

MirrorZ: 从镜像站到镜像站们

Zenithal

`hongren.zheng@tuna.tsinghua.edu.cn`

2021-09-25



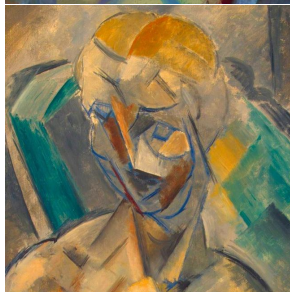
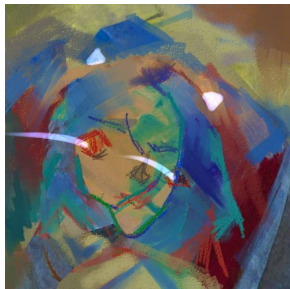
■ 关于我

- 头像是綾波レイ
- 曾经的头像是初音未来
- GitHub: ZenithalHourlyRate
- GPG: blog.zenithal.me/key
- 更多介绍在 GitHub 的 README 中

■ 在金枪鱼 (TUNA) 中

- 潜水摸鱼两三年
- 「(2020 年) 11 月, zenithal 开始加入到镜像站的维护中。」(引自 tuna.wiki/MirrorsHistory)

97 contributions in 2021 in Tsinghua University TUNA Association



镜像站维护，具体在做啥

- tuna/issues: 镜像请求，问题反馈
 - 打打杂、关关 issue
- tuna/mirror-web: 镜像站前端，帮助文档，合 Pull Request
- tuna/tunasync: 同步管理工具
- tuna/tunasync-scripts: 同步脚本
 - 例如修理 homebrew-bottles 同步脚本
- tuna/collection: 讲段子



Created an issue in tuna/collection that received 93 comments

Apr 16

🕒 DNS666事关全國數萬名網友的

Miao Wang, [16.04.21 14:42] 谁知道 .hath.network 是什么东西? Miao Wang, [16.04.21 14:44] 那为什么 666 DNS 从 https 上会收到非常非常多的这个域名的请求呢? Miao Wang, [16.04.21 14:44]...

93 comments

镜像站维护，好做吗

- `tuna/issues`: 镜像请求，问题反馈

- 打打杂、关关 issue
- 其实会参考隔壁镜像站的情况

 [ustclug](#) / [discussions](#) Unwatch ▾ 33

- `tuna/mirror-web`: 镜像站前端，帮助文档，合 Pull Request

- 参考隔壁镜像站的帮助文档

- `tuna/tunasync-scripts`: 同步脚本

- 例如修理 `homebrew-bottles` 同步脚本，具体工作内容就是照抄 USTC 的同步脚本

- workflow 其实有问题

- 遇到新问题，会遍历一次隔壁镜像站（USTC/SJTUG 等）
- 新镜像请求，有时候甚至不知道友镜有没有这个镜子，得遍历主页

镜像站维护带来的启发

- 镜像站管理员想要找东西，都要遍历各个镜像站，更何况是用户
- 如果用户找 rocky 镜像站的时候能有右边的面板，会方便很多
- 更进一步，如果在「换源」时自动选择最优镜像，那该多方便

rocky

4 ✓

NJU

rsync://msync.rockylinux.org/rocky/mirror/pub/rocky/

✓ Success | 2021-09-24 18:35:59 UTC+8
| 37 minutes ago

→ Next | 2021-09-25 18:35:59 UTC+8
| in 23 hours

462.55G

SJTUG - Zhiyuan

✓ Success | 2021-09-24 12:17:13 UTC+8
| 7 hours ago

USTC

rsync://msync.rockylinux.org/rocky/mirror/pub/rocky/

✓ Success | 2021-09-24 14:50:20 UTC+8
| 4 hours ago

→ Next | 2021-09-24 18:43:00 UTC+8
| 30 minutes ago

374.90 GiB

XTOM (EE)

rsync://mirror.netzwerke.de/rocky-linux/

✓ Success | 2021-09-24 15:45:34 UTC+8
| 3 hours ago

→ Next | 2021-09-24 19:43:00 UTC+8
| in 30 minutes

362.04 GiB

rockylinux

1 ✓

cqu

msync.rockylinux.org

✓ Success | 2021-09-24 18:54:55 UTC+8
| 18 minutes ago

unknown


镜像站们

- 基于上述情况，我和各镜像站的管理员、用户，开发了 MirrorZ（中文名暂为「镜像站们」）
- 如果用户找 rocky 镜像站的时候能有上页的面板，会方便很多
 - 前端：可供用户查询「常用软件」「镜像列表」「镜像站状态」
- 更进一步，如果在「换源」时自动选择最优镜像，那该多方便
 - 跳转后端

os	app	font
anthon		BFSU <ul style="list-style-type: none">2021.09.01 (x86_64, CLI-only)
archlinux		DGUT <ul style="list-style-type: none">2021.07.01 archlinux-2021.07.01-x86_64.iso2021.09.01 archlinux-2021.09.01-x86_64.iso
archlinuxarm		NJU <ul style="list-style-type: none">2021.09.01 (x86_64, CLI-only)
archriscv		NJUPT <ul style="list-style-type: none">2021.09.01 (x86_64, CLI-only)
armbian		SUSTech CRA <ul style="list-style-type: none">2021.09.01 (x86_64, CLI-only)


