

OSPP2023: 优化 TUNA 镜像站帮助文档中的 mirrorz-help 支持

郑鈇壬

2023-05-13

- 提供海外常用软件仓库的镜像
- 为校内以及国内用户提供便捷的服务
- Ubuntu, Debian, CentOS 等 204 个软件仓库, 共 115T 的文件
- 目前每天会收到约一亿请求, 平均流量约 10Gb/s, 一天总流量约为 100TiB
 - 之前日常处于满载状态

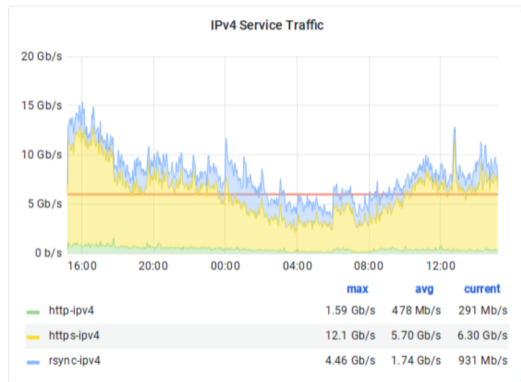


图: TUNA IPv4 流量图

- 如何达到这么多用户？广为流传的文档！
- 搜索「换源」，到处都是 TUNA 镜像站的身影。
- 不过并不推荐随便找一篇文章
- 这是出过问题的！
 - 网上找的 Anaconda 过时配置导致无法使用
 - 「您使用的配置可能是过时配置，请认准帮助文档中的最新配置。」
 - #1734, #1622, #1685, #1572, #1574, #1518
 - 为 Ubuntu focal 替换上了 Ubuntu xenial 的源

- 推荐使用镜像站官方文档
- 有人断言是镜像站最重要的页面
- 目前共有 124 篇镜像帮助文档
- 更新频繁，过去一年有 93 个改动



AOSP

Adoptium

CPAN

CRAN

CTAN

CocoaPods

NetBSD

OpenBSD

OpenMediaVault

alpine

anaconda

anthon

arch4edu

Debian 软件仓库镜像使用帮助

是否使用 HTTPS 是

是否使用 sudo 是

一般情况下，将 `/etc/apt/sources.list` 文件中 Debian 默认的源地址 `http://`

Debian Buster 以上版本默认支持 HTTPS 源。如果遇到无法拉取 HTTPS 源的情况，

```
sudo apt install apt-transport-https ca-certificates
```

Debian 版本:

启用源码镜像: 否

使用官方安全更新软件源: 是

```
# 默认注释了源码镜像以提高 apt update 速度，如有需要可自行取消注释
deb https://mirrors.tuna.tsinghua.edu.cn/debian/ bullseye main cc
# deb-src https://mirrors.tuna.tsinghua.edu.cn/debian/ bullseye
deb https://mirrors.tuna.tsinghua.edu.cn/debian/ bullseye-updates
# deb-src https://mirrors.tuna.tsinghua.edu.cn/debian/ bullseye-
deb https://mirrors.tuna.tsinghua.edu.cn/debian/ bullseye-backpor
```

图: 镜像站帮助文档

- 有的同学可能认为，一个文档而已，并不难
- 需要一些动态内容
 - 不能使用纯 Markdown：变量替换与按钮
- 前端的风格统一
 - 不能随便找一个 Markdown 渲染器

TUNA 前端中的镜像站文档技术发展

- 之前 TUNA 前端中对文档的渲染是「黑魔法」
- 为何会出现黑魔法
- Jekyll Liquid：静态渲染
 - 在生成站点 HTML 时替换
 - TUNA 有姊妹站 BFSU/OpenTUNA，共用前端
 - 不能在 Markdown 中内嵌 mirrors.tuna 的域名
 - `{{ site.hostname }}`
- Marked.js：动态渲染
 - 在用户浏览器中替换
 - 例如 Ubuntu 中用户需要选择版本号，例如 22.04 或 20.04
 - 在 Markdown 中嵌入 HTML 表单与按钮
 - `{{release_name}}`

TUNA 中的 Ubuntu 文档

```
<form class="form-inline">
<div class="form-group">
  <label>选择你的ubuntu版本: </label>
  <select class="form-control release-select" data-template="#apt-template" data-target="#apt-content">
    <option data-release="trusty">14.04 LTS</option>
    <option data-release="xenial">16.04 LTS</option>
    <option data-release="bionic">18.04 LTS</option>
    <option data-release="focal">20.04 LTS</option>
    <option data-release="jammy" selected>22.04 LTS</option>
    <option data-release="kinetic">22.10</option>
  </select>
</div>
</form>
```

```
{% raw %}
<script id="apt-template" type="x-tmpl-markup">
# 默认注释了源码镜像以提高 apt update 速度，如有需要可自行取消注释
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted universe mult
# deb-src https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted univers
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}}-updates main restricted unive
```

