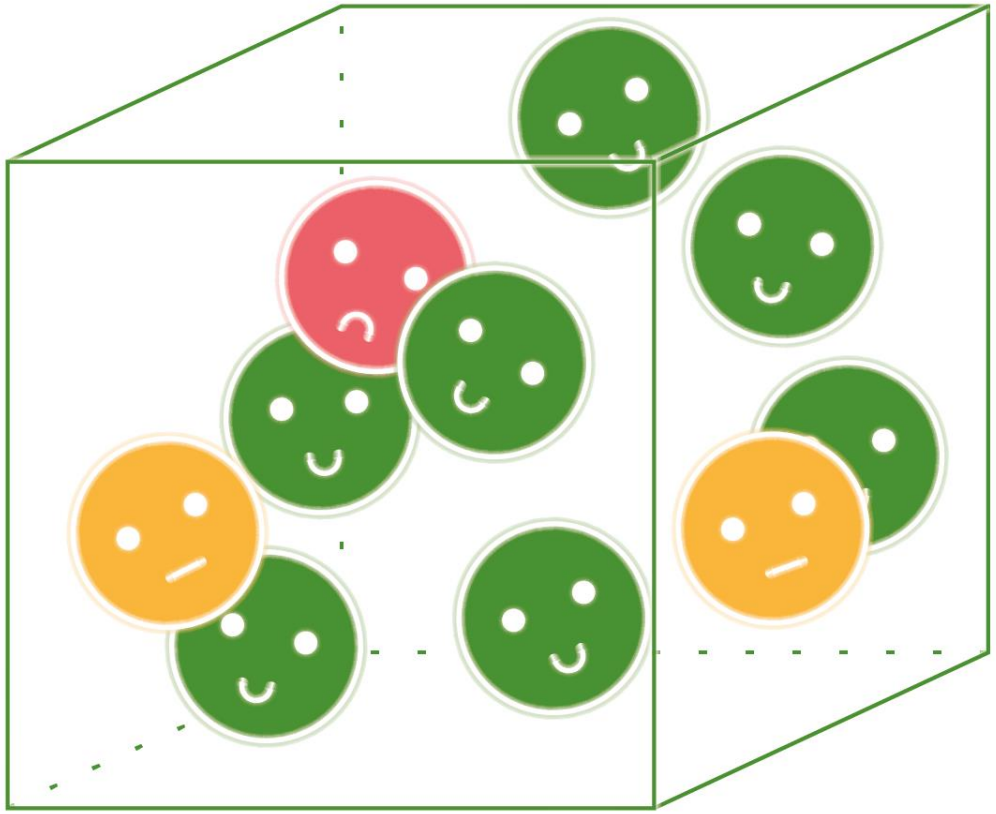


Taaltechnologie

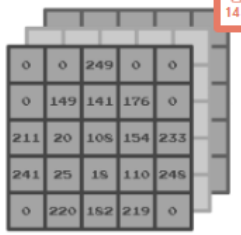
Congres 2LinK^2- 16 juni 2022

UNIVERSITEIT GENT

dwengo

Natacha Gesquière

natacha.gesquiere@ugent.be



Python in STEM - Digitale beelden

Hoe komt een foto tot stand? Hoe herkent een computer kleur? Dit alles en hoe je kleuren kunt aanpassen, leer je in hier.

Python in STEM - Epidemie

Bekijk epidemies vanuit het oogpunt van een wetenschapper. Hoe voorkom je dat een besmettelijke ziekte de hele wereld lamlegt?

Python in STEM - IP-adres

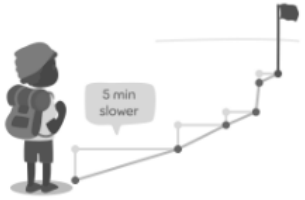
Wat zijn IP-adressen? Hoe worden ze gemaakt en waarvoor worden ze gebruikt? Dit alles leer je a.d.h.v. interactieve Python notebooks.

Python in STEM - Klimaatverandering

Technologie is één van de vele zaken die ons helpen in de strijd tegen de klimaatverandering. Ontdek hier hoe!

Python in wiskunde - Lineaire Regressie

Ontdek m.b.v. lineaire regressie verbanden in data die je anders niet zou kunnen beschrijven! Je gaat aan de slag met datasets en leert hoe je bepaalde soorten data kunt verwerken.

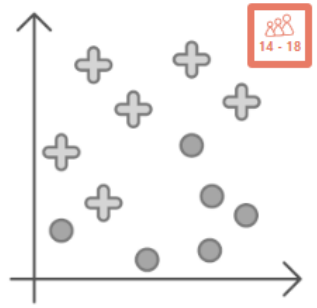


Python in wiskunde - Pythagoras

Leer functies analyseren m.b.v. Python notebooks. Eens je dit kunt, wordt je inzicht op de proef gesteld met enkele interessante oefeningen.

