

Scholieren met dyslexie van de derde graad middelbaar onderwijs ondersteunen bij het intensief lezen van wetenschappelijke artikelen via geautomatiseerde en gepersonaliseerde tekstvereenvoudiging.

De opbouw van een LLM-gedreven prototype voor online geautomatiseerde en gepersonaliseerde tekstvereenvoudiging.

Dylan Cluyse.

Scriptie voorgedragen tot het bekomen van de graad van
Professionele bachelor in de toegepaste informatica

Promotor: Mevr. L. De Mol

Co-promotor: J. Decorte; J. Van Damme;

Academiejaar: 2022–2023

Eerste examenperiode

Departement IT en Digitale Innovatie .

**HO
GENT**

Woord vooraf

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Adaptieve tekstvereenvoudiging helpt deze scholieren bij hun lees- en verwerkingsnelheid. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

Inhoudsopgave

Lijst van figuren	vii
Lijst van tabellen	ix
1 Inleiding	1
1.1 Probleemstelling	2
1.2 Onderzoeksvraag	3
1.3 Onderzoeksdoelstelling	4
1.4 Opzet van deze bachelorproef	4
2 Stand van zaken	6
2.1 Inleiding	6
2.2 Kenmerken van dyslexie en wetenschappelijke artikelen	6
2.2.1 Specifieke noden van scholieren met dyslexie in de derde graad van het middelbaar onderwijs	7
2.2.2 Specifieke kenmerken van wetenschappelijke artikelen	8
2.2.3 Trends van wetenschappelijke artikelen	9
2.3 Aanpakken voor tekstvereenvoudiging	10
2.3.1 Manuele tekstvereenvoudiging	10
2.3.2 Bevoordelende effecten op handmatige tekstvereenvoudiging bij scholieren met dyslexie	11
2.3.3 Aanpak voor ATV	12
2.3.4 Prompt engineering	14
2.4 De verschillende soorten tekstvereenvoudiging	15
2.4.1 Ondersteunende en bevoordelende effecten van ATV voor scholieren met dyslexie	19
2.4.2 Conclusie	20
2.4.3 MTV en ATV combineren	20
2.4.4 Trends bij ATV	20
2.5 De valkuilen bij AI en NLP	20
2.6 Beschikbare tools en taalmodellen	22
2.6.1 Conclusie	30
2.7 Conclusie	30
3 Methodologie	32
3.1 Requirementsanalyse	32

3.2	Vergelijkende studie	33
3.2.1	Vergelijking met referentieteksten	34
3.3	Prototype voor tekstvereenvoudiging.	35
3.3.1	Personalisering van de website	36
3.3.2	Tool voor leerkrachten	36
3.3.3	Tekstinhoud extraheren	36
3.3.4	Zinnen ophalen.	37
3.3.5	PoS-tagging	37
3.3.6	Opties voor gepersonaliseerde tekstvereenvoudiging aanreiken. 37	
3.3.7	Tekstvereenvoudiging met API.	38
3.3.8	Personaliseerbaarheid aanreiken	39
3.3.9	Tekstinhoud uitschrijven naar PDF/DOCX.	39
3.3.10	Conclusie	40
4	Discussie	41
4.1	Requirementsanalyse	41
4.2	Vergelijkende studie	43
4.3	Opbouw van het prototype	44
5	Conclusie	45
A	Onderzoeksvoorstel	49
A.1	Introductie	49
A.2	State-of-the-art	51
A.2.1	Tekstvereenvoudiging	51
A.2.2	Noden van scholieren met dyslexie.	51
A.2.3	Huidige toepassingen.	53
A.2.4	Ontwikkelen met AI	53
A.3	Methodologie	54
A.4	Verwacht resultaat, conclusie	56
	Bibliografie	57

Lijst van figuren

1.1	Figuur van De Meyer e.a. (2019)	2
2.1	Afbeelding uit Plavén-Sigraay e.a. (2017)	9
2.2	Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013)	11
2.3	Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)	14
2.4	Afbeelding uit McFarland (2023)	15
2.5	Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016)	16
2.6	Afbeelding uit Bulté e.a. (2018)	19
2.7	Afbeelding uit Bulté e.a. (2018)	20
2.8	Schermafbeelding van SciSpace	23
2.9	Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast.	24
2.10	Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak.	24
2.12	Afbeelding van Binz en Schulz (2023). Dit toont de <i>mean regret</i> aan tussen de vier engines en de menselijke antwoorden.	25
2.11	Afbeelding van Gooding 2022.	25
2.13	Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De performantie van de modellen ten opzichte van de grootte volgt een lineaire functie.	27
2.14	Afbeelding van Ribas (2023).	29
2.15	In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee.	29

A.1 (Readable, 2021)	55
--------------------------------	----

Lijst van tabellen

List of Listings

1

Inleiding

Iedereen wordt dagelijks geconfronteerd met lezen. Deze vaardigheid strekt zich uit tot elk aspect van ons dagelijks leven, en dit geldt des te meer in het onderwijs, waar leraren worden aangemoedigd om diverse leesmaterialen te gebruiken om lesinhouden op een authentieke manier over te brengen. Wetenschappelijke artikelen kunnen ingezet worden als leesvoer voor scholieren in de derde graad van het middelbaar onderwijs, maar de leesgraad van deze artikelen brengt een nieuwe uitdaging met zich mee voor zowel scholieren als leerkrachten.

Het Amerikaanse onderwijs stampte C.R.E.A.T.E.¹ uit de grond. Dit initiatief zet scholieren tussen 12 en 18 jaar aan om wetenschappelijke artikelen te lezen in plaats van enkel boeken. Daarmee komen zij in direct contact met wetenschappelijk onderzoek. Zo heeft deze doelgroep meer grip op hoe wetenschappers experimenten uitvoeren, plannen en resultaten analyseren en interpreteren. In Vlaanderen bestaat er geen gelijkaardig initiatief, maar er wordt wel vanuit de lerarenopleidingen sterk ingezet op divers didactisch leesmateriaal. Vlaamse STEM-leerkrachten in de derde graad middelbaar onderwijs moeten volgens het M-decreet en de leerplannen van zowel het katholiek² als het gemeenschapsonderwijs³ hun theorielessen op een toegankelijke manier aanbieden, zodat alle scholieren ongeacht leesachterstand worden meegenomen in het verhaal.

Vlaanderen is met een jaarlijks budget van 32 miljoen een pionier in Europa op het gebied van artificiële intelligentie (AI) op de werkvloer (Crevits, 2022). Zo stampte de Vlaamse overheid verschillende AI-projecten uit de grond, om Vlaamse AI-ontwikkelingen te ondersteunen en om AI-softwarebedrijven te inspireren. Het amai!-project⁴ brengt

¹<https://teachcreate.org/>

²<https://pro.katholiekonderwijs.vlaanderen/basisoptie-stem/ondersteunend-materiaal>

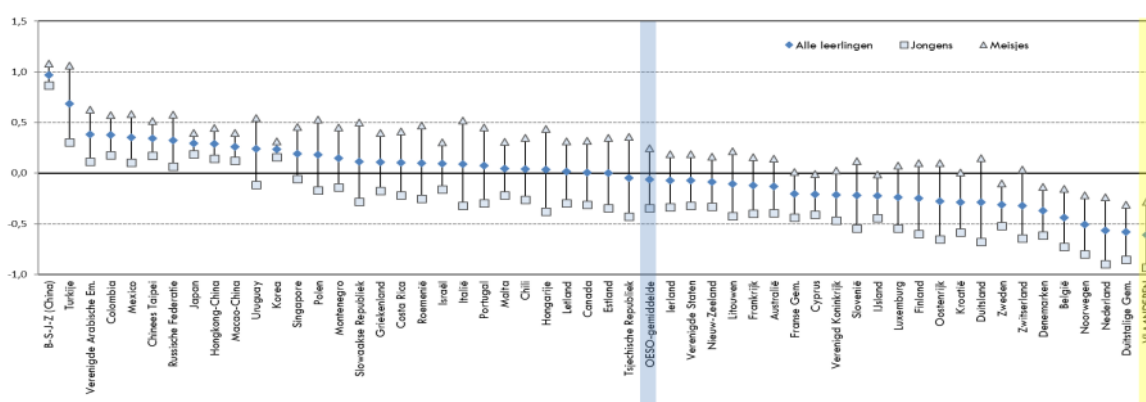
³<https://g-o.be/stem/>

⁴<https://amai.vlaanderen/>

AI-softwarebedrijven uit diverse domeinen, waaronder het onderwijs om processen te automatiseren en de werkdruk verminderen, zoals met real-time ondertiteling in de klas en een taalassistent voor leerkrachten in meertalige klasgroepen.

1.1. Probleemstelling

Iedere drie jaar wordt in ongeveer 79 geïndustrialiseerde landen de PISA-test afgenomen, die waaronder leesvaardigheid en de wetenschappelijke geletterdheid van 15-jarige scholieren, meet. Uit de enquête blijkt dat deze doelgroep zeer negatief is over leesplezier en daarmee scoort Vlaanderen het slechtst van alle bevroagde landen. Zoals aangegeven in 1.1 beschouwt bijna de helft van de bevroagden intensief lezen als tijdverspilling en slechts 17% beschouwt lezen als een hobby. Boeiende topics kunnen belemmerd worden door deze vorm van media bij deze doelgroep.



Figuur (1.1)

Figuur van De Meyer e.a. (2019)

Intensief leesbegrip valt niet te omzeilen in onze huidige samenleving, maar dit leesbegrip verschilt enorm in het middelbaar onderwijs. Een doelgroep die hier extra mee geconfronteerd wordt zijn scholieren met dyslexie. Onderzoeken van Bonte (2020) en van der Meer (2022) schatten dat ongeveer 15% van de Vlaamse scholieren in het middelbaar onderwijs een vorm van dyslexie heeft. Zo kunnen scholieren met dyslexie bij het intensief lezen geconfronteerd worden op een moeilijke en stoeve automatisering bij het lezen en spellen. Hoewel scholieren met dyslexie ondersteuning kunnen krijgen, mag de impact van leesstoornissen niet onderschat worden. De gevolgen hiervan kunnen zich namelijk doorzetten na het middelbaar onderwijs. Leesvaardigheid blijft daarmee cruciaal voor succes op school en in het werkveld. Scholieren met dyslexie kunnen hebben met spelling, wat kan leiden tot onzekerheid en stress. Daarnaast zijn vooroordelen nog steeds een probleem en kunnen ze leiden tot stigmatisering. Echter toont onderzoek aan dat scholieren met dyslexie doorzettingsvermogen hebben en goede probleemop-

lossers zijn (Bonte, 2020; Ghesquière, 2018; Lissens e.a., 2020).

Het leerplan voor STEM-vakken stimuleert het gebruik van wetenschappelijke artikelen, maar houdt niet altijd rekening met de bijhorende complexe leesgraad. De ingewikkelde woordenschat en syntax in wetenschappelijke artikelen kunnen een hindernis vormen voor de begrijpelijkheid van een tekst, waardoor scholieren met dyslexie de kerninhoud moeilijk kunnen doorgronden. Het handmatig vereenvoudigen van wetenschappelijke artikelen echter planning, tijd en energie van leerkrachten in de derde graad middelbaar onderwijs opsloppen. Het Vlaamse middelbaar onderwijs staat onder druk en docenten hebben moeite om met deze werkdruk boven water te blijven.

