



Portfolio scholingsactiviteit

Naam van bijeenkomst	Webinar Taaltechnologie, 'Automatische vertaling en ChatGPT: een kijkje onder de motorkap' door Vincent Vandeghinste
Datum en locatie	Donderdag 11 mei 2023 Online-bijeenkomst 11.00 – 12.00 uur
Doelgroep	Vertalers (al dan niet lid van de VZV)
Leerdoel	De deelnemer leert over de laatste technieken voor automatische vertaling en tekstgeneratie, neurale netwerken en het trainen daarvan, de inzet daarvan in het vertaalproces, wat AI wel en niet kan en de rol van de vertaler.
Competentie(s)	vertaalcompetentie tolkcompetentie linguïstische en tekstuele competentie culturele competentie zoekstrategieën en onderzoekscompetentie X technische competentie zakelijke en ondernemingsgerichte competentie
Beschrijving van de globale inhoud	Abstract: Automatische vertaling en ChatGPT: Een kijkje onder de motorkap Automatische vertaling of machinevertaling (MT) vindt plaats met behulp van een neurale netwerk (een netwerk van knooppunten en verbindingen gebaseerd op de werking van onze hersenen). Dit moet worden gevoed met teksten in twee of meer talen die vertalingen zijn van elkaar (parallele data). Dat is de inputlaag van het algoritme. Met die data (batches) wordt het netwerk getraind. In het neurale netwerk worden woorden omgezet in vectoren (een reeks getallen) met een bepaald gewicht. Dit gebeurt in een zogenaamde 'hidden layer'. Vervolgens wordt de informatie doorgegeven als outputlaag van het algoritme. Het vertalen door een neurale netwerk was tot voor kort simpel gezegd niet meer dan het voorspellen van het volgende woord.



Vereniging Zelfstandige Vertalers keurmerk voor kwaliteit

	<p>Bij het trainen van het netwerk, waarbij gebruik wordt gemaakt van enorme hoeveelheden tekst, wordt gekeken in hoeverre de voorspelling afwijkt van de gewenste waarde. Afhankelijk van de afwijking kan het toegekende gewicht iets worden aangepast, waarna het netwerk verder kan worden getraind totdat de behaalde score niet meer verbetert.</p> <p>Neurale netwerken zijn in de loop van de tijd verbeterd. Bij recurrente neurale netwerken (RNN's) wordt informatie bijvoorbeeld niet alleen rechtstreeks naar de outputlaag gestuurd, maar ook teruggestuurd in de 'hidden layer', waardoor het systeem zich recente woorden beter herinnert.</p> <p>Een grote verbetering was het toepassen van Transformers in plaats van RNN's. Het belangrijkste verschil tussen deze twee systemen is het aandachtsmechanisme. Met Transformers (waarbij ieder woord een 'hidden layer' heeft) kan het systeem focussen op alle ingevoerde woorden, niet alleen op de meest recente. Zo leert het welke woorden nauw gerelateerd zijn aan elkaar en kan het langere zinnen en context verwerken. Dit is waarom AI nu zo veel beter werkt dan vroeger.</p> <p>De voorgetrainde (pretrained) taalmodellen worden steeds groter. Taalmodel mBart wordt getraind door gemaskeerde zinnen in te voeren, waarbij de decoder leert de volledige zinnen te voorspellen in de andere taal, inclusief de gemaskeerde woorden.</p> <p>ChatGPT is een Transformer-taalmodel voor conversatie (chat). Het kan tekst genereren in verschillende stijlen en voor verschillende doeleinden (ook voor vertaling). Het is ook 'alleen maar' een taalmachine en geen zoekmachine. Het controleert geen feiten, het voorspelt alleen woorden op basis van de voorgaande woorden en de door de gebruiker gegeven instructie (prompt). Het lijkt misschien dingen te begrijpen, maar dat doet het niet. Het kan niet redeneren. ChatGPT zegt zelf dat het kan helpen bij het maken van vertalingen, maar dat het geen vervanger is voor de menselijke vertaler.</p> <p>Taalmodellen kunnen MT niet vervangen, maar wel een instrument zijn in combinatie met MT. Je zou ChatGPT bijvoorbeeld kunnen gebruiken als extra plugin in een CAT-tool. Maar ChatGPT heeft beperkingen: het kent niet 'alle' talen en is beter in het vertalen naar het Engels dan vanuit het Engels, heeft een enorm grote</p>
--	---



Vereniging Zelfstandige Vertalers
keurmerk voor kwaliteit

	<p>ecologische voetafdruk en kan niet makkelijk aan een nieuw domein worden aangepast.</p> <p>Trends op het gebied van MT: De markt wordt steeds groter, MT is steeds meer geaccepteerd, de kwaliteit is verbeterd en zal nog beter worden. Voor vertalers wordt post-editing steeds belangrijker. Voorsnog kunnen niet alle soorten vertalingen worden geautomatiseerd (literatuur, poëzie, teksten op speciale vakgebieden).</p> <p>Conclusie: MT-kwaliteit is sterk verbeterd, hoofdzakelijk door het gebruik van de Transformers-technologie en de combinatie met grote taalmodellen. Deze verbetering zal doorgaan in de toekomst. ChatGPT is voorsnog geen specifieke bedreiging voor de vertaler. Het is slechts een systeem dat grote taalmodellen gebruikt die gefinetuned zijn voor chatten.</p>
Lesmateriaal	PowerPointpresentatie die achteraf ter beschikking wordt gesteld van de deelnemers.
Aantal contacturen (online)	1
Aantal uren studiebelasting	1
Totaal aantal PE-punten	1
Informatie over deskundige (incl. curriculum vitae)	<p>Vincent Vandeghinste studeerde theoretische psychologie (psycholinguïstiek) aan de KU Leuven en taalwetenschap aan de KU Leuven en de Universiteit van Antwerpen. Voor zijn doctoraat aan de KU Leuven schreef hij zijn proefschrift over automatische vertaling zonder parallelle data.</p> <p>Hij werkt sinds 2018 bij het Instituut voor de Nederlandse Taal als onderzoeker, hoofdzakelijk rond taalkundige infrastructuur (www.clarin.eu) en taaltechnologie voor inclusie.</p> <p>Daarnaast is hij verbonden aan de KU Leuven, waar hij betrokken is bij de vakken Computerlinguïstiek, Language Engineering Applications, Machine Translation and Post-Editing en Computationale Lexicografie.</p>
Wijze waarop scholingsactiviteit wordt geëvalueerd	Door middel van evaluatieformulieren, ingevuld terug te sturen naar de activiteitencommissie van de VZV.
Verdere informatie	www.vzv.info