Python in wiskunde - Rechten

Leer rechten tekenen met Python.

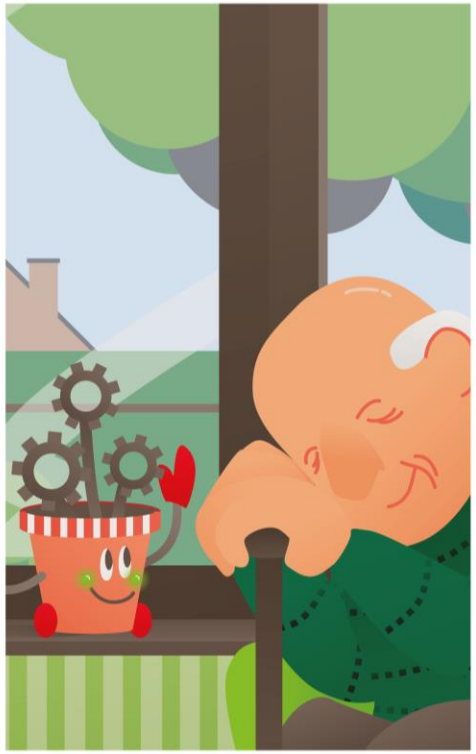


Python in wiskunde - Spreidingsdiagram

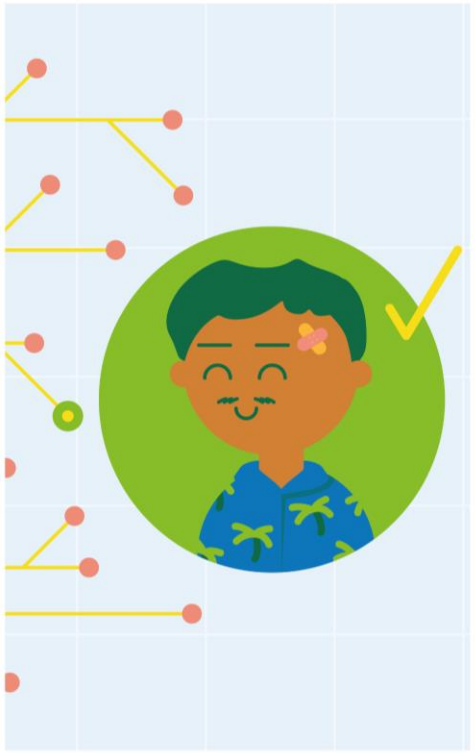
Een basisleerpad waarin je leert hoe je assenstelsels kunt tekenen in Python. Hierna ontdek je hoe jij eenvoudige datasets voor kan stellen in zo'n assenstelsel.