Nu is AI technologisch hoogstaand om tekstvereenvoudiging te automatiseren en om een baanbrekende oplossing aan te bieden aan het middelbaar onderwijs. Enerzijds om de werkdruk bij leerkrachten te verminderen, anderzijds om scholieren in de derde graad te ondersteunen bij het lezen van complexe wetenschappelijke artikelen. Voor het gebruik van taalmodellen of API's is uitgebreide informaticakennis nodig, die de meeste scholieren en leraren niet hebben. Anderzijds zijn tools te beperkt en eerder gericht op samenvatten wat niet noodzakelijk bijdraagt tot een eenvoudigere tekst. Er is nood aan een intuïtieve en gebruikersvriendelijke toepassing die taalmodellen of API's kan integreren en aanpassen naargelang de specifieke behoeften van een student met dyslexie.

1.2. Onderzoeksvraag

De volgende onderzoeksvraag is opgesteld: Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de unieke noden van scholieren met dyslexie in de derde graad middelbaar onderwijs?. Om deze onderzoeksvraag te kunnen beantwoorden, moet een antwoord gezocht worden op de volgende deelvragen:

1. Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten? Aanvullend hierop:
 - Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
2. Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging?
 - Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?
 - Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandse geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?

- Hoe kunnen geautomatiseerde tekstvereenvoudiging en gepersonaliseerde tekstvereenvoudiging gecombineerd worden?
3. Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
 4. Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs?
 - Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?”
 5. Welke taalmodellen of LLM's zijn geschikt voor automatische tekstvereenvoudiging voor vereenvoudigde wetenschappelijke artikelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs met dezelfde of gelijkaardige kwaliteiten als manuele tekstvereenvoudiging?
 6. Hoe kan een intuïtieve lokale webtoepassing worden ontwikkeld die zowel scholieren met dyslexie als docenten helpt bij het vereenvoudigen van wetenschappelijke artikelen met behoud van semantiek, jargon en zinsstructuren?

1.3. Onderzoeksdoelstelling

Het onderzoek heeft als doel om de technologische en logopedische aspecten te identificeren die AI-ontwikkelaars in overweging moeten nemen bij het creëren van een op maat gemaakte AI-toepassing voor geautomatiseerde tekstvereenvoudiging, specifiek ontwikkeld voor scholieren in de derde graad. Het resultaat van dit onderzoek is een prototype voor een AI-toepassing voor tekstvereenvoudiging in de vorm van een webtool. De webtool heeft twee functies. Enerzijds kan de tool de inhoud van wetenschappelijke artikelen vereenvoudigen op basis van de specifieke behoeften van scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Anderzijds biedt de tool een geautomatiseerde benadering voor lectoren om wetenschappelijke artikelen te vereenvoudigen op basis van geselecteerde parameters en deze vervolgens in een bruikbaar formaat (pdf of word) terug te geven. De invoer bij dit prototype is een wetenschappelijk artikel in tekst- of PDF-formaat.

1.4. Opzet van deze bachelorproef

De rest van deze bachelorproef is als volgt opgebouwd:

In Hoofdstuk 2 wordt een overzicht gegeven van de stand van zaken binnen het onderzoeksdomein, op basis van een literatuurstudie.

In Hoofdstuk 3 wordt de methodologie toegelicht en worden de gebruikte onderzoekstechnieken besproken om een antwoord te kunnen formuleren op de onder-

zoeksvragen. Eerst wordt er een requirementsanalyse uitgevoerd, gevolgd door de ontwikkeling van een prototype voor tekstvereenvoudiging.

1. Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten? Aanvullend hierop:
 - Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
2. Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging?
 - Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?
 - Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandse geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
 - Hoe kunnen geautomatiseerde tekstvereenvoudiging en gepersonaliseerde tekstvereenvoudiging gecombineerd worden?
3. Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
4. Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs?
 - Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?”
5. Welke taalmodellen of LLM's zijn geschikt voor automatische tekstvereenvoudiging voor vereenvoudigde wetenschappelijke artikelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs met dezelfde of gelijkaardige kwaliteiten als manuele tekstvereenvoudiging?
6. Hoe kan een intuïtieve lokale webtoepassing worden ontwikkeld die zowel scholieren met dyslexie als docenten helpt bij het vereenvoudigen van wetenschappelijke artikelen met behoud van semantiek, jargon en zinsstructuren?

In Hoofdstuk 4 worden de resultaten gegeven op dit onderzoek. In Hoofdstuk 5, tenslotte, wordt de conclusie gegeven en een antwoord geformuleerd op de onderzoeksvragen. Daarbij wordt ook een aanzet gegeven voor toekomstig onderzoek binnen dit domein.

2

Stand van zaken

2.1. Inleiding

Het onderzoek start met een uitgebreide literatuurstudie over de nodige topics binnen het logopedisch, taal- en informatica vakdomein. Om een gepersonaliseerde toepassing voor de tekstvereenvoudiging van wetenschappelijke artikelen aan te reiken, is het van cruciaal belang om de noden van scholieren met dyslexie in de derde graad van het middelbaar onderwijs te benoemen. Het onderzoek benoemt bewezen noden met behulp van een literatuurstudie. Daarnaast moeten de problemen bij huidige wetenschappelijke artikelen ook aangekaart worden.

Wetenschappelijke artikelen vereenvoudigen op maat voor scholieren met dyslexie kan volgens taalexperten op verschillende manieren. Het is belangrijk om stil te staan bij de bestaande en reeds bewezen handmatige tekstvereenvoudigingstechnieken. Vervolgens komen technieken voor geautomatiseerde tekstvereenvoudiging (ATV) aan bod. Om een beter begrip te hebben op deze technieken, wordt de nodige informatie over natuurlijke taalverwerking met AI gegeven.

Een toepassing voor tekstvereenvoudiging met behulp van AI kan hoogstaand zijn, maar het is cruciaal om bij dit onderzoek aandacht te besteden aan de mogelijke problemen die AI-ontwikkelaars moeten vermijden of zich attent op maken.

2.2. Kenmerken van dyslexie en wetenschappelijke artikelen.

Om wetenschappelijke artikelen te vereenvoudigen op maat van de unieke noden van scholieren met dyslexie zijn Deze sectie gaat in op de unieke noden en bespreken hoe mensen met dyslexie kunnen worden geholpen bij het lezen. Daarnaast

worden de belemmeringen en moeilijkheden van wetenschappelijke artikelen aangekaart. Deze sectie beantwoordt de volgende twee onderzoeksvragen:

- Welke specifieke noden hebben scholieren met dyslexie van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
- Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?

2.2.1. Specifieke noden van scholieren met dyslexie in de derde graad van het middelbaar onderwijs.

Leesvaardigheid is geen aangeboren vaardigheid, maar iets dat mensen zelf moeten aanleren (Bonte, 2020; van der Meer, 2022). Hoewel deze herinrichting bij sommige mensen vlot verloopt, kunnen mensen met dyslexie benadeeld worden tijdens dit proces.

Dyslexie wordt gekenmerkt door beperkt lezen en kan het voorlezen traag, radend en letter-voor-letter maken. Mensen met dyslexie kunnen verschillende drempels ervaren; Bonte (2020), Gala en Ziegler (2016), Rivero-Contreras e.a. (2021) en M. Zhang e.a. (2021) sommen de volgende op:

- Trage woordbenoeming
- Problemen met leesbegrip
- Hardnekkig letter-voor-letter lezen
- Problemen met woordherkenning en -herinnering
- Letter- en klankvorming
- Homofonische of pseudo-homofonische woordenschat

Kenmerk	Bron
Trage woordbenoeming	
Leesbegripsproblemen	
Hardnekkig letter-voor-letter lezen	
Problemen met woordherkenning en -herinnering	
Letter- en klankvorming	Duidelijke symbolen gebruiken
Homofonische of pseudo-homofonische woordenschat	
Onregelmatige en infrequente letter-greepcombinaties	Gala en Ziegler (2016)

Onderzoeken van de voorbije twintig jaar, nadrukkelijk met de digitalisering bij kinderen en tieners, halen volgende visuele pijlers aan waaraan een toepassing moet voldoen om rekening te houden met deze doelgroep. Jongeren van tussen 15 en 18 zijn volgens (...) enorm digitaal geletterd, dus de digitale weergave speelt een essentiële rol. Dit artikel omschrijft verder de verschillende technische elementen waaraan een webpagina of toepassing aan moet voldoen om rekening te houden met scholieren met dyslexie. Deze noden worden weergegeven in tabel 2.2.1.

Ingreep	Effect
Zachtgele, -groene of -bruine achtergrondkleur	Consistente lay-out
Woord- en karakterspatiëring	Waarschuwingen geven omtrent formulieren, sessies (login)
Consistente lay-out	Duidelijk zichtbare koppen- of heading-structuur
Koppenstructuur	Inhoud visueel groeperen
Huidige positie benadrukken	Duidelijke symbolen gebruiken

2.2.2. Specifieke kenmerken van wetenschappelijke artikelen

Wetenschappelijke artikelen volgen IMRAD¹ en worden gebruikt als leermiddel voor jongeren in het middelbaar en hoger onderwijs.

Teksten inschatten op leesbaarheid kan niet enkel op intuïtie gebeuren. Mensen beschikken over verschillende achtergrondkennis, daardoor zal de ene persoon eerder geneigd zijn om een woord buiten het gekende jargon als moeilijk te beschouwen. Als oplossing zijn er over de decennia heen leesbaarheidsscores opgesteld om teksten objectief op leesbaarheid te kunnen beoordelen. De twee meest gekende leesbaarheidsscores worden in tabel 2.2.2 weergegeven. Python-libraries, zoals Textstat², maken het herschrijven van de formules obsoleet en bieden een snelle oplossing voor ontwikkelaars om één van de volgende leesgraadsscores te kunnen berekenen. Nadrukkelijk houden deze scores geen rekening met de achtergrondkennis van de mogelijke lezer(s).

¹Uniform formaat voor gepubliceerde wetenschappelijke artikelen. De structuur bestaat uit vijf hoofdstukken: inleiding, methodologie, resultaten en discussie.

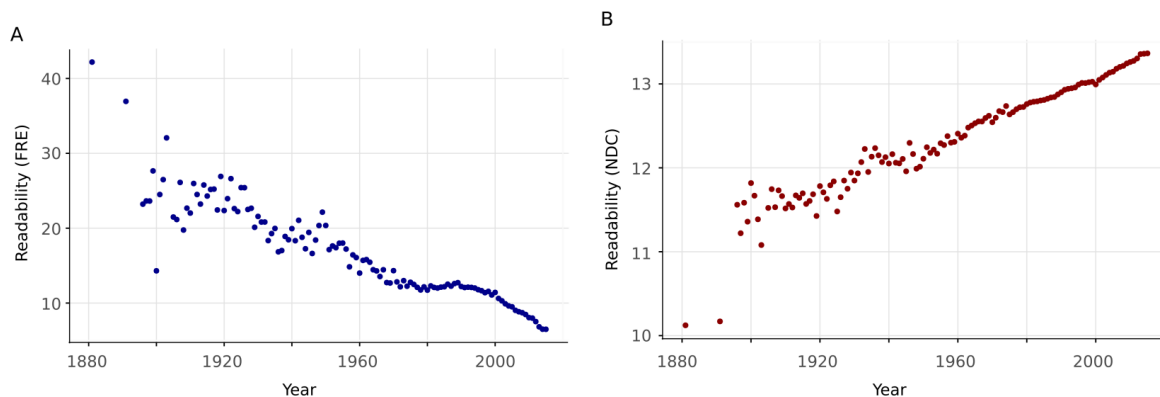
²textstat

Score	Uitleg
Flesch Reading Ease (FRE)	Deze leesbaarheidsscore wordt berekend op zinbasis. Hoe hoger de score, hoe 'eenvoudiger' de tekst.
Gunning FOG (FOG)	Dit wordt berekend op tekstbasis.

