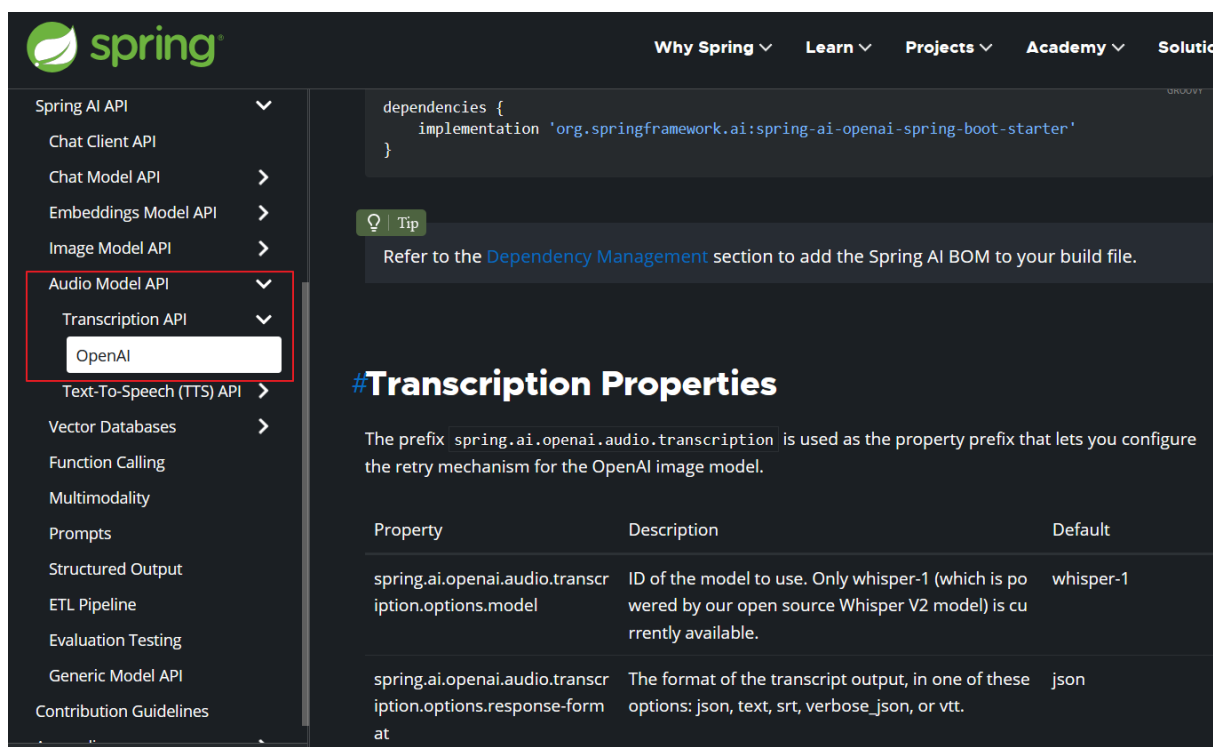


Day9 - 做一個雲端字幕產生器

相信很多人知道 OpenAI 開源了 Whisper 模型，網路上也很多人製作本機端的字幕產生程式，凱文大叔也試過，只能說慘不忍睹，由於電腦沒有獨立顯卡，不到 30 分鐘的影片竟然一個小時還沒完成

這時就很適合用平台的算力來協助了，音頻轉譯在 Spring AI 的文件中目前只支援 OpenAI



The screenshot shows the Spring AI documentation website. On the left sidebar, the 'Audio Model API' section is highlighted with a red box, and 'OpenAI' is selected in the dropdown menu. The main content area displays a code snippet for dependencies, a tip about adding the Spring AI BOM, and a section titled '#Transcription Properties'. This section explains the prefix 'spring.ai.openai.audio.transcription' and includes a table of properties.

Property	Description	Default
spring.ai.openai.audio.transcription.options.model	ID of the model to use. Only whisper-1 (which is powered by our open source Whisper V2 model) is currently available.	whisper-1
spring.ai.openai.audio.transcription.options.response-format	The format of the transcript output, in one of these options: json, text, srt, verbose_json, or vtt.	json

原本想說可以用 Groq 省錢，不過凱文大叔實際測試發現 Groq 闖割了很多功能，只能產生純文字，字幕最重要的時間戳記卻無法產生，若單純產生文字稿還是能用，需要時間戳記就只能花點小錢了（22mb 四分多鐘影片約 0.03 美金）

程式碼跟 Day7 有點像，主要就是上傳影片或聲音檔，再透過 OpenAiAudioTranscriptionModel 生成文字，不過設定檔需要調整一下

```
spring:
  ai:
    openai:
      api-key: ${OPENAI_APIKEY}
      audio:
        transcription:
```

```

        options:
            model: whisper-1
    servlet:
        multipart:
            max-file-size:
                25MB
            max-request-size:
                25MB

```

與 transcription 有關的設定放在 `openai.audio.transcription` 這邊我們只先設定 `model`，其它我們在程式中設定，比較方便進行調整

另外有注意到下方多了 `multipart` 設定嗎？`MultipartFile` 預設最大只能傳 1MB 所以我們必須把每次單檔最大（`max-file-size`）跟每次請求最大（`max-request-size`）都設為 25MB

為什麼設成 25MB 呢？因為這是 `whisper` 的限制，超過就是 `openai` 會回錯誤訊息了

接著看程式吧

```

private final OpenAiAudioTranscriptionModel transcriptionModel;

@PostMapping("/vtt")
public String image(MultipartFile file) {
    OpenAiAudioTranscriptionOptions transcriptionOptions =
        OpenAiAudioTranscriptionOptions.builder()
            .withTemperature(0f)
            .withResponseFormat(TranscriptResponseFormat.SRT)
            .build();
    AudioTranscriptionPrompt transcriptionRequest =
        AudioTranscriptionPrompt.builder()
            .withFile(file)
            .build();
    AudioTranscriptionResponse response = transcriptionModel.transcribe(
        transcriptionRequest, transcriptionOptions);
    return response.getResult().getOutput();
}

```

其中比較重要的參數是 `withResponseFormat(TranscriptResponseFormat.SRT)`

這會直接影響輸出的格式，`srt` 是最常見的字幕檔，除了文字外還會標示出現的時間，其他格式大家有興趣也可以測試看看

下面是轉出的檔案內容，轉譯的內容除了幾個英文字念不標準翻錯以外，中文幾乎都沒問題，凱文大叔覺得最厲害的是一句話有中英文夾雜它也能辨識出來 字幕檔.srt

1

00:00:00,000 --> 00:00:05,600

在學習程式或是求職時, 是否遇過以下狀況

2

00:00:05,600 --> 00:00:10,400

學完Java仍不知如何開發應用程式

3

00:00:10,400 --> 00:00:17,100

在找Java相關工作, 結果超過一半都要求會使用SpringBoot框架

4

00:00:17,100 --> 00:00:23,600

學完了SpringBoot, 卻還是不知道如何整合各個模組完成一個專案

另外這模組也很適合做為會議紀錄，不管是線上會議還是錄音筆，他都能輕易的轉成文字，之後都能作為 RAG 的資料來源，當作企業 KM 系統的一環

回顧一下今天學到的內容:

- 透過 Whisper 可以輕鬆的將影片或音檔轉成文字
- application.yml 以及 Options 的設定方式
- MultipartFile 上限大小調整