图: TUNA 过去的 Ubuntu 文档，两种语法混杂，难以维护

广为抱怨的问题：TUNA 文档的可迁移性

- 由于 TUNA 的体量，TUNA 前端和其中的文档也广为使用
- 不过用户很痛苦
- 直接使用前端的：
 - 内容容易过时，需要时常更新
 - BUPT: [BYRIO/BYRIO-mirrors/issues/19](#)
 - NJUPT: [NJUPT-Mirrors-Group/issues/issues/14](#)
- 只使用文档内容的：
 - 巨大的劳动
 - USTC: 每当 TUNA 有一个改动，USTC 需要同步改动到他们的 ReST 文档中
 - NJU: 前端没有 TUNA 黑魔法，手动替换了变量使得其成为纯 Markdown

- 为了解决上述问题，需要一种通用的镜像帮助文档模板
- 在过去，mirrorz 项目已经收集了各个镜像站的数据（mirrorz.json）
- 需要一种使用可交换的文档格式 MDX（Markdown with JSX）
- 由 SukkaW 于今年寒假设计并实现
 - 镜像站文档模板的具体格式
 - 镜像站文档模板的标准渲染方式

mirrorz-help 中的 Ubuntu 文档

<CodeBlock

```
enableQuickSetup
filepath="/etc/apt/sources.list"
```

```
menus=[
```

```
{
```

```
  title: 'Ubuntu 版本',
```

```
  items: [
```

```
    ['22.04 LTS', { release_name: 'jammy' }],
```

```
    ['23.04', { release_name: 'lunar' }],
```

```
    ['22.10', { release_name: 'kinetic' }],
```

```
    ['20.04 LTS', { release_name: 'focal' }],
```

```
    ['18.04 LTS', { release_name: 'bionic' }],
```

```
    ['16.04 LTS', { release_name: 'xenial' }],
```

```
    ['14.04 LTS', { release_name: 'trusty' }],
```

```
  ]
```

```
},
```

```
{
```

```
  title: '使用官方安全更新软件源',
```

```
  items: [
```

```
    ['是', { security_mirror: '# ', security_official: '' }],
```

```
    ['否', { security_mirror: '', security_official: '# ' }],
```

```
  ]
```

```
},
```

```
]
```

```
````properties
```

```
默认注释了源码镜像以提高 apt update 速度, 如有需要可自行取消注释
```

```
deb {{http_protocol}}{{mirror}}/ {{release_name}} main res
```

```
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{re
```

```
deb {{http_protocol}}{{mirror}}/ {{release_name}}-updates
```

```
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{re
```

```
deb {{http_protocol}}{{mirror}}/ {{release_name}}-backport
```

```
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{re
```

```
{{security_mirror}}deb {{http_protocol}}{{mirror}}/ {{rele
```

```
{{security_mirror}}{{enable_source}}deb-src {{http_protoco
```

```
{{security_official}}deb http://security.ubuntu.com/ubuntu
```

```
{{security_official}}{{enable_source}}deb-src http://secur
```

图: 变量由 CodeBlock 声明, 由 MDX 渲染器替换

# TUNA 中的 Ubuntu 文档

```
<form class="form-inline">
<div class="form-group">
 <label>选择你的ubuntu版本: </label>
 <select class="form-control release-select" data-template="#apt-template" data-target="#apt-content">
 <option data-release="trusty">14.04 LTS</option>
 <option data-release="xenial">16.04 LTS</option>
 <option data-release="bionic">18.04 LTS</option>
 <option data-release="focal">20.04 LTS</option>
 <option data-release="jammy" selected>22.04 LTS</option>
 <option data-release="kinetic">22.10</option>
 </select>
</div>
</form>
```

```
{% raw %}
<script id="apt-template" type="x-tmpl-markup">
默认注释了源码镜像以提高 apt update 速度，如有需要可自行取消注释
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted universe mult
deb-src https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted univers
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}}-updates main restricted unive
```

图: TUNA 过去的 Ubuntu 文档，两种语法混杂，难以维护

- 统一的文档站
- mirrors.help
  - 对应于 mirrorz.org
- help.mirrors.cernet.edu.cn
  - 对应于 mirrors.cernet.edu.cn/list
  - 教育网镜像站
- xtom.help
  - xtom 镜像站



MirrorZ Help 致力于成为一个开源、开放、且持续更新的开源软件镜像的帮助文档整合站点，旨在帮助高校间推广开源软件的使用。

## 热门文档

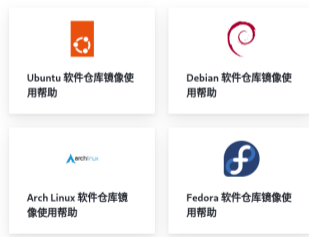
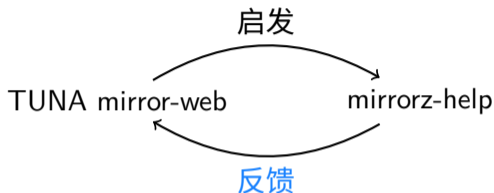


图: MirrorZ Help 页面



- 统一的文档站很好，但并不足够
- 各站前端仍有自己集成文档的需求，基于种种考虑
  - 校内用户无需访问校外
  - 帮助文档已经成了公开 API
  - 前端风格统一
  - 无外链
- TUNA 需要将 MDX 渲染在自己的前端中，并保持原有风格

# 目前的方式