i-Learn

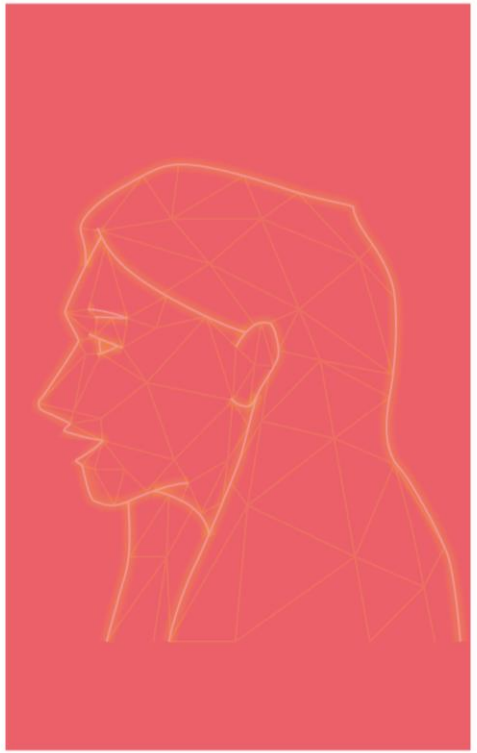
SOCIALE ROBOT



AI IN DE ZORG

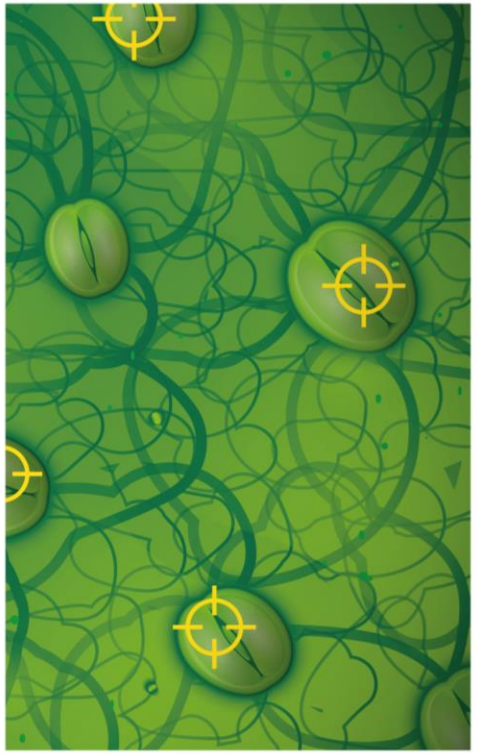


CHATBOT



AI IN DE KUNST

KIKS



AI OP SCHOOL

AI OP SCHOOL

PROJECTEN AGENDA NIEUWS

Chatbot

Waar taal en technologie samenkomen, ontstaat het domein van Natural Language Processing (NLP). Met technieken uit machine learning, zoals het trainen van modellen om tekst, afbeeldingen of geluidsfrequenties te herkennen, kunnen we computers op een intelligente manier laten werken met taal. Bekende NLP-toepassingen van machine learning zijn bijvoorbeeld vertaalprogramma's, spraakbesturingssystemen of virtual assistants zoals Siri, en chatrobots.

We buigen ons over vragen als "Hoe kunnen we automatisch teksten begrijpen, vertalen en zelfs schrijven?", "Kan een computer emoties of ironie herkennen in een tekst?", en vooral "Hoe belangrijk zijn data?". Leerlingen van de tweede en de derde graad secundair onderwijs krijgen inzicht in het domein NLP en zetten hun eerste stappen in de wereld van het programmeren met Python.

UNIVERSITEIT GENT

dwengo

language and translation technology team

Woordsuggestie

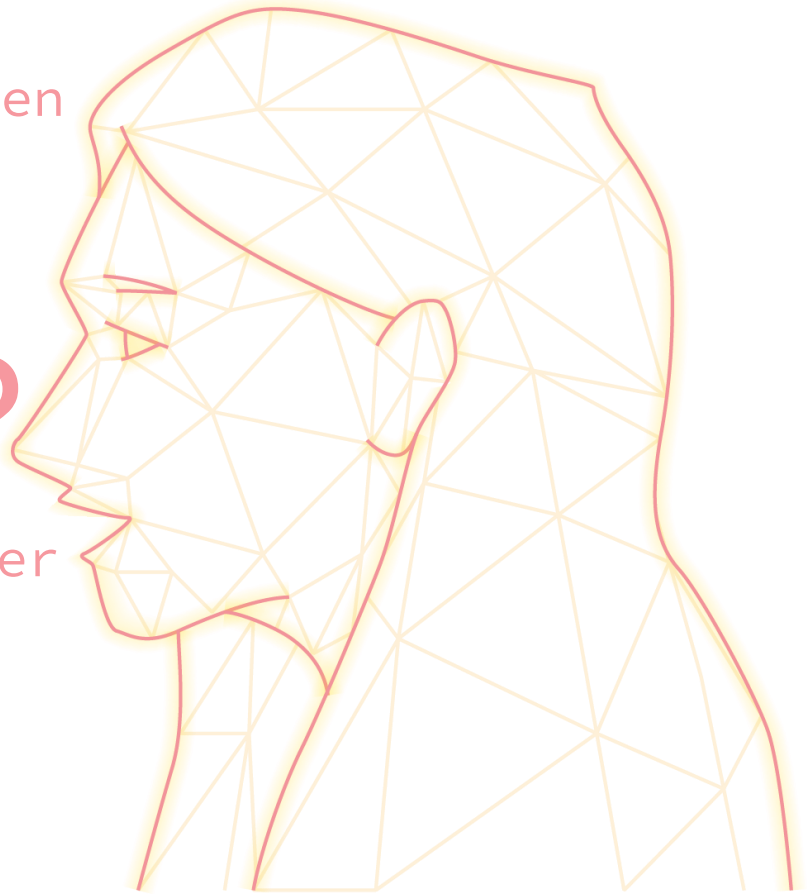
Automatische vertaling

Nieuwsberichten schrijven

Chatbot

Spamdetectie

Taaltechnologie?