2.2.3. Trends van wetenschappelijke artikelen

De leesgraad van wetenschappelijke teksten volgt al sinds de tweede helft van de twintigste eeuw een stijgende trend (Hayes, 1992). Meerdere onderzoeken in de voorbije tien jaar besluiten dat de complexe woordenschat en zinsbouw deze wetenschappelijke artikelen ontoegankelijk maken voor doelgroepen naast onderzoekers (Ball, 2017; Jones e.a., 2019; Plavén-Sigra e.a., 2017).

Het onderzoek van Plavén-Sigra e.a. (2017) beklemtoont de stijgende FRE-scores voor wetenschappelijke artikelen. Door middel van een vergelijkende studie tussen abstracten en de overige inhoud van wetenschappelijke journals, kon de onderzoekers besluiten dat de abstract het meest complexe deel van een wetenschappelijk artikel is. Op figuur 2.1 wordt de evolutie per FRE (links) en NDC-score (rechts) getoond. Het onderzoek schat dat nu 22% van alle wetenschappelijke artikelen gebruik maken van Engels op het niveau van een masterstudent, vergeleken met 14% in 1960.



Figuur (2.1)

Afbeelding uit Plavén-Sigra e.a. (2017)

Onbegrijpelijke en ontoegankelijke zinsstructuren hinderen ook vakexperten. Zo toonde onderzoek van McNutt (2014) aan dat begrip van de methodologie en resultaten cruciaal is in het kader van reproduceerbaarheid; enkel zo kunnen wetenschappers op correcte wijze een studie reproduceren en wetenschappelijke inzichten bevestigen of met verdere resultaten verrijken. Experimenten van Hubbard en

Dunbar (2017) wijzen namelijk uit dat het net vooral de methodologie en resultaten van een wetenschappelijk artikel zijn die een hoge leesgraad vergen. In deze context is ook het onderzoek van Hartley (1999) en Snow (2010) relevant waarin ze aantonen dat het herschrijven van abstracts de begrijpbaarheid ervan kan verhogen.

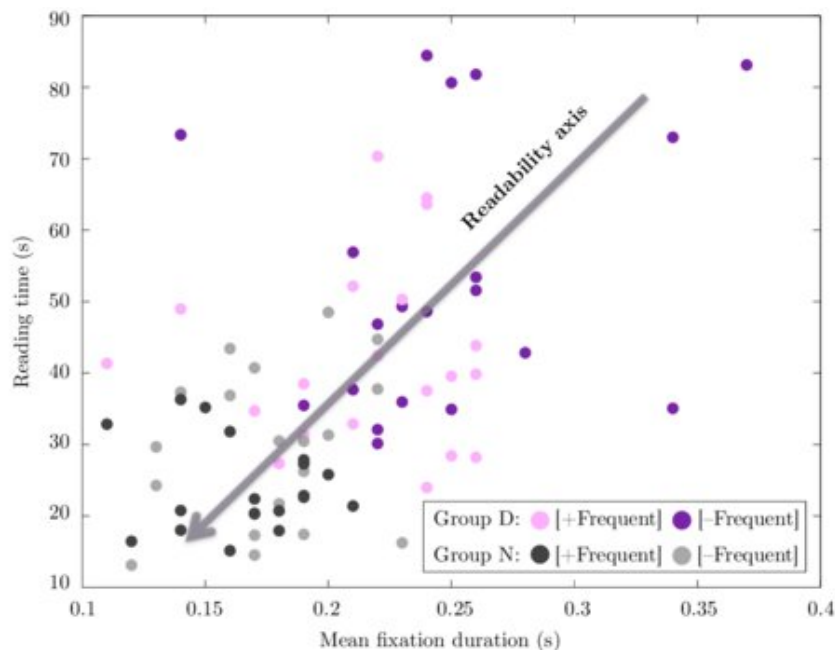
2.3. Aanpakken voor tekstvereenvoudiging

2.3.1. Manuele tekstvereenvoudiging

In alle tijde moet een vereenvoudigd of samengevat wetenschappelijk artikel volgens Hollenkamp (2020) en McCombes (2022) drie vragen kunnen beantwoorden: waarom werd het onderzoek verricht, wat werd er geëxperimenteerd en welke conclusies trekken de onderzoekers? Dit omvat de achtergrondinformatie, hypothesen, methoden, resultaten, implicaties, beperkingen en aanbevelingen. Om de tekst te vereenvoudigen, kan de tekst naar een ander formaat worden omgezet, zoals *post-it*notes, tabelvorm of opsommingen om tekst begrijpelijker te maken (Rijkhoff, 2022).

Manuele tekstvereenvoudiging (MTV).

Lexicale vereenvoudiging	Moeilijke woorden vervangen door eenvoudiger synoniemen Woorden- en synoniemenlijst maken Dubbelzinnige woorden vervangen Idiomen vervangen Rekening houden met het gekende jargon van de doelgroep
Syntactische vereenvoudiging	Tangconstructies aanpassen Zinnen langer dan negen woorden inkorten Verwijswoorden aanpassen Voorzetseluitdrukkingen aanpassen Samengestelde werkwoorden aanpassen Actieve stem gebruiken Onregelmatige werkwoorden vermijden
Formaataanpassingen	Marges aanpassen Lettertype en -grootte aanpassen Woord- en karakterspatiëring aanpassen Herschrijven en -structureren als opsomming of tabelvorm

**Figuur (2.2)**

Afbeelding van Rello, Baeza-Yates, Dempere-Marco e.a. (2013)

2.3.2. Bevoordelende effecten op handmatige tekstvereenvoudiging bij scholieren met dyslexie

Het onderzoek van Rivero-Contreras e.a. (2021) laat zien dat vereenvoudigde teksten de leesnelheid en woordherkenning van kinderen met visuele dysfunctie significant kunnen verbeteren. Uit het experiment van Rello, Baeza-Yates, Dempere-Marco e.a. (2013) blijkt dat frequent woordgebruik de ontcijfertijd bij mensen met dyslexie significant vermindert en dat bevroegden met dyslexie minder leesfouten maken bij teksten met een verminderde lexicale complexiteit, volgens Gala en Ziegler (2016). Het experiment benadrukt ook de moeilijkheden van kinderen met dyslexie bij het lezen van woorden met meer dan zeven karakters en onregelmatige en infrequente lettergreepcombinaties. Volgens de richting van de pijl op figuur 2.2 wordt de ideale situatie benaderd, gekenmerkt door doelwaarden. Deze waarden worden bereikt door mensen zonder dyslexie onder optimale omstandigheden. Het gebruik van vaak voorkomende woorden vermindert de decodeertijd en verbetert de leesbaarheid voor mensen met dyslexie.

Onderzoeken rond de effecten op syntactische vereenvoudiging bij kinderen en scholieren met dyslexie zijn in schaars. Het aanpassen causale structuren bij kinderen en jongeren met een lage leesgraad had een significant effect op het leestempo en de foutenmarge van de bevroegden uit het experiment van Linderholm e.a. (2000). Bij de revisies werden coherentieonderbrekingen hersteld door extra uitleg te voorzien, alsook door tekstgebeurtenissen in een temporele of tijdsaf-

hankelijke volgorde te plaatsen. Zowel vaardige als minder vaardige lezers hadden baat bij de revisies. Verbale parafrasering heeft geen significant effect op lezers met dyslexie volgens Rello, Baeza-Yates en Saggion (2013). De bevroagden waren 13 tot en met 37 jaar oud met een gemiddelde leeftijd van 21 jaar. Het tekstformaat bleef ongewijzigd, maar lettertypes werden wel aangepast.

Geassisteerd samenvatten bevoordeelt de leesbaarheid van een scholier met dyslexie volgens het experiment van Nandhini en Balasundaram (2013). De geassisteerde samenvatting is gebaseerd op onaangepaste zinnen afkomstig uit de oorspronkelijke tekst. Het ontwerp bij dit experiment haalt de belangrijkste zinnen onaangepast uit de oorspronkelijke tekst, herorganiseert deze volgens de structuur van de oorspronkelijke tekst en presenteert deze aan de lezer. Al werd de logische structuur van de gepresenteerde zinnen in vraag gesteld, de leesbaarheid van de bevroagden was significant beter dan bij de oorspronkelijke tekst zonder een nadelig effect op de verstaanbaarheid van de bevroagden.

Experimenten tonen aan dat scholieren met dyslexie gevoeliger zijn voor veranderingen in visuele parameters zoals lettertype, karakterafstand, tekst- en achtergrondkleur en grijswaarden. Aanbevolen kleurencombinaties zijn een lichtgrijze achtergrond met zwart lettertype op een gele achtergrond, of zachtgele, -groene of lichtblauwe achtergrondkleuren (Bezem & Lugthart, 2016; Rello & Baeza-Yates, 2015; Rello & Bigham, 2017). Minimalistische ontwerpen met pictogrammen en afbeeldingen hebben een positief effect op de leesbaarheid van tekst. Lettergrootte groter dan 14pt en een sans-serif, *monospaced* of *roman* lettertype vergroten de leessnelheid. Het gebruik van lettertypes zoals OpenDys heeft geen effect op lezers met of zonder dyslexie, terwijl cursieve lettertypes worden afgeraden (Rello & Baeza-Yates, 2013; Rello & Baeza-Yates, 2015).

2.3.3. Aanpak voor ATV.

Tekstvereenvoudiging is het proces waarin het technisch leesniveau en/of woordgebruik van een geschreven tekst wordt verminderd. Het resultaat van deze fase is een tekst die korter en aangenamer is, zonder het verlies van de kerninhoud. Binnen machinaal leren (ML) is tekstvereenvoudiging een zijtak van natuurlijke taalverwerking. (Siddharthan, 2006) Volgens (Siddharthan, 2014) bestaat een complete en geautomatiseerde tekstvereenvoudiging uit vier verschillende vereenvoudigingen. *Natural Language Processing* (NLP) of natuurlijke taalverwerking is een brede term die zich richt op het verwerken en analyseren van menselijke taal door computers (Eisenstein, 2019). NLP omvat verschillende technieken, zoals tekstanalyse, taalherkenning en -generatie, spraakherkenning en -synthese, en semantische analyse. Computers zijn in staat om op een menselijke manier te communiceren en begrijpen wat er wordt gezegd. De volgende begrippen worden aangehaald

in Eisenstein (2019) en Sohom (2019) en zijn fundamenteel voor de concepten die volgen.

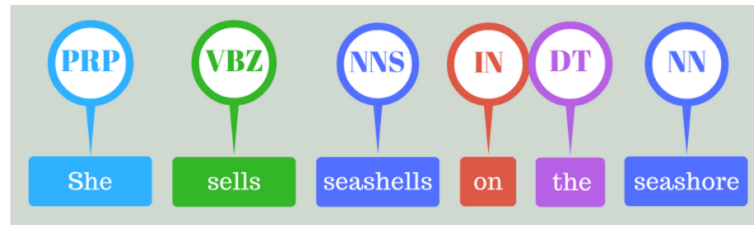
Tokenisatie splitst tokens in een tekst en bouwt zo een woordenschat voor een taalmodel op. Dit kan volgens Menzli (2023) op vier manieren: woord-, karakter-, subwoord- en zinniveau gebeuren. Bij karakter-tokenisatie wordt de inputlengte groter en heeft daarmee volgens Ribeiro e.a. (2018) weinig betekenis. Zeldzame woorden worden opgesplitst in kleinere stukken om een woordenschat op te bouwen. Dit biedt voordelen ten opzichte van word-level tokenisatie (Iredale, 2022).