USTC
132 ✓ 4 1 1 3


WSYU
3


XJTU
40 ✓ 4 4

XTOM (EE)

URL: <https://mirrors.ustc.edu.cn>

Name: 中国科学技术大学开源软件镜像

Homepage: <https://lug.ustc.edu.cn>

Disk: 54T/60T

Issue: <https://github.com/ustclug/discussion>

Mirror Request: <https://github.com/ustclug/mirror-request>

Email: lug@ustc.edu.cn

Group: See <https://lug.ustc.edu.cn/wiki/lug>

AdoptOpenJDK

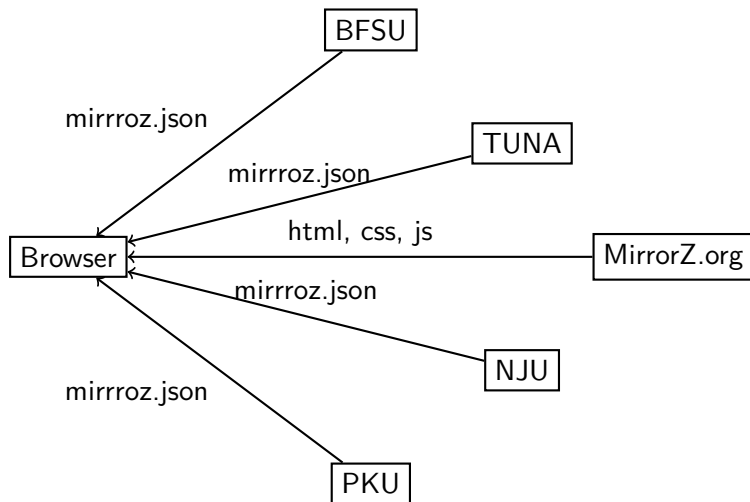
anthon

MirrorZ 名字的来历

- 本质原因：没有 mirrors 的域名（例如 mirrors.moe），我只能想一个别的名字
- Z 来自我的 ID：Zenithal
- Mirrors 的复数形式，与其用 Mirrorses，不如用 Mirrorz
 - Your next MirrorS is not MirrorS, nor MirrorSes, it's MirrorZ.
 - 你的下一个镜像站何必是镜像站
 - (跳转后端确实不是镜像站)
- 字母表的最后一个字母：A final site for Mirror sites.
- 最终决定因素：mirrorz.org 还没有被注册

- 各个站的前端、后端各不相同
- 同步管理器
 - tuna/tunasync: tunasync.json
 - ustclug/Yuki: meta
 - sjtug/mirror-clone: summary.json
 - ideal/mirror: task_status.json
- 首页是 Nginx autoindex: DOMParser 启动!
- 把上述数据转化为统一的 mirrorz.json, 即可被前端使用
- (该数据格式还被不少 MirrorZ 的下游项目使用, 后续会展开)

前端架构



联邦式前端



- 主站: mirrorz.org
- 教育网分站: mirrors.cnki.edu.cn
 - 只包含以 `edu.cn` 结尾的镜像站
- 各省份分站 mirrors.xx.cn, 例如 mirrors.bj.cn
 - 如果那个省有镜像站且被 `mirrorz` 收录
- DN42 站: mirror.z.dn42
 - 虽然里面只有一个镜像站 mirrors.nia.dn42