Elektronisch patiëntendossier

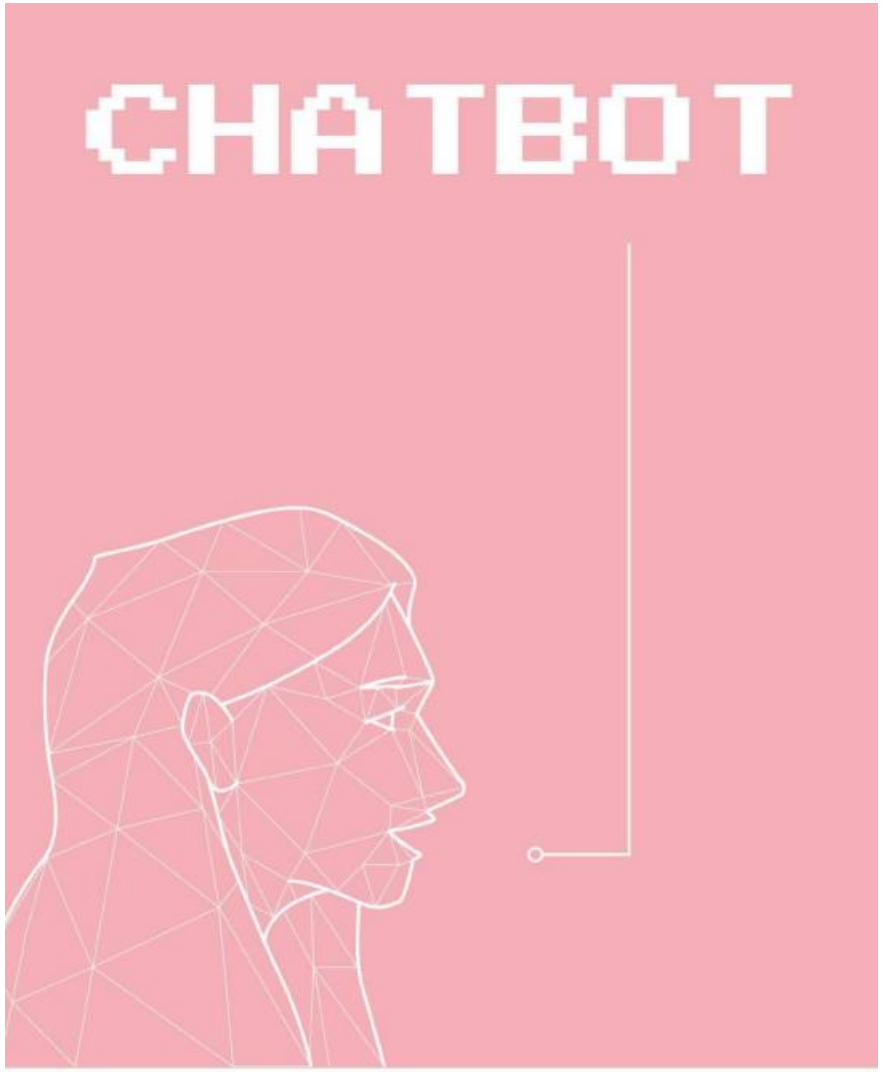
Home assistant

Spraakherkenning

Cyberpestdetectie

Auteursherkenning

Plagiaatdetectie



Gaan robots de job van journalist vervangen?

Dit artikel werd geschreven door een AI-systeem.

Gaan robots de job van journalist vervangen?

Een nieuw tijdperk, een nieuw tijdperk aan informatie, een nieuwe manier van denken, een nieuw tijdperk aan online toepassingen en een nieuw tijdperk voor de digitale infrastructuur.

Dit is niet zomaar een baan. Veel mensen denken dat robots ook met dezelfde vraag kunnen worden aangesproken. Dit is niet zo. De robot die het best scoort, wint. Een betere wereld gaat nooit boven een computer. Het is zelfs zo dat een robot gemiddeld beter wordt. Daarnaast zullen robots steeds groter worden naarmate ze meer en vaker in de wereld worden geplaatst. Wat betekent dat voor de mens als we dat willen. Het antwoord is eenvoudig: robots zijn de perfecte vervanging voor ons. Wij doen dit met passie.

Door middel van slimme software en een perfecte service is het mogelijk om de robot te laten groeien. Een wereld vol data, voor iedereen toegankelijk en open. En een wereld vol technologie. Dat is onze passie. In de bouw van het nieuwe tijdperk zitten we vol met slimme technologie. We zijn ervan overtuigd dat samenwerken nog steeds een van de grootste uitdagingen van onze tijd is.

