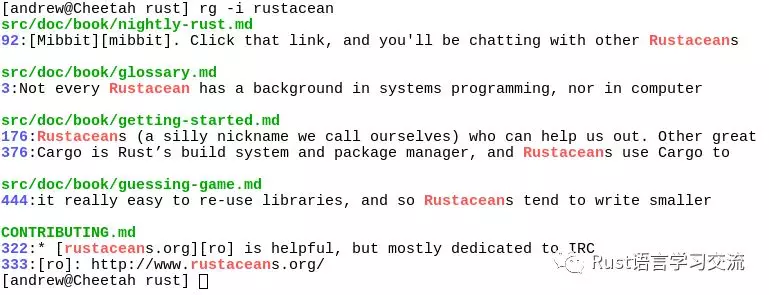
万众期待的 ripgrep 来了。

`grep` 是咱们 Linuxer 几乎每天都会用到的行搜索工具，几乎所有发行版都自带有这个工具。多少年来，没有什么改变，如一潭死水。`ripgrep`的出现，给这个领域带来了一场轰动。

ripgrep 很牛，现在其 github 已经接近 14000 stars 了。仓库地址是：https://github.com/BurntSushi/ripgrep。

ripgrep 超越 grep 的**第一点**就是 ripgrep 是跨平台的，不再歧视 Windows 用户，Windows， Linux，macOS 效果完全一致。

**第二点**，拥有更漂亮的输出。



当然，美这个问题，众口难调。

**第三**，强力的性能。

是骡子是马，拿出来溜溜。

对比 1 ：

搜索 Linux 内核源文件（在执行完 make defconfig && make -j8 之后），CPU Intel i7-6900K 3.2 GHz，ripgrep 开了 SIMD 支持。

| **工具** | **命令** | **行统计** | **耗时** |
| --- | --- | --- | --- |
| ripgrep (Unicode) | rg -n -w '[A-Z]+\_SUSPEND' | 450 | **0.106s** |
| git grep | LC\_ALL=C git grep -E -n -w '[A-Z]+\_SUSPEND' | 450 | 0.553s |
| The Silver Searcher | ag -w '[A-Z]+\_SUSPEND' | 450 | 0.589s |
| git grep (Unicode) | LC\_ALL=en\_US.UTF-8 git grep -E -n -w '[A-Z]+\_SUSPEND' | 450 | 2.266s |
| sift | sift --git -n -w '[A-Z]+\_SUSPEND' | 450 | 3.505s |
| ack | ack -w '[A-Z]+\_SUSPEND' | 1878 | 6.823s |
| The Platinum Searcher | pt -w -e '[A-Z]+\_SUSPEND' | 450 | 14.208s |

(上图, ack 有一个 bug ：）)

对比 2：

还是搜索 Linux 内核源代码，搜索范围和参数有些变化。

| **工具** | **命令** | **行统计** | **耗时** |
| --- | --- | --- | --- |
| ripgrep | rg -L -u -tc -n -w '[A-Z]+\_SUSPEND' | 404 | **0.079s** |
| ucg | ucg --type=cc -w '[A-Z]+\_SUSPEND' | 390 | 0.163s |
| GNU grep | egrep -R -n --include='\*.c' --include='\*.h' -w '[A-Z]+\_SUSPEND' | 404 | 0.611s |

（ucg 在处理符号链接的时候，策略有所不同）

对比3：

搜索一个约 9.3G 的大文件。

| **工具** | **命令** | **行统计** | **耗时** |
| --- | --- | --- | --- |
| ripgrep | rg -w 'Sherlock [A-Z]\w+' | 5268 | **2.108s** |
| GNU grep | LC\_ALL=C egrep -w 'Sherlock [A-Z]\w+' | 5268 | 7.014s |

**为什么推荐使用 ripgrep**

1. 因为它覆盖了其它类似搜索工具的大部分特性，并且更快。
2. ripgrep 会识别 .gitignore 文件并忽略里面指定的文件和目录。默认不会搜索隐藏文件和目录。默认不会搜索二进制文件。
3. ripgrep 可指定搜索某些类型的文件。比如，rg -tpy foo 只搜索 py 文件，rg -Tjs foo 在搜索时排除 js 文件。
4. ripgrep 支持 Unicode。
5. 可选择 PCRE2 作为正则引擎。
6. 支持非 UTF-8 文件的搜索。
7. 支持压缩包里面文件的搜索。
8. 支持任意预处理过滤器。

还等什么，马上执行

$ cargo install ripgrep

安装使用吧。