Lemmatiseren in NLP bouwt verder op *stemming*, maar de betekenis van ieder woord wordt in acht genomen. Voor het lemmatiseren bestaan er Nederlandstalige modellen, waaronder JohnSnow³. Bij omgekeerd lemmatiseren wordt er een afgeleide achterhaald vanuit de stam. Voor zelfstandige naamwoorden, zoals 'hond', is dit enkelvoud of meervoud (Eisenstein, 2019). Bij de parsingfase wordt een label aan ieder woord of zinsdeel toegekend. Voorbeelden van labels zijn zelfstandig naamwoord, bijwoord, werkwoord, bijzin of stopwoord. Het herkennen van zinsdelen wordt *chunking* genoemd. Parsing heeft een dubbelzinnigheidsprobleem, want een 'plant' staat niet gelijk aan de vervoeging van werkwoord 'planten' (Eisenstein, 2019).

Een machine moet de betekenis achter ieder token kunnen vatten. Hier komt *sequence labeling* aan de pas volgens Eisenstein (2019). Elk woord in een tekst wordt gekoppeld aan een *Part-of-Speech* (PoS) of *Named-Entity-Recognition* (NER) label. Deze NLP-fase achterhaalt de structuur van een tekst. PoS-tagging richt zich op grammaticale categorieën van woorden, terwijl NER-labeling instaat voor het herkennen van specifieke entiteiten in een tekst. Bij PoS-tagging worden de woorden in een zin geanalyseerd. Elk woord wordt gekoppeld aan een grammaticale categorie, zoals een zelfstandig naamwoord, werkwoord, bijvoeglijk naamwoord of bijwoord. *PoS-tagging* helpt bij het achterhalen van de syntactische structuur van een zin. Deze taak komt van pas bij parsing en machinevertaling. *PoS-tagging* wordt aanschouwelijk gemaakt op 2.3. Namen van personen, organisaties en locaties worden herkend en geclassificeerd met NER-labeling. Met NER-labeling wordt volgens Jurafsky e.a. (2014) specifieke informatie uit tekst gehaald, zoals het identificeren van de namen van personen, plaatsen of bedrijven die in nieuwsartikelen worden genoemd, of het extraheren van belangrijke data of getallen uit financiële rapporten. Dit wordt aanschouwelijk gemaakt ?? J. Li e.a. (2018) benoemt vier vormen voor NER-labeling: *dictionary-based*, *rule-based*, *ML-based* en *deep learning-based*. De eerste twee gebruiken vooraf gedefinieerde woordenboeken en regels, terwijl de laatste twee gebruik maken van statistische of neurale netwerken om

³https://nlp.johnsnowlabs.com/2020/05/03/lemma_nl.html

te leren hoe entiteiten te herkennen. Elke vorm gebruikt verschillende kenmerken en representaties om entiteiten te modelleren. Poel e.a. (2008) onderzocht *PoS-tagging* met een neurale netwerk voor Nederlandstalige teksten. Het model behaalde een nauwkeurigheid van 97,88% voor bekende woorden en 41,67% voor onbekende woorden en gebruikte de Corpus Gesproken Nederlands (CGN) als trainingsdata.



Figuur (2.3)

Voorbeeld van PoS-labeling op de Engelstalige zin "She sells seashells on the seashore". Afbeelding van Bilici (2021)

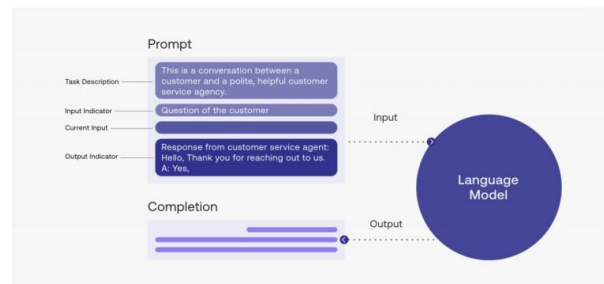
NLP-systemen gebruiken embeddings om woorden numeriek te representeren en tekst te verwerken. Traditionele word embeddings bouwen een woordenschat op zonder de betekenis ervan op te volgen, terwijl contextual word embeddings wel de context van een woord begrijpen. BERT is een meertalig LLM dat contextual word embeddings gebruikt en getraind is op 110 miljoen parameters uit 104 verschillende talen, waaronder Nederlands. Voor de Nederlandse taal zijn er twee varianten van BERT⁴, namelijk RobBERT en BERTje, waarvan RobBERT als krachtiger wordt beschouwd. Om de beste vervanging van woorden te bepalen, gebruikt het model de Substitution Ranking (SR) stap om substituties op basis van relevantie te rangschikken.

2.3.4. Prompt engineering

Large Language Models of LLM's genereren tekst en karakters op basis van de waarschijnlijke uitkomst van een gegeven input. De meest gekende LLM's zijn GPT-3, BERT en T5 en dienen regelmatig als de fundering voor ge-fine-tuned modellen. Deze modellen maken gebruik van een neurale netwerk om patronen in de input te herkennen en deze patronen te gebruiken om voorspellingen te doen over de uitvoer (Liu e.a., 2020). Iedereen kan volgens McFarland (2023) een input of prompt schrijven. Deze tools zoals chatbots zijn ontworpen om zo intuïtief mogelijk te zijn voor een algemeen doelpubliek. Prompt engineering is een steeds belangrijkere vaardigheid die nodig is om effectief te communiceren met LLM's, zoals ChatGPT (Harwell, 2023). Deze vaardigheid wordt geïllustreerd in 2.4.

Deze prompts kunnen volgens Liu e.a. (2020) gebruikt worden om werk te produceren dat is aangepast aan het doel. Een concrete en geoptimaliseerde prompt

⁴<https://github.com/google-research/bert>

**Figuur (2.4)**

Afbeelding uit McFarland (2023)

omvat een concrete scope, duidelijke vraagstelling, specifieke sleutelwoorden, de context en ten slotte gepersonaliseerde keuzes (McFarland, 2023). Bij een zoekopdracht moeten voldoende parameters in de prompt worden opgenomen. Zo niet zal het model te algemeen blijven en mogelijks afwijken van de intentie van de gebruiker. Effectieve AI prompt engineering leidt tot hoogwaardige trainingsgegevens die het AI-model in staat stellen om nauwkeurige voorspellingen en beslissingen te maken (Liu e.a., 2020).

Prompt patterns is samen met prompt engineering naar boven gekomen en is vergelijkbaar met software patterns. Deze patronen zijn herbruikbare oplossingen voor veelvoorkomende problemen in een bepaalde context, waaronder vooral de interactie bij het werken met LLM's. White e.a. (2023) benoemt vier *prompt patterns*:

- Intent-prompts waarbij een LLM een instructie krijgt met een specifiek verwacht antwoord.
- Restriction-prompts die het antwoord van een LLM inperkt. Deze pattern is noodzakelijk om een LLM binnen de lijnen te houden.
- Contextualization-prompts verzekeren dat de output van een LLM relevant is. Een context wordt aan de LLM meegegeven.
- Expansion/reduction-prompts genereren een beknopte output met voldoende details.

2.4. De verschillende soorten tekstvereenvoudiging

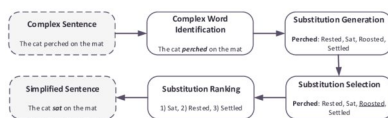
Tekstvereenvoudiging bestaat volgens Siddharthan (2014) uit vier soorten transformaties: lexicale, syntactische en semantische vereenvoudiging en samenvatten.

Bij *lexical simplification* (LS) of lexicale vereenvoudiging worden complexe woorden vervangen door eenvoudigere synoniemen. Bijvoorbeeld, het woord 'adhesief'

wordt vervangen door 'klevend'. Kandula e.a. (2010) haalt twee manieren aan om lexicale vereenvoudiging mogelijk te maken, namelijk het vervangen door een synoniem en het aanmaken of genereren van extra uitleg. De zinsstructuur verandert niet en er is garantie dat de kerninhoud en benadrukking in een tekst identiek blijft. Het doel van lexicale vereenvoudiging is om de moeilijkheidsgraad van de woordenschat in een zin of tekst te verlagen.

Complex word identification (CWI) is een gesuperviseerde NLP-taak. In een pipeline voor lexicale tekstvereenvoudiging is CWI de eerste stap. Moeilijke woorden of *multi-word expressions* (MWE) in een tekst worden achterhaald (Gooding & Kochmar, 2019; Shardlow, 2013). Na CWI kan LS gebruikt worden om deze woorden te vervangen door eenvoudigere synoniemen of om verdere elaboratie te voorzien met behulp van voorbeelden of definities (Kandula e.a., 2010; Zeng e.a., 2005). CWI is volgens Shardlow (2013) een cruciale stap, want een lage *recall* van dit component zal een uitvoertekst geven waar moeilijke woorden niet worden vereenvoudigd. Het model zal moeilijke woorden laten staan.

Substitutiegeneratie wordt gedaan door synoniemen te zoeken voor een doelwoord in lexicale databanken zoals WordNet, BERT, context2vec, nPIC of OOC.



Figuur (2.5)

Afbeelding van Althunayyan en Azmi (2021). Deze pipeline wordt in meerdere onderzoeken rond lexicale vereenvoudiging toegepast, zoals Bingel e.a. (2018), Bulté e.a. (2018) en Paetzold en Specia (2016)

Syntactische vereenvoudiging vermindert de complexiteit van een tekst door de grammatica en zinsstructuur aan te passen. Dit kan door het combineren van twee zinnen tot één eenvoudigere zin of door de syntax te vereenvoudigen. Deze transformaties verbeteren de toegankelijkheid van de tekst zonder de inhoud te verliezen. Kandula e.a. (2010) ontwikkelden een model om medische informatie te vereenvoudigen. Dit model omvat drie modules, die zinnen met meer dan tien woorden vereenvoudigen en eventueel vervangen door kortere zinnen. De architectuur omvat een PoS Tagger, een Grammar Simplifier en een Output Validator.

- Voor de *PoS Tagger*-fase gebruikten Kandula e.a. (2010) beschikbare functies uit het open-source pakket OpenNLP⁵.

⁵<https://opennlp.apache.org/>

- De *Grammar Simplifier* module splitst de lange zin in twee of meer kortere zinnen door POS-patronen te identificeren en een set transformatieregels toe te passen.
- De *Output Validator* module controleert de output van de Grammar Simplifier op grammatica en leesbaarheid.

Geautomatiseerde tekstvereenvoudiging is geen nieuw concept. Volgens onderzoeken van Canning e.a. (2000) en Siddharthan (2006) waren de eerste aanpakken op geautomatiseerde tekstvereenvoudiging gebouwd op rule-based modellen. Deze modellen bewerken de syntax door zinnen te splitsen, te verwijderen of de volgorde van de zinnen in een tekst aan te passen. Lexicale vereenvoudiging kwam hier niet aan de pas. Enkel bij recentere onderzoeken van Bulté e.a. (2018) en Coster en Kauchak (2011) werd het duidelijk hoe lexicale en syntactische vereenvoudiging gecombineerd kon worden.

Lexicale, conceptuele en/of syntactische vereenvoudiging van teksten leidt niet altijd tot een kortere tekst. Technologieën zoals full-text-search en gepersonaliseerde informatiefiltering benadrukken het belang van gebruikersgerichte samenvatting. De architectuur van een samenvattingssysteem omvat drie fases: analyse van de brontekst, identificatie van kernpunten en het samenvoegen van de punten tot één uitvoertekst. Teksten machinaal samenvatten is geen nieuw concept en kan op twee manieren gebeuren: extraherend en abstraherend (DuBay, 2004; Hahn & Mani, 2000).

Bij extraherende samenvatting worden de belangrijkste zinnen gemarkeerd en opnieuw neergeschreven, maar dit kan leiden tot onsamenhangende uitvoertekst. Kernzinnen achterhalen gebeurt volgens Khan (2014) op basis van woordfrequentie, zinpositie -en gelijkenissen, de *cue method*, titels, de rest van het document, *proper nouns* woordgebruik en ten slotte de afstand tussen *text units* waarin entiteiten voorkomen. Verma en Verma (2020) onderzocht verschillende manieren om een tekst extraherend samen te vatten, waaronder graafgebaseerd, maximal marginal relevance (MMR) en metaheuristiek-gebaseerde ES. Deze worden verder verwoord in 2.4.

