

Chmurowe usługi AI

TTS & STT - ćwiczenia

Piotr JANKOWSKI

1 Rozpoznawanie i synteza mowy

1.1 Wstęp

1.2 Wymagania

Wymagania poniżej dotyczą wszystkich zadań (patrz niżej).

1. Klucze uwierzytelniające nie mogą w żadnym momencie być umieszczone w kodzie aplikacji.
2. Poprawność działania kodu powinna być sprawdzana testem integracyjnym.

Dokładne wymagania znajdują się w opisie poszczególnych zadań.

1.3 Zadania

1.3.1 Rozpoznawanie mowy z Azure STT

Celem aplikacji jest wykonanie transkrypcji z pliku dźwiękowego do tekstu.

Uzupełnij klasę `AzureStt` tak, aby test `AzureSttTest` wykonywał się poprawnie.

Do uwierzytelnienia w usłudze wykorzystaj klucz API. Wartość klucza `API SPEECH_KEY` oraz konfigurację regionu w `Azure SPEECH_REGION` prześlij do aplikacji za pomocą zmiennych środowiskowych jak w przykładzie:

```
private static final String SPEECH_KEY = Objects.requireNonNull(  
    System.getenv("SPEECH_KEY"),  
    "Missing environment variable SPEECH_KEY");
```

Konkretne wartości tych zmiennych otrzymasz od prowadzącego w trakcie zajęć.

Przekazane wartości możesz użyć do uwierzytelnienia dzięki metodzie `fromSubscription` z klasy `SpeechConfig` (`com.microsoft.cognitiveservices.speech`).

Konfigurację źródła dźwiękowego wykonaj przy użyciu metody `fromWavFileInput` z klasy `AudioConfig` (`com.microsoft.cognitiveservices.speech.audio`).

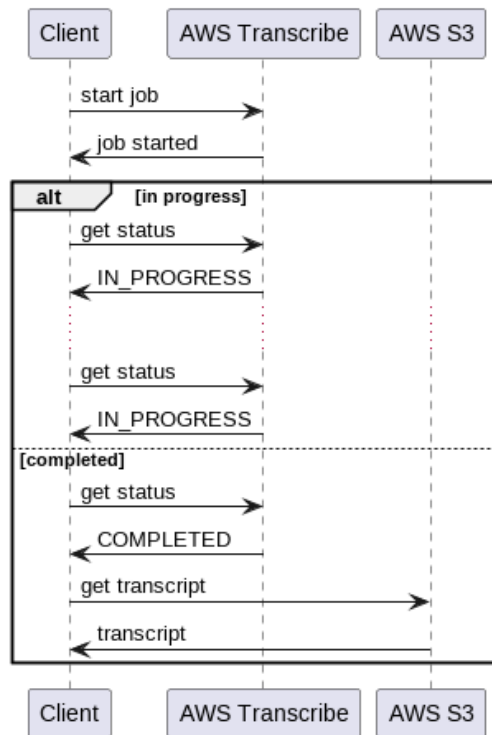
Tak przygotowaną konfigurację prześlij jako parametry do konstruktora klasy `SpeechRecognizer` (`com.microsoft.cognitiveservices.speech`). Następnie użyj tej klasy do wywołania usługi w Azure.

Gdy test będzie wykonywał się poprawnie, napisz kolejny test, który sprawdzi wykonanie transkrypcji pliku `src/test/resources/Cytat-Byc-moze-szkola_Eco.wav`.

1.3.2 Rozpoznawanie mowy jako batch job (Amazon Transcribe)

1.3.2.1 Wprowadzenie W tym zadaniu wykorzystamy przetwarzanie batchowe w Amazon Transcribe. Schemat poniżej przedstawia ogólną koncepcję:

Do usługi Amazon Transcribe zostaje wysłane żądanie uruchomienia nowego zadania przetwarzania. Jako parametry przekazane są: nazwa zadania `jobName` oraz lokalizacja pliku źródłowego w usłudze Amazon S3 `s3Source`. W odpowiedzi otrzymujemy potwierdzenie utworzenia zadania.



Rysunek 1: Diagram sekwencji przetwarzania batchowego w Amazon Transcribe

Odpytanie usługi Amazon Transcribe o status utworzonego zadania może zwrócić:

- **QUEUED** - gdy zadanie oczekuje na rozpoczęcie
- **IN_PROGRESS** - gdy zadanie jest w trakcie realizacji
- **FAILED** - gdy zadanie nie wykonało się poprawnie
- **COMPLETED** - gdy zadanie zakończyło się sukcesem

Patrz `com.amazonaws.services.transcribe.model.TranscriptionJobStatus`.

Po zakończeniu zadania z sukcesem, klient otrzymuje lokalizację pliku wynikowego w usłudze Amazon S3, skąd może go pobrać.

