

Advanced Language Processing with Python (A704065)

Cursusomvang *(nominale waarden; effectieve waarden kunnen verschillen per opleiding)*

Studiepunten 5.0

Studietijd 150 u

Aanbodsessies in academiejaar 2024-2025

A (semester 2)

Engels

Gent

Lesgevers in academiejaar 2024-2025

Tezcan, Arda

LW22

Verantwoordelijk lesgever

Singh, Pranaydeep

LW22

Medewerker

Aangeboden in onderstaande opleidingen in 2024-2025

Master of Arts in Technology for Translation and Interpreting

stptn

aanbodsessie

5

A

Master of Arts in Advanced Studies in Linguistics (afstudeerrichting Natural Language Processing: Theory and Practice)

6

A

Postgraduate Certificate Computer-Assisted Language Mediation

5

A

Onderwijstalen

Engels

Trefwoorden

Programmeren, Python, natuurlijke taalverwerking, machine learning

Situering

Het vak Advanced Language Processing with Python bouwt voort op eerder verworven programmeercompetenties in Python en focust op geavanceerdere NLP-toepassingen, die in het vak NLP aan bod komen. In dit vak leren studenten hoe ze beschikbare bibliotheken (zoals de Natural Language Toolkit (NLTK), spaCy, NumPy and Scikit-Learn) kunnen integreren in Pythoncode. Ze maken ook gebruik van gesuperviseerde machine learning technieken zoals lineaire en logistische regressie om NLP-problemen op te lossen.

Inhoud

De volgende onderwerpen komen aan bod:

- zoeken en manipuleren van tekst met behulp van reguliere expressies
- integratie van bestaande NLP-bibliotheken in Pythoncode
- tekstverwerking: tokenisatie, part-of-speech tagging, named entity recognition, lemmatisering, dependency parsing
- feature engineering voor NLP taken
- inleiding tot gesuperviseerde machine learning
- visualiseren van gegevens
- evaluatie van machine learning modellen

Begincompetenties

Basiskennis NLP (Natuurlijke Taalwerking)

Basis programmeervaardigheden met Python, zoals:

1. Over algemene kennis beschikken over de werking van computerprogramma's
2. Beschikken over de praktische kennis en vaardigheden die nodig zijn om een eenvoudig programma zelf te implementeren
3. Een grotere opdracht kunnen opsplitsen in deelproblemen
4. Fouten in programmacode kunnen identificeren en verwijderen

De begincompetenties kunnen verkregen worden door het volgen van de vakken Introduction to Processing Language with Python - A704064 (eerste semester) en Natural Language Processing - A704066 (tweede semester).

Eindcompetenties

- 1 Beschikken over de praktische kennis en vaardigheden die nodig zijn om bestaande NLP-bibliotheeken te integreren in Pythoncode
- 2 Beschikken over de praktische kennis en vaardigheden die nodig zijn om machine learning modellen te bouwen en te evalueren
- 3 NLP tools and machine learning kunnen toepassen in een grotere programmeeropdracht

Creditcontractvoorwaarde

Dit opleidingsonderdeel kan niet via creditcontract gevolgd worden

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Werkcollege, Zelfstandig werk

Studiemateriaal

Type: Software

Naam: Google Colab Notebooks

Richtprijs: Gratis of betaald door opleiding

Optioneel: nee

Referenties

- Python Software Foundation. *Official Python documentation*. <http://www.python.org/doc/>
- Aurélien Géron: *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (1st ed.), O'ReillyMedia, 2017.
- Steven Bird, Ewan Klein, & Edward Loper. *Natural Language Processing with Python. Analyzing Text with the Natural Language Toolkit*. <http://www.nltk.org/book>
- D. Jurafsky, J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics* (2nd ed.), Prentice-Hall, 2009.

Vakinhoudelijke studiebegeleiding

Ondersteunend discussieforum via de Ufora-leeromgeving

Individuele begeleiding na afspraak via e-mail met de lesgever

Evaluatiemomenten

niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Evaluatievormen bij niet-periodegebonden evaluatie

Mondelinge evaluatie, Vaardigheidstest, Werkstuk

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

Het inleveren van alle (5) opdrachten zal de leerlingen belonen met 5% van de totale score (1/20).

Het werkstuk bestaat uit een eindproject (codingproject) dat de student in samenspraak met de docent definieert. Elke student moet in staat zijn om de volledige code van het eindproduct uit te leggen.

Het werkstuk wordt in twee delen geëvalueerd: de Python-code (50%) en de mondelinge evaluatie (presentatie, beantwoorden van vragen over de code en het project) (45%). Om te slagen moet de student een gemiddelde score van 10 of meer behalen, met een minimale score van 50% voor de Python-code en een minimale score van 50% voor de mondelinge evaluatie. Indien deze minimumscore niet gehaald wordt op elk afzonderlijk onderdeel, kan een student maximaal 9/20 halen voor het totaal van het opleidingsonderdeel.

Tweede examenkans:

Werkstuk: zelfde eindproject (met mondelinge evaluatie); de studenten dienen een (Goedgekeurd)

nieuwe, verbeterde of aangepaste versie in.

Eindscoreberekening

Vaardigheidstests (5%)

Werkstuk:

- Coding project (50%)
- Mondelinge evaluatie (45%)

Faciliteiten voor werkstudenten

Aanwezigheid in de lessen is ten sterkste aanbevolen.

Beperkte mogelijkheid tot feedback per e-mail voor concrete vragen.