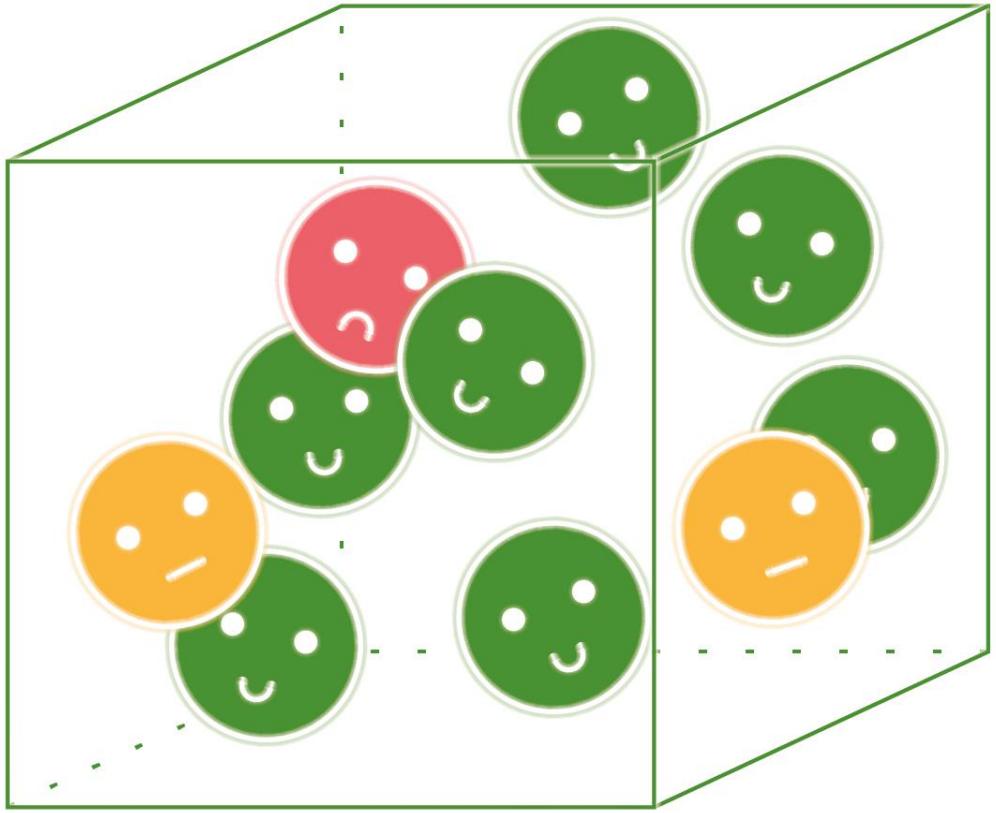
(Artikel hier opgenomen met toestemming van ML6 en Data News).

Een tijdschriftartikel genereren met GPT-2

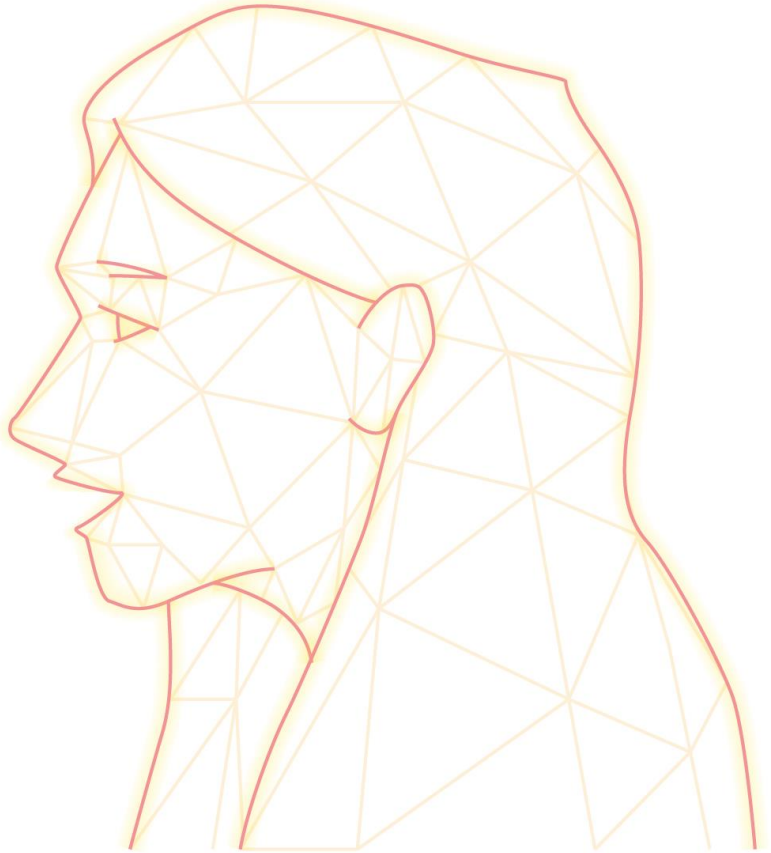
Bekijk op de website `https://kanaalz.knack.be/embed/video/1658795?sad=_Site-KanaalZ-NL-nl_` hoe het artikel 'Gaan robots de job van journalist vervangen?' tot stand kwam.

Probeer het AI-systeem van ML6 op de webpagina `https://gpt2.ml6.eu/nl`.