MMR-gebaseerde ES	Deze traditionele techniek gebruikt de maximaal marginale relevantiescore (MMR) om de relevantie en diversiteit van gemarkeerde zinnen te bepalen. Zorgt ervoor dat de geselecteerde zinnen niet te veel overlappen in inhoud en relevantie. Kan leiden tot betere samenvattingen, maar vereist meer rekenkracht en tijd.
Graafgebaseerde ES	Deze techniek vertegenwoordigt een document als een graaf van zinnen en gebruikt algoritmen om de belangrijkste zinnen te bepalen en redundantie te vermijden. Kan zowel voor lange wetenschappelijke artikelen als korte nieuwsartikelen goede resultaten opleveren (Lin & Bilmes, 2010; McDonald, 2007).
Metaheuristiek-gebaseerde ES	Deze techniek maakt gebruik van optimalisatie-algoritmen zoals genetische algoritmen en zwermoptimalisatie om de belangrijkste zinnen in een tekst te vinden (Premjith e.a., 2015; Verma & Verma, 2020). Evaluatiefunctie kan in een lokaal optimum vastlopen afhankelijk van de gebruikte criteria. (Rani & Kaur, 2021).

McKeown e.a. (1999) voerden experimenten uit op extraherende samenvattingen van nieuwsartikelen. De resultaten wijzen erop dat deze vorm vatbaar is op vooroordelen of *bias* van de auteur. De zinnen worden genomen zoals ze zijn. Hahn en Mani (2000) bouwde verder op dit experiment door knowledge-rich en knowledge-poor methoden te combineren met significante resultaten ter gevolg. De nadruk bij extraherend samenvatten ligt in het kiezen van de *salient text units*. Deze punten zijn typisch in de vorm van zinnen. Er is nood aan een manier om de lexicale en statistische relevantie van een zin te kunnen aanduiden. Hiervoor haalt Hahn en Mani (2000) twee manieren aan:

- Een lineair gewicht model. Iedere teksteenheid wordt gewogen op factoren zoals de *location weight* en het aantal voorkomens.
- Een gewicht model op basis van de statistische opvallendheid van een eenheid. Zo wordt er rekening gehouden met de aanwezigheid van een woord in (sub)titels.

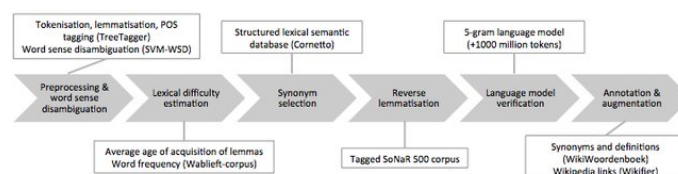
Nallapati e.a. (2017) wilden de nauwkeurigheid van deze modellen overbruggen. Dit doen ze met *SummaRuNNer*⁶, een oplossing voor het extraherend samenvatten van teksten met een neurale netwerk. De toepassing werd opgebouwd met *PyTorch* in en bestaat uit een combinatie van drie modellen: een recurrent neurale

⁶<https://github.com/hpzhao/SummaRuNNer>

netwerk, een convolutioneel recurrent neurale netwerk en een *hiërarchische attention network*. Al behoort samenvatten niet tot het vereenvoudigen van een tekst, het is een techniek die nodig is om met zo min mogelijk karakter- of woordtokens de semantiek achter een tekst te kunnen bevatten.

De kernzinnen ophalen kan leiden tot een incoherente tekst. In de taalwereld worden kernzinnen vaak geparafraseerd om zo samenhang tussen de zinnen te maken. Er zijn twee manieren om een tekst abstraherend samen te vatten: semantisch en structuurgebaseerd. De structuurgebaseerde benadering gebruikt regels om belangrijke informatie in de tekst te vinden en kan leiden tot samengevatte zinnen met lage linguïstische kwaliteit en grammaticale fouten. De semantisch-gebaseerde benadering gebruikt de betekenis van de tekst om korte en duidelijke samenvattingen te maken met minder redundante zinnen en betere linguïstische kwaliteit, hoewel een extra parsingfase nodig kan zijn. Cao (2022) heeft verder onderzoek gedaan naar deep learning methoden om automatisch abstraherende samenvattingen te genereren. Deep learning-modellen zoals RNN's, CNN's en Seq2Seq kunnen worden gebruikt voor abstraherend samenvatten door de betekenis van de tekst te begrijpen en belangrijke informatie over te brengen (Suleiman & Awajan, 2020). Het Pegasus-model, beschreven in J. Zhang e.a. (2020), maakt gebruik van pre-trained modellen voor samenvatting met NLP en handelt gap-zinnen af, en is getraind en beoordeeld op verschillende soorten samenvattingstaken. LED of Longformer Encoder-Decoder is specifiek ontworpen om lange documenten te verwerken, waardoor het geschikt is voor het samenvatten van langere wetenschappelijke artikelen. In het best denkbare geval wordt abstraherende en extraherende samenvatting gecombineerd volgens Hsu e.a. (2018) en Huang e.a. (2019). Zo omvat een pipeline voor hybride samenvatting twee onderdelen: een *content selection* fase waarin de kernzinnen met extraherende samenvatting worden opgehaald en *paraphrasing*-fase waarbij de gemarkeerde kernzinnen abstraherend worden samengevat.

2.4.1. Ondersteunende en bevoordelende effecten van ATV voor leerlingen met dyslexie



Figuur (2.6)

Afbeelding uit Bulté e.a. (2018)

**Figuur (2.7)**

Afbeelding uit Bulté e.a. (2018)

Al zijn er onderzoeken over lexicale, syntactische en semantische vereenvoudiging voor kinderen en scholieren met dyslexie, het aantal onderzoeken over samenvatten voor deze doelgroep is schaars. Zoals eerder aangehaald is er wel onderzoek gedaan naar de verschillende manieren om een tekst samen te vatten, maar er is geen toepassing of onderzoek dat dit concreet uitwerkt.

2.4.2. Conclusie

2.4.3. MTV en ATV combineren.

(...) wijzen erop dat toepassingen voor tekstvereenvoudiging regelmatig als *showcase* van de technologie ontwikkeld worden en zelden tot weinig rekening houden met personalisatieopties die nodig zijn om een vereenvoudiging op maat te kunnen genereren.

Een ondersteunende toepassing moet met een individuele analyse van de specifieke behoeften en uitdagingen van elke leerling in gedachten worden ontworpen (Gooding, 2022). Het is belangrijk om te erkennen dat dyslexie zich bij verschillende kinderen op verschillende manieren kan uiten. Een bijkomende stoornis heeft bijvoorbeeld geen impact op de spellingprestaties van een kind. Het is daarom belangrijk om een toepassing te ontwerpen met de diversiteit van dyslexie in het achterhoofd.

2.4.4. Trends bij ATV

2.5. De valkuilen bij AI en NLP.

AI en ML zijn volop in groei. NLP gebruikt AI en ML om menselijke taal te verwerken, terwijl NLU deze technologieën gebruikt om menselijke taal te begrijpen. Hoewel deze technologieën veelbelovend zijn, moeten AI-ontwikkelaars rekening houden met veelvoorkomende en genegeerde uitdagingen en valkuilen (Khurana e.a., 2022; Roldós, 2020; Sciforce, 2020). Deze sectie beantwoordt de volgende onderzoeksvraag: "Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?"

NLP- en NLU-toepassingen behoren tot de duurste om te ontwikkelen, wat een

obstakel kan vormen voor veel IT-professionals. Het gebrek aan NLP-expertise, de kwaliteit en kwantiteit van data, de integratie en deployment van modellen en de transparantie van modellen zijn allemaal factoren die bijdragen aan deze hoge kosten (IBM, 2022). Software-ontwikkelaars verkiezen volgens voor *black-box* modellen bij de ontwikkeling en finetuning van een NLP-toepassing met AI. AI is het verschil qua nauwkeurigheid minimaal, de afweging wordt gemaakt bij de transparantie van het model. Na een transformatie wordt er niet aangegeven waarom specifieke transformaties werden uitgevoerd, bijvoorbeeld het vervangen van een woord door een eenvoudiger synoniem. White-box taalmodellen zijn er in schaarse hoeveelheden (Sikka & Mago, 2020).

Homoniemen kunnen volgens Roldós (2020) problemen veroorzaken bij sequence labeling of het labelen van tokens in een doorlopende tekst. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. Word Sense Disambiguation (WSD), PoS-tagging en contextual embeddings kunnen de betekenis van een woord achterhalen op basis van de context (Eisenstein, 2019; Liu e.a., 2020). Het gebruik van synoniemen en antoniemen in NLP-systemen kan verbeterd worden door het gebruik van candidate generation en synonym detection, en meertalige transformers zoals BERT bieden een oplossing voor het gebrek aan niet-Engelstalige toepassingen (Dandekar, 2016; Roldós, 2020).

Tekstvereenvoudiging is bedoeld om gelijke kansen te bieden aan iedereen, maar ethische overwegingen en bewustzijn van de behoeften van de eindgebruiker zijn belangrijk bij het ontwikkelen van adaptieve tekstvereenvoudigingstoepassingen, zoals beschreven in onderzoeken van Gooding (2022), Niemeijer e.a. (2010) en Xu e.a. (2015). De eindgebruiker moet de keuze hebben om te kiezen welke delen van de tekst vereenvoudigd moeten worden, wat kan worden bereikt door synoniemen te kiezen of zinnen te markeren die moeilijk te begrijpen zijn.

Iedereen kan converseren met een chatbot, maar de gepaste en verwachte antwoorden krijgen vergt een doordachte input. Een onnauwkeurige prompt of gebrek aan trainingsdata kan leiden tot onjuiste output, terwijl het gebruik van conditionele expressies of finetunen van hyperparameters kan helpen de betrouwbaarheid van het antwoord te vergroten (Jiang, 2023; Miszczak, 2023).

Vergeleken met andere ML- of NLP-taken vergt het beoordelen van een vereenvoudigde tekst voldoende aandacht en opvolging van de ontwikkelaar. Evaluatiemetrieken zoals ROUGE en BLEU zijn beperkt, omdat ze geen rekening houden met de semantiek tussen een referentietekst en een vereenvoudigde tekst. Als valnet kunnen ontwikkelaars menselijk evaluatie inschakelen om de vereenvoudigde

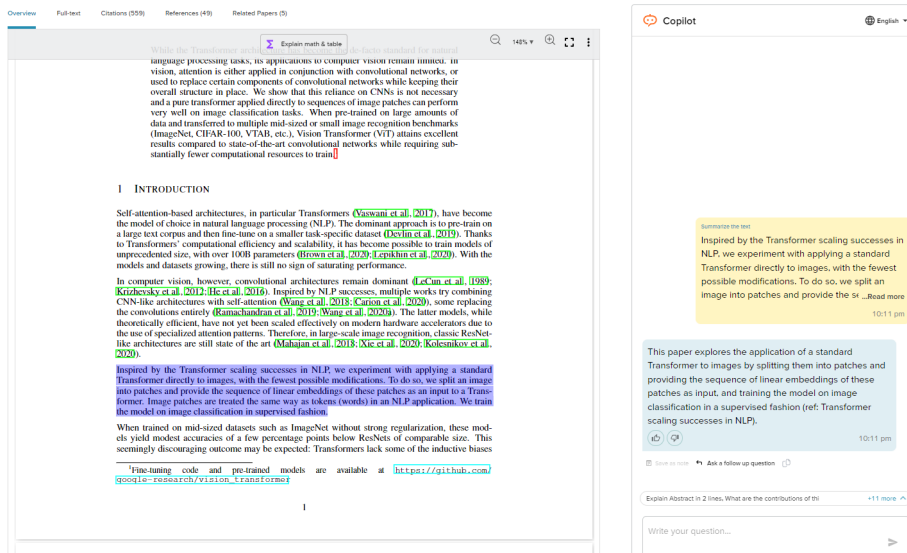
tekst van een taalmodel te beoordelen volgens Fabbri e.a. (2020). De onderzoekers stimuleren verder onderzoek naar nieuwe standaarden en *best practices* voor betrouwbare menselijke beoordeling. De doelgroep waarvoor een tekst wordt vereenvoudigd, moeten nauw in het proces worden opgenomen (Iskender e.a., 2021).

