

Dit project is een automatiseringspipeline voor nieuwsverzameling, vertaling, en publicatie. Het project maakt gebruik van verschillende technologieën, waaronder een RSS-feed voor nieuwsaggregatie, Natural Language Processing (NLP) voor samenvattingen en vertalingen via de Cohere AI API. Het doel is om nieuwsartikelen te verzamelen, automatisch samen te vatten en op een gebruiksvriendelijke manier te presenteren via een eenvoudige website of WordPress.

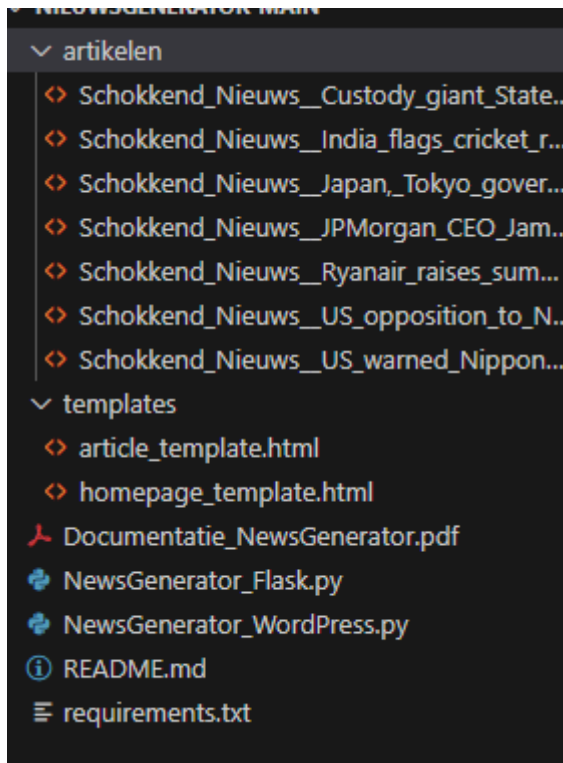
Documentatie - Nieuws Automatisering Pipeline

door Goudantov Bers en Van Camp Loïc.

Inhoud

Projectstructuur.....	2
Technologieën en API's.....	2
RSS-Feeds	2
Cohere API	2
Flask.....	2
BeautifulSoup	2
Functionaliteiten.....	3
Nieuwsaggregatie	3
Volledige Artikelinhoud ophalen	3
Automatische Samenvattingen	3
Opslag als HTML.....	3
Website.....	3
Flask Webserver	3
Hoe te gebruiken.....	4
Vereisten.....	4
Stappen om het project uit te voeren	4
1. Start de Flask-app	4
2. Kies artikelen of genereer nieuwe	4
3. Genereer nieuwe samenvattingen	4
Ethische Overwegingen	4
Accuraatheid van Informatie	4
Copyright en Bronvermelding	4
Transparantie	4
Toekomstige Verbeteringen	5
Meer geavanceerde samenvattingstechnieken	5
Verbeterde UI	5
Meertalige ondersteuning	5
Extra nieuwsbronnen	5
besluit	5

Projectstructuur



Technologieën en API's

RSS-Feeds

Het project maakt gebruik van RSS (Really Simple Syndication) om nieuwsartikelen te verzamelen van externe bronnen. RSS-feeds bieden een gestructureerd formaat om artikelen te verzamelen zonder dat de hele webpagina gescrapet hoeft te worden.

Cohere API

Voor het genereren van samenvattingen en het vertalen van teksten wordt de Cohere API gebruikt. Deze API voert NLP-taken uit zoals:

- Samenvatten van nieuwsartikelen.
- Genereren van inzichten in de implicaties van het nieuws (politiek, economisch, enz.).
- Genereren van gepaste titel op basis van de context van de artikel.

Flask

De webapplicatie is gebouwd met Flask, een lichtgewicht Python webframework. Flask biedt de mogelijkheid om dynamische inhoud te serveren via templates en biedt een intuïtieve manier om routes te beheren voor de website.

BeautifulSoup

Voor het scrapen van volledige artikelen van externe websites gebruiken we BeautifulSoup, een Python-bibliotheek voor het parsen van HTML en XML. Hiermee kunnen we de volledige tekst van een artikel extraheren.