Uitgewerkt - notebooks



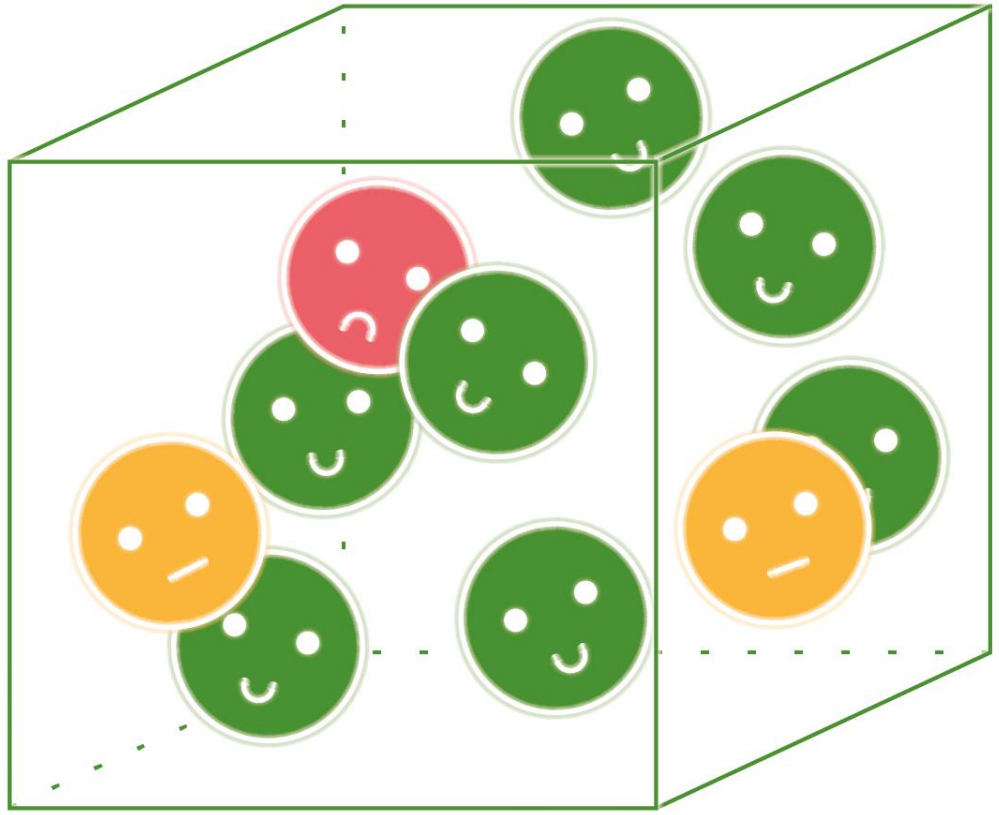
Sentimentanalyse



Chatbot

NA DEZE WORKSHOP

DEZE WORKSHOP SENTIMENTANALYSE



Sentimentanalyse

Tekst	Label
Een mooi, gezellig restaurantje in hartje Gent❤️	POSITIEF
De kroketten zijn niet slecht, maar dat mag ook wel voor een fortuin.	NEGATIEF
Het restaurant heeft een Scandinavisch interieur.	NEUTRAAL
Ze stuurden ons van het kastje naar de muur en de bediening was te traag!	NEGATIEF
De sommelier bood een heerlijk wijntje aan!	POSITIEF

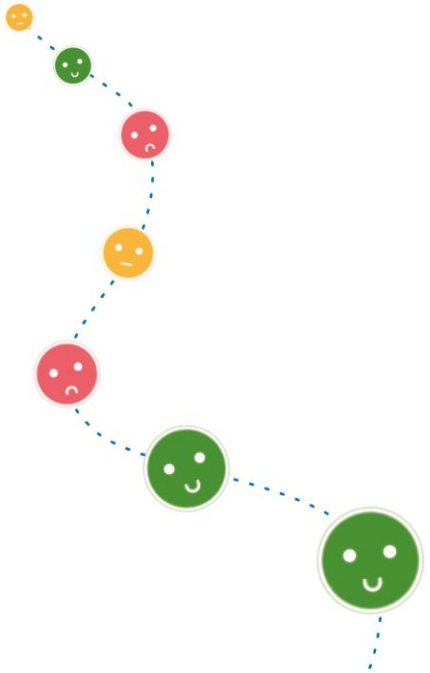
Sentimentanalyse

Socialemediaposts
bv. reviews

POSITIEVE REVIEW

NEUTRALE REVIEW

NEGATIEVE REVIEW



Sentimentanalyse

Socialemediaposts
bv. reviews

POSITIEVE REVIEW

NEUTRALE REVIEW

NEGATIEVE REVIEW

We willen een AI-systeem ontwikkelen dat reviews op socialemediaposts classificeert volgens sentiment (positief, neutraal, negatief).

Verkennen van het probleem. Wat heb je nodig?
Subtaken (**decompositie**):

1. Keuze maken tussen **regelgebaseerd** of datagebaseerd AI-systeem
2. Sentimentwoordenboek in Nederlands inladen
3. De reviews voorverwerken (tokeniseren, hoofdletters en leestekens verwijderen)
4. De reviews representeren a.d.h.v. woordenboekvorm van de woorden erin, samen met hun woordsoort
5. Matching tussen de woorden en het sentiment-woordenboek om sentimentscore elk word te bepalen
6. Bepaal de sentimentscore en sentiment de reviews.