2.6. Beschikbare tools en taalmodellen

Dyslexie is een veelvoorkomende aandoening die de lees- en schrijfvaardigheden van scholieren kan belemmeren. Om deze scholieren te ondersteunen, worden er verschillende softwareprogramma's en tools ontwikkeld. In dit hoofdstuk zal worden gekeken naar mogelijke nationale en internationale software die specifiek is ontworpen om scholieren met dyslexie te helpen bij het lezen van teksten. Er zal met name worden gekeken naar de beschikbare software in Vlaamse middelbare scholen, chatbots, zoals Bing AI en ChatGPT, en software die speciaal is ontwikkeld om dyslexie te ondersteunen bij het lezen. Deze sectie beantwoordt de volgende onderzoeksvraag: "Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?"

In het middelbaar onderwijs wordt lees- en studieondersteuning voor scholieren met dyslexie enkel in de vorm van voorleessoftware voorzien (De Craemer e.a., 2018; OnderwijsVlaanderen, 2023). OnderwijsVlaanderen (2023) leent licenties voor de volgende softwarepakketten uit SprintPlus, Kurzweil3000, Alinea Suite, IntoWords en TextAid. Naast luister- en schrijfopties kunnen scholieren deze toepassingen gebruiken om zinnen te markeren om deze zinnen vervolgens samen te vatten. Enkel de gemarkeerde zinnen worden betrokken in de samengevatte versie, dus de zinnen blijven lexicaal, syntactisch en semantisch identiek. Alle vermelde softwarepakketten bieden echter geen onafhankelijke samenvat- of vereenvoudigfunctie aan. Tops e.a. (2018) benadrukt de handige aspecten van deze software, maar deze software moet zo vroeg mogelijk in een schoolcarrière worden ingezet. Zo raken de scholieren snel vertrouwd met het gebruik, wat kan leiden tot een optimaal gebruik in verdere studies. Volgens Tops e.a. (2018) is het te laat om deze software pas in het hoger onderwijs te introduceren.

Online zijn er tools beschikbaar om teksten generiek samen te vatten. Resoomer, Paraphraser en Scholarcy zijn oorspronkelijk Engelstalige tools, met ondertussen de mogelijkheid om een abstraherende samenvatting te maken van Nederlandstalige teksten. Proof-of-concepts zijn in schaarse hoeveelheid beschikbaar. De taalmodellen waar deze applicaties op werken, is niet gekend. Daarnaast zijn er ook geen API's beschikbaar om mee te werken. Gepersonaliseerde toepassingen zijn er in mindere mate. Bingel e.a. (2018) omschrijft een proof-of-concept voor een webtoepassing dat teksten vereenvoudigd, met oog op mensen met dyslexie. Deze



Figuur (2.8)
Schermabbeelding van SciSpace

software noemt nu Hero en bevindt zich in betafase.

Toepassingen om wetenschappelijke artikelen te vereenvoudigen zijn schaars, maar er zijn enkele gratis en betalende toepassingen beschikbaar. SciSpace⁷ is gratis. Scholarcy⁸ is betalend.

Generative Pretrained Transformer 3 of GPT-3 is een taalmodel ontworpen door OpenAI. Dit taalmodel gebruikt een tweestapsleerparadigma waarbij het eerst ongesuperviseerd wordt getraind met een taalmodelleringsdoel en daarna gesuperviseerd wordt gefinetuned. Over drie versies heen is het model aanzienlijk vergroot, van anderhalf miljard parameters bij GPT-2 naar 175 miljard parameters bij GPT-3. Het model is getraind op niet-gecategoriseerde data van het internet en gebruikt datasets waaronder Common Crawl, WebText2, Books1, Books2, and Wikipedia (C. Li, 2022; Radford e.a., 2019).

⁷<https://typeset.io/>

⁸<https://www.scholarcy.com/?ref=theresanaiforthat>



Can you simplify the following paragraph:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.



Here is a simpler version of the paragraph:

When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.

Figuur (2.9)

Afbeelding van Gooding 2022. De invoertekst is een paragraaf uit een niet-vermeld boek van de Russische schrijver Dostoevsky. Het resultaat van de meegegeven prompt is een transformatie dat iedere vorm van vooraf aangehaalde vereenvoudiging weergeeft. Lexicale, conceptuele en syntactische vereenvoudiging worden op de invoertekst toegepast.



Can you simplify the following paragraph in three different versions:

At some ideas you stand perplexed, especially at the sight of sin. And you must ask yourself whether to combat it with force or with humble love. Always decide to combat it with humble love. If you decide this once and for all, you can conquer the whole world! Humble love is a frightening force, of which there is nothing stronger.

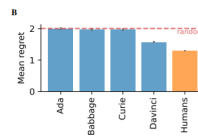


Here are three simplified versions of the paragraph:

1. When you see something bad, like sin, you might be confused. Ask yourself if you should use force or love to fight it. Always choose love. If you do, you can win against anything. Love is the strongest power.
2. When you encounter ideas that confuse you, especially when it comes to sin, you may be unsure how to react. You must ask yourself if it is better to fight back with force or with humble love. It is always better to choose humble love. If you make this decision once and for all, you can overcome anything. Humble love is a powerful and intimidating force that is stronger than anything else.
3. If you see something bad, like sin, and you don't know what to do, remember that love is always the answer. Choose love over force. If you do, you will be able to conquer anything. Humble love is the most powerful force in the world.

Figuur (2.10)

Afbeelding van Gooding 2022. Gooding haalt verder aan dat modellen zoals ChatGPT op twee vlakken de leesbaarheid van een tekst kan bevorderen. Allereerst door het verlenen van verschillende mogelijke versies van een vereenvoudigingstaak.

**Figuur (2.12)**

Afbeelding van Binz en Schulz (2023). Dit toont de *mean regret* aan tussen de vier engines en de menselijke antwoorden.



Which of the three simplified versions is the simplest?



Of the three simplified versions, the first one is the simplest. It uses the shortest and simplest words and phrases to convey the main idea of the original paragraph, which is that humble love is the strongest force and can be used to overcome anything. The other two versions add more details and examples, which make them slightly more complex than the first version. However, all three versions are simpler than the original paragraph and effectively convey the same idea.

Figuur (2.11)

Afbeelding van Gooding 2022.

Lisowski (2023) vergelijkt de twee OpenAI taalmodellen met een *mixed-methods* onderzoek. Al blijken de twee heel gelijkaardig, het experiment benadrukt dat het ChatGPT-model gericht is op conversationele doeleinden met voorkeur als chatbot, terwijl GPT-3 een ML-model is bedoeld om met hoogstens één prompt te werken. De grootte van het GPT-3 model met 175 miljard parameters imposanter dan ChatGPT. Daarnaast is de limiet bij het meest recente GPT-3 model is 4000 tokens. Verder haalt Lisowski aan dat de kwaliteit bij beide modellen sterk afhankelijk is van de invoer. De prompts moeten concreet genoeg zijn, om zo niet af te wijken van wat de gebruiker wilt (Lisowski, 2023). Deze twee API's zijn nu vrij beschikbaar voor ontwikkelaars als betalende API (Greg e.a., 2023).

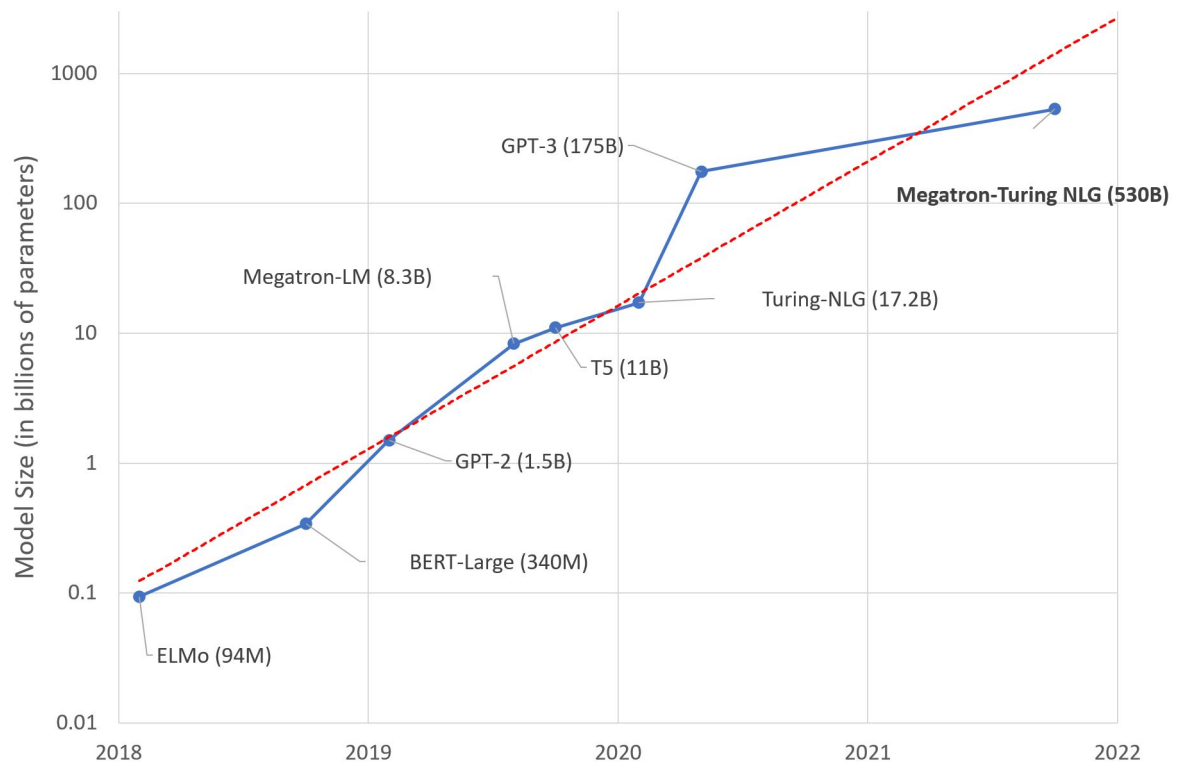
De documentatie van OpenAI⁹ reikt vier verschillende engines voor het GPT-3 taalmodel aan, namelijk Davinci, Curie, Babbage en Ada. In Maart 2023 voegde een vijfde engine zich toe, namelijk GPT-3 Turbo wat de basis is achter Chat-GPT. Davinci-003 is het meest geavanceerde model dat alles kan wat de andere engines ook kunnen, met de meest menselijke antwoorden en geschikt voor taken zoals essays schrijven en code genereren. Curie is goed voor nuance maar minder menselijk dan Davinci, terwijl Ada en Babbage minder krachtig zijn en aangeraden worden voor eenvoudige taken zoals tekst aanvullen en sentiment analyse (Greg e.a., 2023).

