



fontconfig & libpinyin 简介

Presented by
Peng Wu
Red Hat

今日主题

1. fontconfig 简介
2. libpinyin 简介
3. libpinyin 解密

fontconfig 简介

字体渲染组件

- Gtk+
- Pango
- Cairo
- Fontconfig
- Freetype

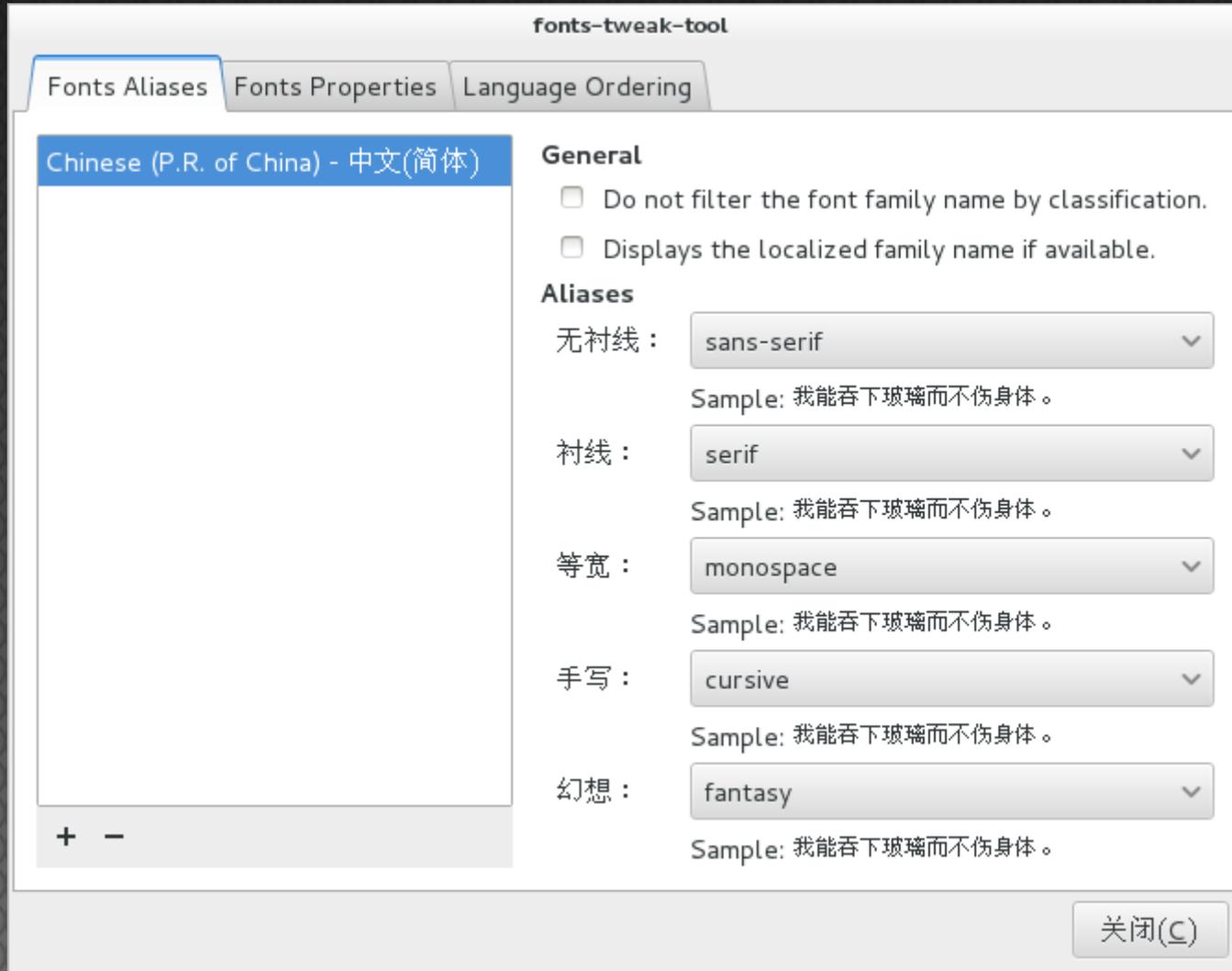
Fontconfig 目标

- 根据查询条件选择最好的字体
 - fc-match
- 根据查询条件列出匹配的字体
 - fc-list

Fedora 上的 fontconfig 配置

- 为每一种语言提供一个默认字体
 - 参考
 - <http://tagoh.fedorapeople.org/fonts/f18-defaultfonts.html>
- 绝大部分字体都使用统一的 fontconfig 配置格式
 - <http://fedoraproject.org/wiki/Packaging:FontsPolicy>
 - 示例
 - `/usr/share/fontconfig/templates/`

fonts-tweak-tool



libpinyin 简介

libpinyin 项目信息

- 主页：<https://github.com/libpinyin>
 - 包含 libpinyin 和 ibus-libpinyin 两个子项目
- 下载镜像：
<http://sourceforge.net/projects/libpinyin/>
- 简介：Library to deal with pinyin

libpinyin 简介

- libpinyin
 - 基于 n-gram 自然语言处理模型的输入法后端
 - 已经进入 Fedora 18 中
 - 仍然在改进当中 ;)
- libpinyin 的前端
 - ibus-libpinyin
 - fcitx-libpinyin

假想的拼音输入法



记录所有的拼音 / 句子到 2TB 关系数据库中。

- 优点
 - 接近 100% 的正确率
- 缺点
 - 没有那么大的硬盘
 - 也没有那么强悍的 CPU
- 使用自然语言处理的技术来达到正确率和资源使用量的折中

libpinyin 解密

开源输入法数学模型

- Uni-gram:
 - scim-pinyin
- N-gram:
 - libpinyin

scim-pinyin 模型示例

例子 : zhong'guo'ren

$P(\text{中国人} | \text{zhong'guo'ren})$

$= P(\text{中国人})$

$= P(\text{中国}) * P(\text{人})$

$= 0.001 * 0.001$

