


Speechly

 Speechly	Voice Interface API	
	Systeemvereisten:	Browser, javascript
	Ontwikkeld door:	Speechly
Commerciële licentie	Contactpersoon:	Bert.Vanhalst@smals.be

Functionaliteiten

[Speechly](#) laat toe om spraak toe te voegen aan een toepassing, bijvoorbeeld voor het ingeven van gegevens of zoeken op basis van de stem. Daartoe biedt het een systeem voor spraakherkenning en *natural language understanding* (NLU): het herkennen van *intents* (wat wil de gebruiker?) en entiteiten (parameters zoals een datum of getal). Op vlak van talen zijn momenteel enkel Engels en Fins ondersteund. Enkele andere talen staan op de roadmap: Frans, Spaans, Duits en Japans. Nederlands zou rond Q2 2023 ondersteund worden.

Een online dashboard (of command line interface) laat toe om de NLU te configureren zodat intents en entiteiten kunnen herkend worden. Dat gebeurt op basis van een specifieke Markdown-achtige syntax, de Speechly Annotation Language (SAL).

De oplossing heeft verschillende [deployment modellen](#): SaaS, on-premise en on-device, wat het aantrekkelijk maakt voor scenario's waarbij gegevensbescherming en privacy belangrijk zijn.

Om spraak toe te voegen aan een toepassing biedt Speechly verschillende integratiemogelijkheden. Er zijn [client libraries](#) beschikbaar voor web browser (Javascript), React, iOS (Swift) en Android (Kotlin). Daarnaast zijn er twee gRPC [API's](#) voor lower level integratie: een streaming API en een batch API. Enkele [kant-en-klare visuele interface-elementen](#) maken het eenvoudig om spraakfunctionaliteit toe te voegen aan een toepassing: een *push-to-talk knop* om de spraakinput te starten, een *transcript drawer* voor het in realtime tonen van het speech-to-text transcript en de herkende entiteiten, en een *intro pop-up* voor het tonen van een boodschap bij de vraag naar toelating voor het gebruik van de microfoon.

Conclusies & Aanbevelingen

Speechly is een interessante tool om spraak als een feature toe te voegen aan een toepassing. De gebruikerservaring is zeer goed. Het is snel en eenvoudig op te zetten en te integreren in een toepassing. Maar we stellen vast dat de herkenning van entiteiten niet zo intelligent is en de voorbeeldzinnen vrij strikt moeten gevolgd worden in de spraak-input. We hebben dan ook wat voorbehoud bij het gebruik van Speechly voor complexere scenario's. Daarenboven is de taalondersteuning momenteel heel beperkt (enkel Engels en Fins). De on-premise en on-device deployment opties zijn interessant in het kader van gegevensbescherming en privacy.

Testen & Resultaten

Als test gaan we aan de slag met de gratis SaaS-versie van Speechly. Als scenario kiezen we een Dimona-aangifte die één intent vertegenwoordigt en 4 parameters kan bevatten: datum, startuur, einduur en persoonsnaam. Allereerst maken we in het online dashboard een configuratie aan. We maken de 4 entiteiten aan door ze een naam te geven en een type toe te kennen: date(Date), startTime(Time), endTime(Time) en worker(Person name). Daarna stellen we de trainingsdata op voor het machine learning model. De syntax laat toe om beknopt en overzichtelijk templates toe te voegen van voorbeeldzinnen:

```
name = $SPEECHLY.PERSON_NAME  
worker_phrase = [for $name(worker) | {the} worker {name} {is} $name(worker)]
```

Vervolgens moet het model getrained en gedeployed worden. Die fase neemt al snel enkele minuten in beslag met uitschieters tot 13 minuten, wat toch wel lang is, zelfs voor een iets complexer model.

Na het trainen van het model kunnen we de configuratie effectief integreren in een toepassing. We gaan aan de slag met de kant-en-klare web componenten en tonen de resultaten in een eenvoudig webformulier op basis van Javascript code. Door het instellen van de parameters van de UI elementen en wat finetuning in de Javascript code slagen we erin om de gewenste gebruikerservaring te bekomen, met name een microknop die je ingedrukt houdt zolang je iets inspreekt en het tonen van de parameters op een streaming manier, d.w.z. naargelang ze ingesproken worden.

... A DECLARATION FOR JOHN FOR 2022-10-19 FROM 06:00 TO 22:00

Worker name:
Date:
Start time:
End time:



De templates uit de configuratie worden via *template expansion* omgezet in voorbeeldzinnen waarop het model getrained wordt. Die template expansion is niet exhaustief, wat betekent dat het kan voorvallen dat het model niet getrained wordt op bepaalde combinaties of volgordes van parameters. Het gevolg is dat bepaalde logica verloren gaat en bepaalde entiteiten dus niet goed herkend zullen worden. Een eenvoudiger model met minder opties en combinaties in de voorbeeldzinnen lost dit probleem weliswaar op, maar beperkt de mate waarin varianten op uitdrukkingen herkend worden.

Het systeem is uit zichzelf niet zo slim in het omgaan met variante uitdrukkingen. Zo wordt op basis van de voorbeeldzin “the start time is 6am” het uur in de uitdrukking “6am is the start time” niet correct herkend, terwijl we dit toch zouden verwachten bij een machine learning model.

De spraakherkenning (omzetting van spraak naar tekst) op zich ervaren we dan weer wel als heel snel en performant.

Gebruiksvoorwaarden & Budget

Speechly is gratis uit te proberen, met een beperking van 50 audio-uur per maand. Het Professional plan biedt o.a. meer verbruik en ondersteuning. Het Enterprise plan voegt daar customisatie en meer deployment opties (on-premise, on-device) aan toe. Meer informatie is te vinden op de [pricing pagina](#).