De mogelijkheden van OpenAI's ChatGPT en GPT-3 modellen zijn nog volop in ont-

⁹<https://platform.openai.com/docs/>

wikkeling, maar er zijn al enkele vergelijkende onderzoeken uitgevoerd. Uit het experiment van Tanya Goyal (2022) blijkt dat *zero-shot* samenvattingen met GPT-3 beter presteren dan *fine-tuned* modellen. Daarnaast haalt Mottes (2023) verschillende tools aan die gebruik maken van de GPT-3 API, waaronder Jasper AI en ChatSonic. Ook voor het onderwijs zijn er mogelijkheden, zoals de hoge toegankelijkheid en granulaire personalisatie van het GPT-3 model (Garg, 2022; Roose, 2023). Echter, GPT-3 is niet geschikt voor alle taken, zoals sentimentanalyse en -classificatie, waarvoor een kleinschaliger taalmodel beter presteert (C. Li, 2022). Bovendien is er aandacht voor de ecologische effecten van de grote omvang van deze modellen, waarvoor alternatieve oplossingen zoals het gebruik van Cloud-infrastructuur en geschikte model finetuning worden voorgesteld (Simon, 2021; Strubell e.a., 2019).

De architectuur tussen GPT-3 en BERT is volgens Mottes (2023) het meest opvallende verschil. GPT-3 is een autoregressief model en houdt daarmee enkel rekening met de linkercontext bij het voorspellen of genereren van tekst. BERT daarentegen is bidirectioneel en neemt zowel de linker- als de rechtercontext in overweging. De bidirectionele werking is geschikt voor sentimentanalyse waarbij begrip van de volledige zincontext noodzakelijk is. GPT-3 heeft toegang tot meer informatie (45TB) dan BERT (3TB), wat het een voordeel kan geven bij het samenvatten of het vertalen. Ten slotte zijn er ook verschillen in grootte. Hoewel beide modellen erg groot zijn, is GPT-3 aanzienlijk groter dan de voorganger vanwege de uitgebreide trainingsdatasetgrootte (Brown e.a., 2020). LLaMa of Large Language Model Meta AI is een generatief taalmodel met potentieel dat sterker is dan GPT-3 en soortgelijke modellen, terwijl het van tien keer minder parameters gebruik maakt, maar is nog niet beschikbaar als online webtoepassing of API (Hern, 2023; Touvron e.a., 2023).

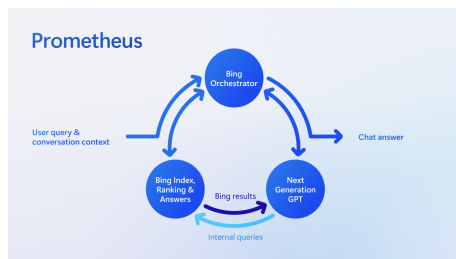
**Figuur (2.13)**

Afbeelding van Simon (2021). De evolutie van pre-trained taalmodellen wordt hier weergegeven tot eind 2022. De performantie van de modellen ten opzichte van de grootte volgt een lineaire functie.

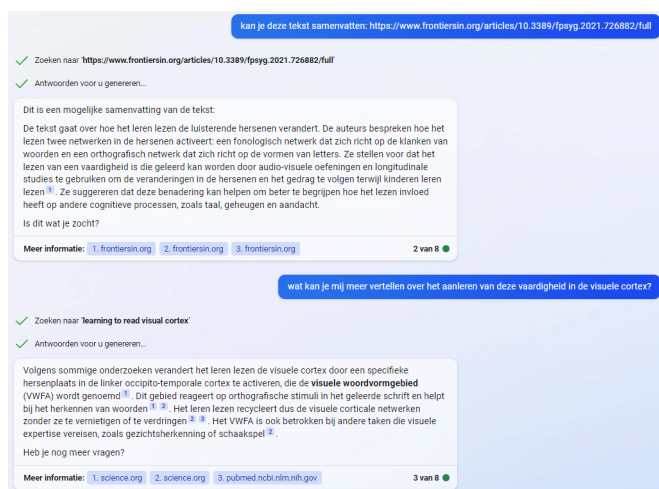
Parameter	Omschrijving	Mogelijke waarden
model	Het GPT-3 model om te gebruiken	davinci, curie, babbage, ada, text-davinci-002, text-curie-001, text-babbage-001, text-ada-001, davinci-codex
temperature	De gulzigheid van een generatief model. Een lagere waarde zal conservatieve en voorspelbare tekst teruggeven. Hogere waarden zullen meer gevarieerde en onverwachtse tekst teruggeven, wat beter werkt bij creatieve toepassingen.	Een kommagetal tussen 0 en 1.
max_tokens	Het maximaal aantal tokens (woorden of subwoorden) dat het generatief model kan teruggeven.	Een getal tussen 1 and 2048.
top_p	Vergelijkbaar met temperature, maar deze waarde onderhoudt de probability distribution voor common tokens. Hoe lager de waarde, hoe waarschijnlijker de woordenschat dat het model zal overwegen bij het genereren van tekst. Een hoge waarde is toepasselijker wanneer een toepassing gericht is op nauwkeurigheid en correctheid.	Een kommagetal tussen 0 en 1.
stop	Een tekstwaarde (woord/symbool) tot waar het model zal genereren. When the model generates a string that matches any of the specified strings, it stops generating text.	Een lijst van stringwaarden, of een enkele string.
presence_penalty	Factor die bepaalt hoe regelmatig woorden voorkomen.	Een kommagetal tussen 0 en 1

Microsoft en OpenAI werken nauw samen. Zo maakt het conversationele taalmodel van Bing ook gebruik van GPT-3. Deze chatbot bouwt verder en biedt zo verwijzingen en referenties aan naar andere websites. Deze verwijzingen zijn volgens mogelijk door de Prometheus-technologie van Microsoft (Ribas, 2023). Prometheus is een eigen technologie die door Bing is ontwikkeld. Het AI-model is volgens Ribas (2023) de eerste van zijn soort die de Bing-index-, ranking- en antwoordresultaten

combineert met het redeneervermogen van OpenAI's GPT-modellen. Prometheus maakt gebruik van de kracht van Bing en GPT om iteratief via een component genaamd *Bing Orchestrator* een set interne queries te genereren met als doel binnen gegeven gesprekscontext een nauwkeurig antwoord op gebruikersqueries te bieden (Ribas, 2023). Bing AI is nu in testfase met wachtlijst en bestaat in de vorm van een webpagina en een browserextensie voor Microsoft Edge. Onderzoek naar deze chatbot staat nog in de kinderschoenen en er is nood aan onderzoek naar de credibiliteit en correctheid van de verwijzingen. Deze chatbot gebruikt een combinatie van extraherende en abstraherende samenvattingen. In tegenstelling tot GPT-3 is er geen officiële API beschikbaar. Daarnaast is de limiet ook lager met 2000 tokens per bericht tijdens een conversatie.



Figuur (2.14)
Afbeelding van Ribas (2023).



Figuur (2.15)
In deze afbeelding wordt er een online wetenschappelijk artikel meegegeven. Er wordt geen titel of onderwerp meegegeven, maar de Bing AI chatbot is in staat om een abstraherende samenvatting te maken van het artikel. Daarna geeft de chatbot verder uitleg over een bepaald onderwerp en geeft het extra referenties mee.

In recente literatuur is Huggingface beschreven als een platform of portaalsite voor het delen van ML-modellen en datasets. De bibliotheek biedt een scala aan API's en tools die gemakkelijk te downloaden en trainen zijn voor pretrained modellen

voor prevalentie NLP-taken, zoals tekstclassificatie, taalmodellering en samenvatting. Deze modellen kunnen worden gefinetuned op specifieke datasets, waardoor ontwikkelaars snel modellen kunnen bouwen en inzetten voor vereenvoudigings- en samenvattingstaken. Voor wetenschappelijke documenten en artikelen bestaan er al enkele modellen en datasets: ^{10, 11}

2.6.1. Conclusie

Experten halen het GPT-3 model en ChatGPT aan als de toekomst voor gepersonaliseerde en adaptieve uitleg aan scholieren. Bing AI biedt een extra dat revolutionair kan zijn bij het opzoeken van uitleg voor zoektermen, zonder het verlies aan bronvermelding. Huidige toepassingen staan mogelijks in een spreekwoordelijke schaduw eenmaal leessoftware voor scholieren met dyslexie worden ontwikkeld met AI. De mogelijkheden van GPT-3 zijn eindeloos en toepassingen die hiervan gebruik maken, kunnen in het onderwijs ingezet worden als ondersteunende software.

2.7. Conclusie

De noden van scholieren met fonologische dyslexie in de derde graad van het middelbaar gaan verder dan gewoon moeizaam lezen. Het ontcijferen en automatiseren van wordeherkenning gebeurt langzaam. Er zijn bewezen voordelen van MTV en aangepaste visuele weergaven op kinderen en jongeren met dyslexie. De leesbaarheid van wetenschappelijke artikelen bevindt zich in een dalende trend. Het formaat, gebruik van vakjargon en ingewikkelde woordenschat en ten slotte de moeizame syntax en zinsbouw sluiten een algemene doelgroep uit bij het lezen van wetenschappelijke artikelen. Enkel wetenschappelijk geletterden zijn in staat om deze artikelen te lezen. Het uniforme formaat van een wetenschappelijk artikel biedt kansen aan voor een geautomatiseerde aanpak tot het vereenvoudigen van een tekst.

Experten halen meerdere bewezen tactieken aan om teksten automatisch te vereenvoudigen op maat voor een scholier met dyslexie. Handmatig worden teksten vereenvoudigd aan de hand van leesbaarheidsformules of intuïtie. Zinnen moeten lexicaal, syntactisch en semantisch worden vereenvoudigd. Teksten samenvatten maakt de tekst korter zonder het verlies van de kernboodschap. Voor deze vier transformaties zijn er taalmodellen beschikbaar in de vorm van API's of open-source software. Huidige software dat de overheid uitleent aan scholieren met dyslexie in het middelbaar onderwijs fungeert voornamelijk als voorleessoftware. Nieuwe en opkomende technologieën en taalmodellen zoals GPT-3 blinken uit om tekstvereenvoudiging mogelijk te maken. De ontwikkeling met LLM's is in opmars,

¹⁰<https://huggingface.co/sambydl0/bart-large-scientific-lay-summarisation>

¹¹https://huggingface.co/haining/scientific_abstract_simplification

maar ontwikkelaars moeten bewust zijn dat andere taalmodellen zoals BERT voor taken zoals semantische analyse minder rekenkracht vereisen voor eenzelfde en soms beter resultaat.

3

Methodologie

Met de verworven kennis uit de literatuurstudie in Hoofdstuk 2 over de noden van scholieren met dyslexie, het complexe karakter van wetenschappelijke artikelen en ten slotte de technieken voor ATV en MTV kunnen de onderzoeksmethoden toegepast worden om een antwoord te kunnen vormen op de onderzoeksvraag.

3.1. Requirementsanalyse

Deze onderzoeksfase doelt op het uittesten van verschillende tools op hun functionaliteiten voor geautomatiseerde en gepersonaliseerde tekstvereenvoudiging. De requirementsanalyse biedt een antwoord op de volgende twee subvragen, aan de hand van de technieken voor tekstvereenvoudiging beschreven in de literatuurstudie 2.

- Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs?
- Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?

De requirementsanalyse analyseert toepassingen die enerzijds de overheid aan scholieren met dyslexie in het onderwijs aanbiedt, anderzijds online tools die leerkrachten kunnen gebruiken om teksten te vereenvoudigen. Om de lijst van beschikbare tools in te perken, beperkt het onderzoek zich op de volgende tools:

- Simplish
- Resoomer
- ChatGPT

- Bing Chatbot

Vijf verschillende aspecten, namelijk lexicale en syntactische vereenvoudiging, samenvatting, personalisatie en ontwikkelaarsmogelijkheden dienen als de houvast bij deze requirementsanalyse. De resultaten van deze onderzoeksfase dienen als vereisten om een prototype voor tekstvereenvoudiging te ontwikkelen, specifiek op maat om wetenschappelijke artikelen gepersonaliseerd te vereenvoudigen op maat voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs. De tools ondergaan een test om hun vermogen om een wetenschappelijk artikel in PDF-formaat te vereenvoudigen, indien dat mogelijk is. Als dit niet mogelijk is, gebeurt het kopiëren en plakken van de tekst in de toepassing. Op deze manier vindt een evaluatie van de prestaties van de verschillende tools plaats en worden eventuele beperkingen of tekortkomingen geïdentificeerd.