1.3.2.2 Zadania Zaimplementuj metodę `runTranscribeJob` w klasie `TranscribeJob` tak, aby startowała process batchowy transkrypcji dźwięku w usłudze Amazon Transcribe. Skorzystaj z metody prywatnej `getAmazonTranscribe`.

WAŻNE: użyj nazwy bucketa S3 `ug2223` (tylko do tego bucketa będą uprawnienia).

WAŻNE: użyj tzw. *output key* jako wyniku wywołania metody `getOutputKey`.

WAŻNE: użyj unikatowej nazwy *transcription job*: *pierwsza litera imienia + nazwisko + licznik*: `jkowalski01`. W kolejnym wywołaniu z tą samą nazwą *transcription job* może być zwracany błąd, że taki *job* już istnieje. Zwiększ wówczas licznik, np. `jkowalski02` itd.

Następnie zaimplementuj metodę `checkJobStatus` w tej samej klasie. Jej celem jest pobranie stanu utworzonego zadania (patrz wyżej). Również skorzystaj z `getAmazonTranscribe`.

Ostatecznie zaimplementuj metodę `getTranscript`, której celem jest pobranie wyniku przetwarzania pliku dźwiękowego. Ten wynik przetwarzania będzie plikiem JSON umieszczonym na S3 w ścieżce zgodnej z użytymi *transcription job nam* oraz *output key*. Np. dla *transcription job* `jkowalski02` wynik będzie dostępny pod ścieżką `ug2223/jkowalski02/output.json`. Ponownie wykorzystaj metodę `getOutputKey`. Upewnij się również, że pobierasz plik wynikowy tylko wtedy, gdy status *joba* to **COMPLETED**. W tym zadaniu skorzystaj z metody `getAmazonS3`.

(Opcjonalnie) Dodatkowo efekt działania aplikacji można sprawdzić używając interfejsu linii poleceń AWS (`aws-cli`)

Wylistowanie pliku wynikowego:

```
$ aws s3 ls ug2223/jkowalski02/output.json
2022-11-02 00:33:48          9872 output.json
```

Skopiowanie pliku wynikowego z S3 do lokalnego folderu:

```
$ aws s3 cp s3://ug2223/jkowalski02/output.json ./
download: s3://ug2223/jkowalski02/output.json to ./output.json
```

1.3.2.3 Debugowanie Jeśli zajdzie taka potrzeba, możesz włączyć dodatkowe informacje w logu. W tym celu zmodyfikuj plik `src/main/resources/log4j2.xml`. W szczególności dla AWS zmień

```
<Logger name="org.apache.http.wire" level="WARN" />
```

na

```
<Logger name="org.apache.http.wire" level="DEBUG" />
```

1.3.3 (Opcjonalnie)

1.3.3.1 Transkrypcja w czasie rzeczywistym (Azure) W zadaniu z Azure Speech-to-text zamień `AudioConfig.fromWavFileInput()` na `AudioConfig.fromDefaultMicrophoneInput()`. Usługa powinna wykonać transkrypcję głosu z mikrofonu komputera, na którym uruchomiony zostanie program.

1.3.4 Text to speech z Google Cloud

Celem zadania jest wygenerowanie pliku dźwiękowego (mp3) na podstawie zadanego tekstu przy użyciu usługi Google Cloud Text-to-Speech.

Zaimplementuj metodę `textToSpeech` w klasie `GcpTts`. Wykorzystaj metodę `synthesizeSpeech` z klasy `TextToSpeechClient`. Jako dane wejściowe użyj dowolnego krótkiego tekstu w języku angielskim.

Następnie wykorzystaj plik SSML do wygenerowania dźwięku z dodatkowymi parametrami. Aby załadować zawartość pliku, użyj metody `getSsml()`.

1.3.4.1 (Opcjonalnie) Poeksperymentuj z innymi konfiguracjami zwracanymi przez metodę `getVoiceSelectionParams`, np. z innymi językami i głosami.

Poeksperymentuj z innymi ustawieniami w pliku SSML. Patrz również dokumentacja SSML w GCP Text-to-Speech.

1.3.4.2 (Opcjonalnie) Porównaj sposoby implementacji w różnych językach programowania na stronie z dokumentacją do usługi Cloud Text-to-Speech.

1.4 Źródła

Użyte na podstawie licencji Creative Commons Attribution 3.0 Unported

https://commons.wikimedia.org/wiki/File:Martin_Sheen_BBC_Radio4_Desert_Island_Discs_3_April_2011_b00zzn2c.flac

https://commons.wikimedia.org/wiki/File:Cytat-Byc-moze-szkola_Eco.ogg

https://commons.wikimedia.org/wiki/File:Valerie_Beral_in_The_Life_Scientific_b01qdw1k.flac