

Soporte de idiomas: Fuentes y Texto a Voz

Mi blog ahora admite nueve idiomas: japonés (ja), español (es), hindi (hi), chino (zh), inglés (en), francés (fr), alemán (de), árabe (ar) y chino tradicional (hant). Puedes encontrar el sitio en <https://lzwjava.github.io>.

Cuando se trata de múltiples idiomas en un entorno informático, hay varios aspectos que deben considerarse.

Manejo de Fuentes

Diferentes idiomas requieren fuentes específicas para su correcta representación, especialmente al generar PDFs con LaTeX. El siguiente código en Python muestra cómo seleccionar las fuentes apropiadas según el sistema operativo y el idioma:

```
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'
]

```

Es importante tener en cuenta que esta solución no es perfecta. Por ejemplo, el texto en hindi dentro de los comentarios de los bloques de código podría no renderizarse como se espera.

Texto a Voz

Utilizo Google Text-to-Speech para generar versiones de audio de mis publicaciones en el blog. El siguiente fragmento de código ilustra cómo selecciono el código de idioma adecuado para el motor de texto a voz:

```

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  
    )
```

Actualmente, el audio se genera para contenido en chino e inglés. Para extender el soporte a otros idiomas, se deben configurar los códigos de idioma correspondientes.