Het gebruik van een lexicon (woordenboek specifiek voor de taak) en tokeniseren is een techniek die veel wordt toegepast in taaltechnologie, bv. cyberpest detectie, auteursherkenning, automatisch vertalen, tekst genereren … (**patroonherkenning**)

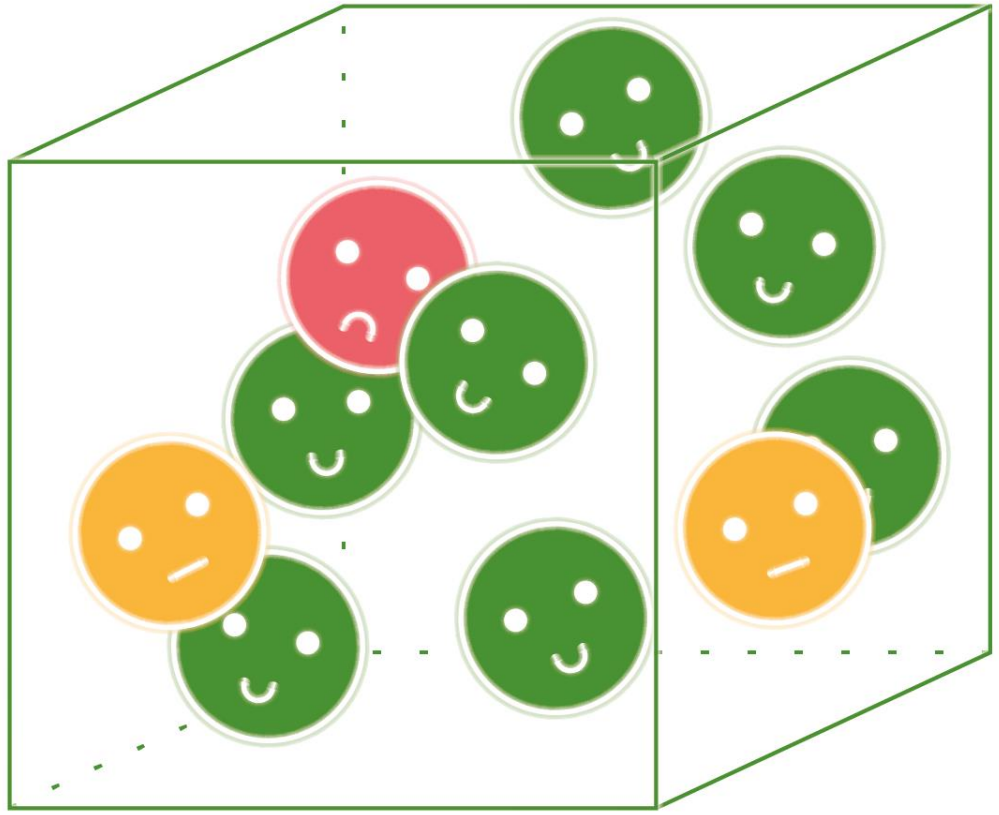
Door te tokeniseren worden de zinnen herleid tot de woorden die ze bevatten.
De hoofdletters en leestekens worden genegeerd. De woorden worden herleid tot hun woordenboekvorm en aangevuld met hun woordsoort.(**abstractie**)
De voorverwerking laat toe om een datastructuur die geschikt is om efficiënt te zoeken, te gebruiken.

Een **algoritme** om het sentiment van een review te bepalen

- Maak een lijst van de woorden die in de zin voorkomen, met hun woordsoort. Negeer leestekens.
- Zoek elk woord op in het sentimentwoordenboek.
- Sla de score van elk woord op in een lijst.
- Tel de scores op.
- Bepaal het sentiment a.d.h.v. van een wiskundige uitdrukking (groter dan 0, gelijk aan 0, kleiner dan nul).

UNIVERSITEIT GENT dwengo

NOTEBOOKS



```
i = 0   # index
for token in lexiconmatches:
    if polariteiten[i] > 0: # overeenkomstige polariteit is positief
        review_highlighted = review_highlighted.replace(" " + token + " ", " " + Back.GREEN + token + Back.RESET + " ")   # posit
    elif polariteiten[i] == 0.0: # overeenkomstige polariteit is neutraal
        review_highlighted = review_highlighted.replace(" " + token + " ", " " + Back.BLUE + token + Back.RESET + " ")    # neutr
    elif polariteiten[i] < 0: # negatieve polariteit
        review_highlighted = review_highlighted.replace(" " + token + " ", " " + Back.RED + token + Back.RESET + " ")     # negat
    i = i + 1

print(review_highlighted)
```

```
 nieuw concept in gent ,  maar dat kan volgens mij toch beter .  de meeste cornflakes waren gewoon de basic soorten .  ook wat
duur voor de hoeveelheid die je krijgt ,  vooral met de toppings zijn ze zuinig .  en als je ontbijt aanbiedt ,  geef de mensen
dan toch ook wat meer keuze voor hun koffie .
```

7. Oefening

Bekijk de volgende reviews.

- Kies er een uit.
- Welk sentiment koppel jij eraan?
- Leidt het lexicon tot hetzelfde resultaat?

