

Advanced Language Processing with Python (A704065)

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

Studiepunten 5.0 **Studietijd** 150 u **Contacturen** 45.0 u

Aanbodsessies en werkvormen in academiejaar 2023-2024

A (semester 2) Engels Gent zelfstandig werk

Lesgevers in academiejaar 2023-2024

Tezcan, Arda	LW22	Verantwoordelijk lesgever
Singh, Pranaydeep	LW22	Medewerker

Aangeboden in onderstaande opleidingen in 2023-2024

	stptn	aanbodsessie
Master of Arts in Technology for Translation and Interpreting	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-bibliotheken 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

Practicum, zelfstandig werk

Leermateriaal

Leermateriaal wordt via Ufora beschikbaar gesteld.
Studenten dienen over een laptop te beschikken en die mee te brengen naar de les.
Geraamde totaalprijs leermateriaal: 0 €

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

Werkstuk

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

De studenten dienen minstens 80% van de oefeningen.
Het werkstuk bestaat uit een eindproject 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.
Tweede examenkans:
Werkstuk: zelfde opdracht; de studenten dienen een nieuwe, verbeterde of aangepaste versie in.

Eindscoreberekening

Werkstuk (100%) [zie de uitzondering zoals beschreven in Details Evaluatievormen]

Faciliteiten voor werkstudenten

Aanwezigheid in de lessen is ten sterkste aanbevolen.
Beperkte mogelijkheid tot feedback per e-mail voor concrete vragen.