Functionaliteiten

Nieuwsaggregatie

De pipeline verzamelt nieuwsartikelen van een opgegeven RSS-feed. De verzamelde artikelen bevatten informatie zoals:

- Titel
- Link naar het volledige artikel
- Beschrijving
- Publicatiedatum
- Auteur

Volledige Artikelinhoud ophalen

Naast de samenvattingen haalt het systeem ook de volledige inhoud van elk artikel op. Dit wordt gedaan door middel van web scraping met BeautifulSoup.

Automatische Samenvattingen

Met behulp van de **Cohere API** genereert het systeem automatisch samenvattingen van de artikelen, samen met een diepere analyse van de mogelijke implicaties (economisch, politiek, enz.).

Opslag als HTML

Verwerkte artikelen worden opgeslagen als individuele HTML-bestanden in de map `artikelen`. Deze bestanden bevatten de artikeltekst, de AI-gegenereerde samenvatting, en metadata zoals auteur en publicatiedatum.

Website

De gegenereerde artikelen worden gepresenteerd op een eenvoudige website. De homepage toont een lijst met artikelen, en elke artikelpagina bevat:

- Titel
- Volledige artikelinhoud
- AI-samenvatting
- Metadata (auteur, publicatiedatum)

Flask Webserver

Met de Flask-webserver kun je lokaal een website draaien die de gegenereerde artikelen toont. Deze server is eenvoudig op te starten vanuit het hoofdsript `NewsGenerator.py`.

Hoe te gebruiken

Vereisten

Zorg ervoor dat je de volgende technologieën hebt geïnstalleerd:

- Python 3.x
- Pip (Python package installer)

Installeer de vereiste pakketten met het volgende commando:

```
pip install -r requirements.txt
```

Stappen om het project uit te voeren

1. Start de Flask-app of WordPress

Voer het volgende commando uit in de terminal om de Flask-server te starten:

```
python NewsGenerator_WordPress.py / NewsGenerator_Flask.py
```

De server zal draaien op `http://127.0.0.1:5000/`. Navigeer hierheen om de gegenereerde artikelen te bekijken. Of navigeer naar het WordPress pagina (<https://yapnews.meubel-centrum.be/>) om het inhoud te kunnen zien.

2. Kies artikelen of genereer nieuwe

Bij het starten van de server krijg je de optie om bestaande artikelen te gebruiken of nieuwe artikelen te genereren van een RSS-feed. Volg de instructies in de terminal om je keuze te maken.

3. Genereer nieuwe samenvattingen

Als je ervoor kiest om nieuwe artikelen op te halen, zal het systeem de RSS-feed uitlezen, de volledige inhoud van de artikelen scrapen en AI-samenvattingen genereren.

Ethische Overwegingen

In dit project zijn we ons bewust van enkele belangrijke ethische overwegingen.

Accuraatheid van Informatie

- De gegenereerde AI-samenvattingen worden naast het originele artikel weergegeven, zodat de gebruiker kan controleren of de samenvatting accuraat is.

Copyright en Bronvermelding

- Elk artikel bevat duidelijke bronvermeldingen en metadata die de oorspronkelijke auteur en publicatiedatum weergeven. Hiermee zorgen we ervoor dat auteurs de erkenning krijgen die ze verdienen.

Transparantie

- Het is voor de gebruiker duidelijk welke delen van het artikel door AI zijn gegenereerd. Dit verhoogt de transparantie en geeft de lezer de mogelijkheid om geïnformeerde beslissingen te nemen.

Toekomstige Verbeteringen

Meer geavanceerde samenvattingstechnieken

- Het integreren van meer geavanceerde AI-modellen voor samenvattingen die specifieke nuances of emoties uit artikelen kunnen halen.

Verbeterde UI

- De website kan verder worden uitgebreid met verbeterde navigatie en visuele elementen om de gebruikerservaring te optimaliseren.

Meertalige ondersteuning

- Het toevoegen van automatische vertaling voor een breder scala aan talen, zodat het systeem wereldwijd ingezet kan worden.

Extra nieuwsbronnen

- Ondersteuning voor meer dan één RSS-feed en de mogelijkheid om artikelen uit meerdere bronnen te combineren en te vergelijken.

Besluit

Dit document biedt een compleet overzicht van de technische opzet en de functionaliteiten van het project. Door de ethische overwegingen en de focus op accuraatheid en transparantie, streeft dit project ernaar om een verantwoorde en efficiënte oplossing te bieden voor nieuwsautomatisering.