> Sunrisewater is net wat België nodig heeft. Met het obesitasprobleem dat toch wel aan een opmars bezig is, kunnen we alle initiatieven

Naar de leerpaden met de notebooks

https://dwengo.org

Werken met notebooks

Sentimentanalyse

- Operatoren
- Dataypes
- Programmeerstructuren

Chatbot

Naar het leerpad

https://www.dwengo.org/

> Leerpad Sentimentanalyse

Overig materiaal

https://www.aiopschool.be/chatbot/

> Leerpad Chatbot

Naar het leerpad

https://www.dwengo.org/

> Leerpad Sentimentanalyse

Overig materiaal

https://www.aiopschool.be/chatbot/

> Leerpad Chatbot

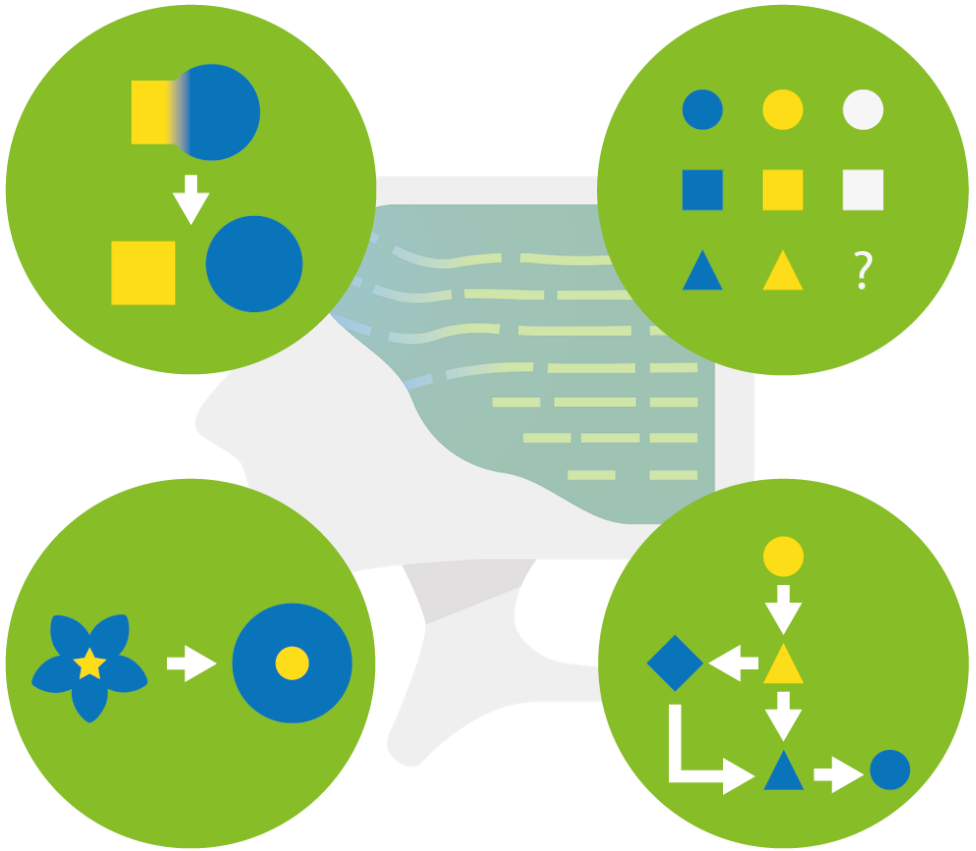
Basisconcepten van computationeel denken

Decompositie

Een probleem doordacht opsplitsen in goedgekozen deelproblemen, zodat elk deel afzonderlijk kan worden aangepakt, en een oplossing van het probleem gemakkelijker bekomen wordt.

Patroonherkenning

Achterhalen dat bepaalde aspecten van een probleem gelijkenissen vertonen, waardoor het vereenvoudigd kan worden.
Herkennen dat een probleem gelijkenissen vertoont met een eerder opgelost probleem.



Abstractie

Negeren van informatie die niet nodig is om een probleem op te lossen.
Bepaalde details verbergen om in grote lijnen over een probleem te kunnen nadenken.
Bij het aanpakken van complexe problemen is er nood aan abstractie omdat het redeneren vlotter verloopt zonder de ballast van irrelevante details.

Algoritmisch denken

Expliciteren van een reeks eenduidige instructies die stapsgewijs moeten worden uitgevoerd.
Inzien dat deze reeks van instructies en de volgorde ervan essentieel zijn om het gewenste resultaat op te leveren.

Voor een bepaald probleem zijn deze vier basisconcepten vaak verweven met elkaar.

Waar vind je ons materiaal?

dwengo.org
blockly.dwengo.org
aiopschool.be
aiopschool.be/chatbot
aiopschool.be/agenda
wegostem.be

facebook.com/dwengo
@dwengo
@NachaGesquiere

UNIVERSITEIT GENT