

Opleiding : NLP, natuurlijke taalverwerking met Python

Natuurlijke taalverwerking met Python-tools en -bibliotheeken

Praktijkcursus - 3d - 21u00 - Ref. PTS

Prijs : 1650 € V.B.



Deze cursus leert het gebruik van Python voor natuurlijke taalverwerking: datavoorbereiding, tekstrepresentatie en modellering. De deelnemer gebruikt Python tools en bibliotheken om veelvoorkomende NLP taken uit te voeren en NLP modellen te implementeren en toe te passen.

Pedagogische doelstellingen

Aan het einde van de training is de deelnemer in staat om:

- ✓ Python gebruiken om tekstuele gegevens te verwerken
- ✓ De Python-gereedschappen en -bibliotheeken kiezen die nodig zijn voor verwerking
- ✓ De verschillende stadia van voorbewerking en vectorisatie instellen
- ✓ Gebruik geschikte technieken afhankelijk van de doelstellingen: classificatie / onderwerpmodellering / sentimentanalyse
- ✓ Modellen toepassen en evalueren op echte gegevens

Voorafgaande vereisten

Programmeervaardigheden in Python.

Opleidingsprogramma

DEELNEMERS

VOORAFGAANDE VEREISTEN

Programmeervaardigheden in Python.

VAARDIGHEDEN VAN DE CURSUSLEIDER

De deskundigen die de cursus leiden zijn specialisten op het betreffende vakgebied. Zij werden geselecteerd door onze pedagogische teams zowel om hun vakkennis als hun pedagogische vaardigheden voor elke cursus die zij geven. Zij hebben minstens vijf tot tien jaar ervaring in hun vakgebied en oefenen of oefenden verantwoordelijke bedrijfsfuncties uit.

BEOORDELINGSMODALITEITEN

De cursusleider beoordeelt de pedagogische vooruitgang van de deelnemer gedurende de gehele cursus aan de hand van meerkeuzevragen, praktijksituaties, praktische opdrachten, ...
De deelnemer legt ook van tevoren en naderhand een test af ter bevestiging van de verworven kennis.

1 Python-omgeving voor NLP

- De ontwikkelomgeving Python / Anaconda / Jupyter Notebook.
- De belangrijkste gegevenstypen: strings, booleans, getallen, lijsten, tuples en woordenboeken.
- Besturingsstructuren: for en while-lussen, if/elif/else-tests.
- Functies: aanmaken, parameters doorgeven, standaardwaarden, variabele argumenten.
- Numpy: vectoren, matrices, snijden, aaneenschakelen.
- Pandas: analyse van gegevens in tabelvorm (CSV, Excel, enz.), statistieken, pivots, joins, filters.

Praktisch werk

Omgaan met Python in een Jupyter-notebook. Praktische oefening met pandas en numpy.

2 Tekstgegevens voorbereiden

- Identificeren wat tekstuele gegevens zijn en de bibliotheken spaCy en nltk presenteren.
- Woord tokenisatie.
- Verwijdering van stopwoorden, interpunctie en elementen die niet essentieel zijn voor de analyse.
- Lemmatisering vs stemming.

Praktisch werk

Vorbewerking van tekstcorpora met de 2 bibliotheken, vergelijking van resultaten en implementatiemethoden. Aanmaken van stopwoordenlijsten, vergelijking van lemmatisering en rootisering.

3 Informatie-extractie

- Identificatie van de grammaticale aard van woorden met behulp van Part Of Speech Tagging.
- Personen, plaatsen enz. identificeren met Named Entity Recognition.

Praktisch werk

Implementeren van Part Of Speech Tagging en Named Entity Recognition. Analyse van resultaten, filters op bepaalde grammaticale categorieën en eigennamen.

4 Vectorrepresentatie van tekstgegevens

- Zak met woorden.
- Weging tf-idf.
- Benadering met n-grammen.
- Inbeddingen: word2vec, gloVe, fastText ...

Praktisch werk

Transformation d'un corpus de texte en utilisant différentes approches : bag of words, tf-idf, word2vec, gloVe. Comparaison des vecteurs.

PEDAGOGISCHE EN TECHNISCHE MIDDELEN

- De gebruikte pedagogische middelen en cursusmethoden zijn voornamelijk: audiovisuele hulpmiddelen, documentatie en cursusmateriaal, praktische oefeningen en correcties van de oefeningen voor praktijkstages, casestudies of reële voorbeelden voor de seminars.
- Na afloop van de stages of seminars verstrekt ORSYS de deelnemers een evaluatievragenlijst over de cursus die vervolgens door onze pedagogische teams wordt geanalyseerd.
- Na afloop van de cursus wordt een presentielijst per halve dag verstrekt, evenals een verklaring van de afronding van de cursus indien de stagiair alle sessies heeft bijgewoond.

TOEGANGSMODALITEITEN EN TERMIJNEN

De inschrijving dient 24 uur voor aanvang van de cursus plaatsgevonden te hebben.

TOEGANKELIJKHEID VOOR MINDERVALIDEN

Is voor u speciale toegankelijkheid vereist? Neem contact op met mevr. FOSSE, contactpersoon voor mindervaliden, via het adres psh-accueil@ORSYS.fr om uw verzoek en de haalbaarheid daarvan zo goed mogelijk te bestuderen.

5 Machinaal leren op tekstgegevens

- Een herinnering aan de stappen die nodig zijn om een voorspellend model te bouwen.
- Classificatie.
- Sentimentanalyse.
- Onderwerp modellering.

Praktisch werk

Modélisation en utilisant différents types de vecteurs (bag of words vs embeddings). Analyse de sentiment sur des tweets.

6 Modelevaluatieprocedures

- Resamplingtechnieken in training, validatie en testspellen.
- Testen van de representativiteit van trainingsgegevens.
- Prestatiemetingen voor voorspellende modellen.
- Verwarringmatrix.

Praktisch werk

Een NLP-model op een toegepaste manier bouwen en evalueren...

Data en plaats

KLAS OP AFSTAND

2026 : 22 juni, 2 nov.

PARIS LA DÉFENSE

2026 : 22 juni, 2 nov.