

语言支持：字体与文本转语音

我的博客现在支持九种语言：日语 (ja)、西班牙语 (es)、印地语 (hi)、简体中文 (zh)、英语 (en)、法语 (fr)、德语 (de)、阿拉伯语 (ar) 和繁体中文 (hant)。你可以访问 <https://lzwjava.github.io> 查看网站。

在计算机环境中处理多种语言时，有几个方面需要考虑。

字体处理

不同的语言需要特定的字体来正确显示，尤其是在使用 LaTeX 生成 PDF 时。以下 Python 代码展示了如何根据操作系统和语言选择合适的字体：

```
if platform.system() == "Darwin":
    if lang == "hi":
        CJK_FONT = "Kohinoor Devanagari"
    elif lang == "ar":
        CJK_FONT = "Geeza Pro"
    elif lang in ["en", "fr", "de", "es"]:
        CJK_FONT = "Helvetica"
    elif lang == "zh":
        CJK_FONT = "PingFang SC"
    elif lang == "hant":
        CJK_FONT = "PingFang TC"
    elif lang == "ja":
        CJK_FONT = "Hiragino Sans"
    else:
        CJK_FONT = "Arial Unicode MS"
else:
    if lang == "hi":
        CJK_FONT = "Noto Sans Devanagari"
    elif lang == "ar":
        CJK_FONT = "Noto Naskh Arabic"
    elif lang in ["en", "fr", "de", "es"]:
        CJK_FONT = "DejaVu Sans"
    elif lang == "zh":
        CJK_FONT = "Noto Sans CJK SC"
    elif lang == "hant":
        CJK_FONT = "Noto Sans CJK TC"
    elif lang == "ja":
        CJK_FONT = "Noto Sans CJK JP"
    else:
```

```

CJK_FONT = "Noto Sans"

command = [
    'pandoc',
    input_markdown_path,
    '-o', output_pdf_path,
    '-f', 'markdown',
    '--pdf-engine', 'xelatex',
    '-V', f'romanfont={CJK_FONT}',
    '-V', f'mainfont={CJK_FONT}',
    '-V', f'CJKmainfont={CJK_FONT}',
    '-V', f'CJKsansfont={CJK_FONT}',
    '-V', f'CJKmonofont={CJK_FONT}',
    '-V', f'geometry:{GEOMETRY}',
    '-V', 'classoption=16pt',
    '-V', 'CJKoptions=Scale=1.1',
    '-V', 'linestretch=1.5'
]

```

需要注意的是，这个解决方案并不完美。例如，代码块注释中的印地语文本可能无法按预期渲染。

文本转语音

我使用 Google 文本转语音来生成博客文章的音频版本。以下代码片段展示了我如何为文本转语音引擎选择合适的语言代码：

```

if filename.endswith('-zh.md'):
    language_code = "cmn-CN"
    voice_language_code = "cmn-CN"
else:
    language_code = "en-US"
    voice_language_code = "en-US"

text_to_speech(
    text=article_text,
    output_filename=output_filename,
    task=task,
    language_code=language_code,
    dry_run=dry_run,
    progress=progress
)

```

目前，音频生成仅支持中文和英文内容。要扩展支持其他语言，必须配置相应的语言代码。