UNIVERSITEIT GENT

Studiefiche

Vanaf academiejaar 2026-2027

NLP and Linguistic Analysis (A005870)

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

Studiepunten 5.0 **Studietijd 150 u**

Aanbodsessies in academiejaar 2027-2028

A (semester 1) Engels Gent

Lesgevers in academiejaar 2027-2028

Macken, Lieve	LW22	Verantwoordelijk lesgever
Tezcan, Arda	LW22	Medelesgever
Van Hee, Cynthia	LW22	Medelesgever

Aangeboden in onderstaande opleidingen in 2027-2028	**stptn**	**aanbodsessie**
Bachelor of Arts in de toegepaste taalkunde: combinatie van ten minste twee talen(afstudeerrichting Nederlands, Duits, taaltechnologie)	5	A
Bachelor of Arts in de toegepaste taalkunde: combinatie van ten minste twee talen(afstudeerrichting Nederlands, Engels, taaltechnologie)	5	A
Bachelor of Arts in de toegepaste taalkunde: combinatie van ten minste twee talen(afstudeerrichting Nederlands, Frans, taaltechnologie)	5	A
Postgraduate Certificate Computer-Assisted Language Mediation	5	A
Voorbereidingsprogramma tot Master of Arts in de meertalige communicatie: combinatie van ten minste twee talen	5	A
Voorbereidingsprogramma tot Master of Arts in het vertalen: combinatie van ten minste twee talen	5	A

Onderwijstalen

Engels

Trefwoorden

Natuurlijke taalverwerking, Taalkundige analyses

Situering

In deze cursus gaan we in op de complexiteit van natuurlijke taalverwerking (NLP). We leren hoe algoritmes kunnen worden gebruikt om computers verschillende niveaus van tekstbegrip bij te brengen op basis van menselijke taalinput. Omdat taal zo ambigu is, starten we de bespreking op het morfologische en lexicale niveau. Van daaruit bouwen we stapsgewijs een tekstrepresentatie op via het syntactische niveau, met als eindpunt de complexiteit van het semantische en het discoursniveau.
Deze cursus heeft een tweeledig doel. Enerzijds biedt ze studenten theoretisch inzicht in de basisprincipes van natuurlijke taalverwerking. Anderzijds wordt er voortgebouwd op eerder verworven programmeercompetenties in Python en leren studenten hoe ze beschikbare bibliotheken, zoals spaCy, kunnen integreren in Pythoncode. Zo passen ze de verworven kennis actief toe in het ontwikkelen van NLP-toepassingen.
Deze cursus bouwt verder op het vak Inleiding programmeren voor de humanities en wordt gedoceerd gedurende 5 lesweken (weken 4-8).

Inhoud

De cursus behandelt de volgende onderwerpen:

- Basisprincipes van tekstverwerking en preprocessing, waaronder tokenisatie, part-of-speech tagging, named entity recognition, lemmatisering en dependency parsing
- Corpusanalyse met aandacht voor woordfrequenties, lexicale diversiteit, densiteit en sophistication, en topicanalyse

- Toepassing van reguliere expressies, edit distance en dynamic programming
- Syntactische analyse met nadruk op PoS-tagging en parsing
- Inleiding tot taalmodellen en representatie van betekenis via woordvectoren
- Ethische vraagstukken rond taalmodellen en gegevensverwerking, zoals bias
- Integratie van bestaande NLP-bibliotheken, zoals spaCy, in Pythoncode
- Visualisatie van taalkundige gegevens ter ondersteuning van tekstanalyse en -interpretatie

Begincompetenties

Een goede basiskennis Python programmeren (zie eindcompetenties 'Inleiding Programmeren voor de humanities')

Eindcompetenties

1 Studenten beschikken over basiskennis over en inzicht in natuurlijke taalverwerking (NLP) als een fundamenteel subdomein binnen de artificiële intelligentie
2 Studenten beschikken over de nodige praktische kennis en vaardigheden om bestaande NLP-bibliotheken te integreren in Pythoncode, taalkundige analyses uit te voeren van teksten en die informatie te visualiseren.

Creditcontractvoorwaarde

Dit opleidingsonderdeel kan niet via creditcontract gevolgd worden

Examencontractvoorwaarde

Dit opleidingsonderdeel kan niet via examencontract gevolgd worden

Didactische werkvormen

Werkcollege, Hoorcollege, Zelfstandig werk

Toelichtingen bij de didactische werkvormen

Hoorcolleges en hands-on werkcolleges
Zelfstandig werk thuis

Lesopnames worden ter bechikking gesteld

Studiemateriaal

Type: Slides

Naam: NLP and Linguistic analysis
Richtprijs: Gratis of betaald door opleiding
Optioneel: nee
Taal : Engels
Beschikbaar op Ufora : Ja

Type: Software

Naam: Google Colab Notebooks
Richtprijs: Gratis of betaald door opleiding
Optioneel: nee

Referenties

- Daniel Jurafsky & James Martin, "Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition" (ed3).
- Python Software Foundation. Official Python documentation. http://www.python.org/doc
- SpaCy. https://spacy.io/

Vakinhoudelijke studiebegeleiding

Individuele begeleiding aansluitend op de lessen, of na afspraak via e-mail met de lesgever.

Evaluatiemomenten

periodegebonden en niet-periodegebonden evaluatie

Evaluatievormen bij periodegebonden evaluatie in de eerste examenperiode

Schriftelijke evaluatie

Evaluatievormen bij periodegebonden evaluatie in de tweede examenperiode

Schriftelijke evaluatie

Evaluatievormen bij niet-periodegebonden evaluatie

Participatie, Werkstuk

Tweede examenkans in geval van niet-periodegebonden evaluatie

Examen in de tweede examenperiode is mogelijk

Toelichtingen bij de evaluatievormen

Eerste zittijd:

- Werkstuk + aanwezigheid in de les (10%)
- Schriftelijk examen (90%)

Tweede zittijd:

- Schriftelijk examen (100%)

Eindscoreberekening

Een combinatie van niet-periodegebonden (10%) en periodegebonden evaluatie (90%)

Faciliteiten voor werkstudenten

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