```
const FRONT_MATTER_REGEX = /---([\s\S]*?)---/;
const FRONT_MATTER_TITLE_REGEX = /title: (.+)/;
const CODEBLOCK_REGEX = /\<[\s\r\n]*?CodeBlock([\s\S]*?)\>[\s\S]*?```\s*(\w*)[\r\n]([\s\S]*?)```[\s\S]*?\</CodeBlock\>/;
const MENU_REGEX = /menus={([\s\S]*)}/;
```

图: 使用正则表达式提取 CodeBlock 中的变量与代码

- 用正则表达式匹配 CodeBlock 中的变量和代码
- 之后将提取内容转译 (transpile) 成 TUNA 特色 markdown
- 我一个晚上写的灵车
- 这是 PoC, 另外也为了保证兼容性

# 现在 TUNA 的 Ubuntu 文档

```
<form class="form-inline">
<div class="form-group">
 <label>Ubuntu 版本: </label>
 <select id="select-0-0" class="form-control content-select" data-
 <option data-release_name="jammy" selected>22.04 LTS</option>
 <option data-release_name="lunar">23.04</option>
 <option data-release_name="kinetic">22.10</option>
 <option data-release_name="focal">20.04 LTS</option>
 <option data-release_name="bionic">18.04 LTS</option>
 <option data-release_name="xenial">16.04 LTS</option>
 <option data-release_name="trusty">14.04 LTS</option>
 </select>
</div>
</form>
```

```
{% raw %}
<script id="template-0" type="x-tmpl-markup">
默认注释了源码镜像以提高 apt update 速度，如有需要可自行取消注释
deb {{http_protocol}}{{mirror}}/ {{release_name}} main restricted univer
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{release_name}} #
deb {{http_protocol}}{{mirror}}/ {{release_name}}-updates main restricte
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{release_name}}-u
deb {{http_protocol}}{{mirror}}/ {{release_name}}-backports main restric
{{enable_source}}deb-src {{http_protocol}}{{mirror}}/ {{release_name}}-t

{{security_mirror}}deb {{http_protocol}}{{mirror}}/ {{release_name}}-sec
{{security_mirror}}{{enable_source}}deb-src {{http_protocol}}{{mirror}}/

{{security_official}}deb http://security.ubuntu.com/ubuntu/ {{release_na
{{security_official}}{{enable_source}}deb-src http://security.ubuntu.com
```

图: 灵车驱动, 依然使用 Marked.js, Jekyll 部分被处理了

# TUNA 中的 Ubuntu 文档

```
<form class="form-inline">
<div class="form-group">
 <label>选择你的ubuntu版本: </label>
 <select class="form-control release-select" data-template="#apt-template" data-target="#apt-content">
 <option data-release="trusty">14.04 LTS</option>
 <option data-release="xenial">16.04 LTS</option>
 <option data-release="bionic">18.04 LTS</option>
 <option data-release="focal">20.04 LTS</option>
 <option data-release="jammy" selected>22.04 LTS</option>
 <option data-release="kinetic">22.10</option>
 </select>
</div>
</form>
```

```
{% raw %}
<script id="apt-template" type="x-tmpl-markup">
默认注释了源码镜像以提高 apt update 速度，如有需要可自行取消注释
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted universe mult
deb-src https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}} main restricted univers
deb https://{%endraw%}{ { site.hostname } }{%raw%}/ubuntu/ {{release_name}}-updates main restricted unive
```

图: TUNA 过去的 Ubuntu 文档，两种语法混杂，难以维护



- 清除转译「黑魔法」，具体来说，使用正确的 MDX 解析器生成一份使用 TUNA 前端高级功能语法的 Markdown。
  - 保证兼容性
- 清除 TUNA 前端「黑魔法」，一种可能的路线是将 mirrorz-help 中的 MDX 渲染直接搬入 TUNA 前端，另一种可能路线是将 MDX 转译为使用 Vue 的 Markdown
  - shankerwangmiao 给了个使用 Vue 的原型
  - 在 Jekyll 渲染 MDX 时调用 Jekyll 插件从而调用 MDX 渲染器生成使用使用 Vue 的 HTML
- 如果采取第一种路线，可以将 mirrorz-help 中的 MDX 渲染提取成库，供其他镜像站前端使用
- (可选) 在 TUNA 前端实现 mirrorz-help 的高级功能，例如快速配置，命令拷贝，域名选择