$= 1e-6$

scim-pinyin 模型示例 (继续)

$P(\text{种果人} | \text{zhong'guo'ren})$

$= P(\text{种果人})$

$= P(\text{种果}) * P(\text{人})$

$= 0.0001 * 0.001$

$= 1e-7$

$< 1e-6 = P(\text{中国人} | \text{zhong'guo'ren})$

所以最终结果为中国人。

libpinyin 模型示例



例子：zhong'guo'ren

$$\begin{aligned} &P(\text{中国人} | \text{zhong'guo'ren}) \\ &= P(\text{中国人}) * P(\text{zhong'guo'ren} | \text{中国人}) \\ &= P(\text{中国}) * P(\text{人} | \text{中国}) * P(\text{zhong'guo} | \text{中国}) * P(\text{ren} | \text{人}) \\ &= 0.01 * 0.1 * 0.7 * 0.5 \\ &= 3.5 * 10^{-4} \end{aligned}$$

libpinyin 模型示例 (继续)



$$\begin{aligned} & P(\text{种果人} | \text{zhong'guo'ren}) \\ &= P(\text{种果人}) * P(\text{zhong'guo'ren} | \text{种果人}) \\ &= P(\text{种果}) * P(\text{人} | \text{种果}) * P(\text{zhong'guo} | \text{种果}) * P(\text{ren} | \text{人}) \\ &= 0.01 * 0.01 * 0.8 * 0.5 \\ &= 4.0 * 10^{-5} \\ &< 3.5 * 10^{-4} = P(\text{中国人} | \text{zhong'guo'ren}) \end{aligned}$$

所以最终结果为中国人。

libpinyin on Fedora 19 开发计划

- 允许用户导入第三方词库
- 更多 ...

总结

- 请使用 `fonts-tweak-tool` 来定制自己的桌面字体
- 将会在 Fedora 19 继续改进 `libpinyin`，提供更好的输入体验

Demo

问题？



Contact:

pwu at fedoraproject dot org

License statement goes here. See https://fedoraproject.org/wiki/Licensing#Content_Licenses for acceptable licenses.