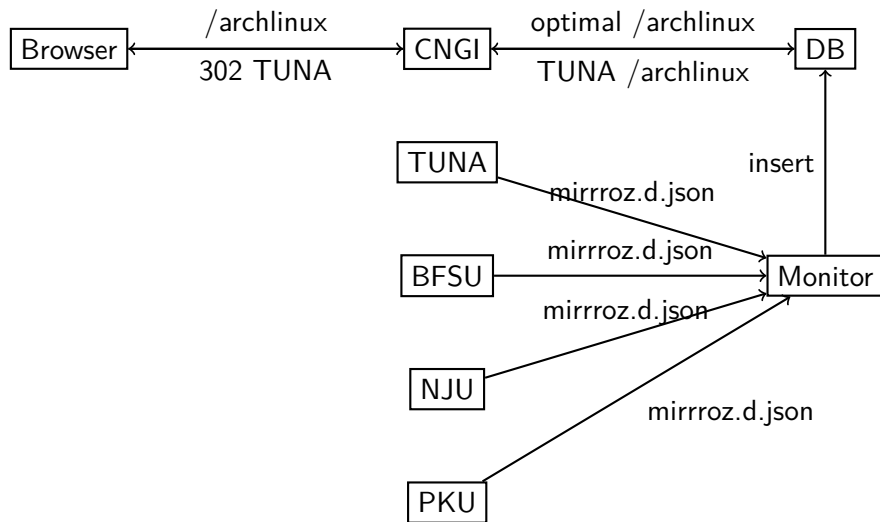
跳转后端：如何使用

- Arch Linux: Server =
`https://mirrors.cngl.edu.cn/archlinux/$repo/os/$arch`
- Debian: deb `https://mirrors.cngl.edu.cn/debian/ bullseye`
`main contrib non-free`
- Ubuntu: deb `https://mirrors.cngl.edu.cn/ubuntu/ focal`
`main restricted universe multiverse`
- CentOS/Fedora: baseurl=`https://mirrors.cngl.edu.cn`
- 以及更多软件源，不一而足。「镜像列表」中所列镜像基本可以使用该功能。

跳转后端：打分机制

- 跳转后端位于镜像站与用户中间，需要考虑两边的因素
- 镜像站端
 - 网络环境：接入点，IPv6/IPv4，HTTP/HTTPS，所在网段与 AS
 - 同步状态：各个镜像源的同步情况（同步落后/同步中）
- 用户端
 - 网络环境：IPv6/IPv4，HTTP/HTTPS，所在网段与 AS
 - 用户喜好（可选）：用户访问时提供喜好列表
 - 用户可以直接使用 `mirrors.cngi.edu.cn`
 - 也可使用 `ustc-tuna.mirrors.cngi.edu.cn` 提供其喜好列表
- `mirrorz.d.json`: 对 `mirrorz.json` 的一个扩展，包含节点的网络环境
- 打分机制：对网络环境的匹配情况、用户喜好、同步状态进行打分，占优者会被选择
- 可以用 `mirrors.cngi.edu.cn/archlinux?trace` 查看技术细节

跳转后端架构



跳转后端：一些指标

- 以上打分机制只适用于单次请求，是用户喜好、网络环境与同步状态等因素的折中与妥协
- 对于多次请求，我们提出以下指标
- 一致性
 - 对于用户单次 `apt-get update`，其跳转的镜像站要保持一致
 - 以保证其元数据的一致性，避免跳转到的两个镜像站的状态不一致导致用户本地元数据损坏
 - 实现方式：只需要对用户短时间内的请求做缓存
- 单调性
 - 对于用户连续两次 `apt-get update` 请求，新请求获得的版本不会比旧请求的版本更低
 - 避免本地软件包降级
 - 实现方式：较难实现或无法实现
- 还有一些性能方面的指标，略过不表

跳转后端：部署

教育网分站：mirrors.cngi.edu.cn

- 跳转后端服务是教育网分站的特色服务

```
[mirrorz@mirrorz mirrorz.d.json]$ ls -1  
bfsu.d.json  
bjtu.d.json  
hit.d.json  
nano.d.json  
neo.d.json  
nju.d.json  
nwafu.d.json  
sustech.d.json  
ustc.d.json
```

镜像站们：下游项目

- 有了 mirrorz.json ，可以玩很多花样
- oh-my-mirrorz：测速脚本！
 - mirrorz.org/oh-my-mirrorz.py
 - mirrors.cngi.edu.cn/oh-my-mirrorz.py
- mirrorz-legacy：静态页面
 - mirrorz.org/_/
 - mirrors.cngi.edu.cn/_/
- mirrorz-monitor：镜像状态监控
 - mirrorz.org/monitor
 - mirrors.cngi.edu.cn/monitor

ABBR	SPEED
PKU	356.71 MiB/s
BFSU	299.22 MiB/s
TUNA	150.06 MiB/s
SJTUG - Siyuan	99.72 MiB/s
SJTUG - Zhiyuan	77.53 MiB/s
USTC	76.72 MiB/s
NEUSOFT	67.35 MiB/s
SUSTech CRA	58.14 MiB/s
NJU	51.33 MiB/s
BUPT	28.81 MiB/s
CQUPT	13.75 MiB/s
NJUPT	11.79 MiB/s
HIT	10.46 MiB/s

镜像站们：后续计划

- 更多镜像站的参与！
 - `mirrorz.org` 包含的大多是国内高校镜像站
 - 仅校内的镜像站也可以参与到跳转后端中
- 更多的下游项目
 - 例如镜像帮助乃至「自动换源」，我们还没有启动 `mirrorz-help` 项目
 - 各个镜像站的镜像帮助其实非常相似，只需要做变量替换
- 更多的用户：跳转后端的功能、性能如何，还需要测试与反馈
- 项目地址，欢迎点星星和贡献代码！
 - 前端：`github.com/mirrorz-org/mirrorz`
 - 跳转后端：`github.com/mirrorz-org/mirrorz-302`