3.2. Vergelijkende studie

Taalmodellen zijn nodig om teksten te kunnen vereenvoudigen, maar een toepassing voor tekstvereenvoudiging binnen de casus van wetenschappelijke artikelen moet gebruik maken van een taalmodel dat het meest aansluit op deze casus. Om het prototype af te stemmen op één taalmodel, vereist een antwoord op de volgende vraag.

- Welk taalmodel of LLM is ideaal voor automatische tekstvereenvoudiging voor vereenvoudigde wetenschappelijke artikelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs met dezelfde of gelijkaardige kwaliteiten als manuele tekstvereenvoudiging?

De literatuurstudie identificeerde de volgende vier taalmodellen voor opname in de vergelijkende studie.

- Een gefinetuned versie van T5¹.
- Een gefinetuned versie van BART², getraind op woordenschat van Engelstalige wetenschappelijke artikelen.
- Scientific Abstract Simplification³
- GPT-3

Zoals aangegeven in het onderzoek van Nenkova en Passonneau (2004) worden leesgraadsformules en referentieteksten ingezet om de kwaliteit van een vereenvoudigde tekst te beoordelen. De taalmodellen en API's ondergaan een vergelijking met elkaar en met een referentietekst om de kwaliteit van de gegenereerde

¹<https://huggingface.co/husseinMoh/t5-small-finetuned-text-simplification>

²<https://huggingface.co/sambydl0/bart-large-scientific-lay-summarisation>

³https://huggingface.co/haining/scientific_abstract_simplification

tekstvereenvoudigingen te beoordelen. De vergelijking gebeurt zowel subjectief als objectief aan de hand van de volgende leesgraadsmetrieken:

- Flesch-Reading-Ease
- Gunning FOG

Deze evaluatie leidt tot de keuze van een taalmodel dat het meest geschikt is voor deze specifieke casus van wetenschappelijke artikelen, die gepersonaliseerd en geautomatiseerd moeten worden vereenvoudigd voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Het resultaat van deze onderzoeksfase is één taalmodel wat zal dienen als de basis en als methode om gekregen wetenschappelijke artikelen te vereenvoudigen op maat voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs.

3.2.1. Vergelijking met referentieteksten

Dit onderzoek vergelijkt twee wetenschappelijke artikelen met een referentietekst die handmatig is vereenvoudigd door twee lectoren en twee scholieren. Deze vier personen baseren zich op vooraf meegekregen richtlijnen, toegelicht in bijlage ...

GPT-3 en de HuggingFace taalmodellen kunnen enkel per API worden aangesproken en zijn daarmee niet in staat om direct PDF's te opladen. De inhoud wordt met de code op (TODO REF TOEVOEGEN) opgehaald en vervolgens direct aan het model gegeven om zo een resultaat te bekomen. Bij het gebruik van het GPT-3 model worden vijf prompts gebruikt en om rekening te houden met de tokenlengte, wordt de tekst tussen de subtitels opgebroken. De prompts zijn zoals volgt opgesteld:

- P1: Simplify this text in Dutch
- P2: Simplify a text for economics students (16-18 years old) by replacing difficult words, keeping technical jargon, replacing words longer than 18 letters, writing acronyms in full, replacing a word only once with a synonym, giving brief explanations when necessary, and avoiding percentages.
- P3: Simplify a text by breaking them into shorter sentences with a maximum of ten words. Change pronouns like 'they', 'their', or 'he' to names. Replace complex sentence constructions and prepositional phrases with simpler alternatives, but leave them unchanged if no simpler option is available.
- P4: Simplify an article for economics students (16-18 years old) by replacing difficult words (except technical jargon learned in the 2nd grade), words longer than 18 letters, and avoiding percentages. Use synonyms only once and give brief explanations if necessary. Write acronyms in full and replace pronouns

with names. Simplify sentence constructions and prepositional phrases, splitting sentences into a maximum of ten words, but leave them unchanged if no simpler alternative is available.

3.3. Prototype voor tekstvereenvoudiging

Dit hoofdstuk omschrijft de ontwikkeling van een prototype voor tekstvereenvoudiging voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Deze onderzoeksmethode geeft een antwoord op de volgende deelvraag:

- Hoe kan een intuïtieve en lokale webtoepassing worden ontwikkeld die zowel scholieren met dyslexie als docenten helpt bij het vereenvoudigen van wetenschappelijke artikelen met behoud van semantiek, jargon en zinsstructuren?

Het prototype is ontwikkeld met de benodigde functionaliteiten en eigenschappen uit de requirementsanalyse en wordt enkel lokaal uitgerold. De uitrol van dit prototype gebeurt enkel lokaal door middel van Docker. Python en het Flask-framework dienen als de basis voor deze webtoepassing. Het onderstaande schema haalt de gebruikte programmeertalen aan waaruit dit prototype bestaat.

Technologie	Functionaliteit
-------------	-----------------

Python-bibliotheek	Functionaliteit
PDFMiner	Tekstinhoud van PDF's inlezen.
EasyOCR	Pagina van een PDF opslaan als afbeelding. Tekst uit een afbeelding extraheren.
Spacy	PoS-tagging en lemmatizing
NumPy	De reshape-functie vereenvoudigt de manier om arrays van zinnen bij elkaar te plaatsen om zo een paragraaf te bekomen.

Het prototype is opgebouwd in Python met het Flask-framework om de webapplicatie te ondersteunen. Aanvullend op Flask maakt het prototype gebruik van Jinja, HTML, CSS en JavaScript (JS) zoals vermeld in 3.3. Tekstvereenvoudigingsfuncties worden eerst in Jupiter notebooks uitgetest, vooraleer deze in het prototype worden geplaatst. Docker wordt gebruikt als opzet voor de ontwikkelaars. Een scriptbestand vereenvoudigt de opstart van deze webapplicatie in tegenstelling tot de opstart per terminal. De nodige Python-bibliotheken worden alvorens opgehaald met Pipreq. Alle taalmodellen worden per API aangesproken, ofwel één Docker-container voor de webapplicatie volstaat voor dit prototype.

3.3.1. Personalisering van de website

Zoals aangewezen in 2 door (..) moeten ontwikkelaars de personalisering van de website in rekening houden, want deze functionaliteit heeft een significant effect op het leesgedrag en -begrip van zowel mensen met dyslexie als mensen zonder dyslexie. Op een apart scherm kan de eindgebruiker de volgende elementen aanpassen naar hun toebehoren:

- Test

De twee verschillende soorten eindgebruikers voor dit prototype, namelijk leerkrachten die een wetenschappelijk willen vereenvoudigen op maat voor een scholier in hun klasgroep of een scholier die dit zelf wil doen, moeten elk een andere taak verrichten. Het prototype volgt daarmee een opsplitsing qua functies.

3.3.2. Tool voor leerkrachten

1. PDF of tekstinhoud opladen met PDFMiner
2. Alle zinnen uit de opgeladen tekst ophalen met Spacy.
3. Woorden taggen met de PoS-functie van Spacy.
4. Opties voor gepersonaliseerde tekstvereenvoudiging aanreiken aan de eindgebruiker.
5. Tekst vereenvoudigen op basis van gepersonaliseerde keuzeopties.
6. Nieuw artikel genereren met Pandoc.

3.3.3. Tekstinhoud extraheren

Eindgebruikers kunnen met het prototype wetenschappelijke artikelen op één van twee manieren inladen: *plaintext* of via een PDF-bestand. De werking om een wetenschappelijk artikel in te lezen is identiek bij zowel de toepassing voor scholieren als die voor lectoren en daarmee is de code herbruikbaar. De Flask-applicatie controleert vooraf de type invoer om vervolgens de Reader-klasse aan te spreken die de tekst verder verwerkt. PDF's worden tijdelijk in-memory opgeslaan. Er wordt rekening gehouden met de splitsing tussen normale upload en geavanceerde upload. Deze twee methoden zijn terug te vinden in de Reader-klasse⁴.

- PDFMiner itereert doorheen de PDF en extraheert vervolgens de tekst op iedere pagina. Deze methode resulteert in een string-object.
- PDF extractors, waaronder PDFMiner, kunnen tekstinhoud verliezen tijdens het extraheren zoals eerder aangewezen in 3.1. Als vangnet biedt het prototype een tweede optie aan waarbij de PDF-pagina's als afbeelding worden

⁴<https://github.com/Dyashen/text-simplification-tool/blob/main/web-app/Reader.py>

opgeslaan. De Python-bibliotheek EasyOCR voorziet een eenduidige en ontwikkelaarsvriendelijke manier om PDF-pagina's op te slaan als JPG of PNG. Tesseract biedt een even eenduidige oplossing aan, maar vergt meer configuraties en daarmee vergroot de omvang van de Docker-container. Vervolgens worden de afbeeldingen per tekst-chunk ingelezen en de tekst wordt opgeslaan. Het gebruik van deze alternatieve methode kan de omvang van de Docker-container vergroten. Daarom verwijdt het prototype deze afbeeldingen nadat deze de tekstinhoud hebben ingelezen en opgeslaan om ruimte te besparen. Net zoals bij de eerste methode resulteert deze methode in een string-object.

3.3.4. Zinnen ophalen

De eerste fase van het prototype slaat de tekstinhoud op in meerdere arrays die zinnen voorstellen. Door middel van een aparte functie wordt de tekst opgesplitst per zin. Het resultaat van deze transformatie is een tweedimensionale array. Deze transformatie bevoordeelt het proces om vervolgens de teksten per zin op de webpagina uit te printen.

3.3.5. PoS-tagging

De key-value paren dienen om de woorden aan hun respectievelijke PoS-tag te koppelen. De sleutel verwijst naar het woord in de zin en de waarde verwijst naar de PoS-tag die aan dit woord toebehoort. Dit prototype houdt enkel rekening met de PoS-tagging van Nederlandstalige en Engelstalige teksten en daarom laadt het prototype enkel twee embeddingsmodellen op, zoals aangeduid in 3.3.5. Hardcoderen is uit den boze en zo maakt het prototype gebruik van een dictionary die de naam van deze embeddingsmodellen bijhoudt. Het prototype hoeft daarmee enkel de taal te herkennen en die vervolgens door te geven aan de dictionary. Deze methode fouttolerant maken kan door Engels als standaardtaal mee te geven of door vooraf de gebruiker te vragen in welke taal de opgelade tekst staat.

Taal	Embeddingsmodel
Nederlands	NL Core News Medium ⁵
Engels	EN Core Web Medium ⁶

3.3.6. Opties voor gepersonaliseerde tekstvereenvoudiging aanreiken.

Leerkrachten hebben baat aan een intuïtieve toepassing die prompts en API-calls voor hen kan regelen. Het prototype regelt dit door de eindgebruiker een formulier te presenteren met daarin alle opties die leerkrachten kunnen gebruiken om een wetenschappelijk artikel te vereenvoudigen op maat van de unieke noden voor een scholier met dyslexie in de derde graad van het onderwijs.

JavaScript komt hier aan bod als api-key herkenner; zodat de eindgebruiker weet of deze een GPT-3 api-sleutel heeft ingegeven.

3.3.7. Tekstvereenvoudiging met API

Om afwijkende resultaten op een GPT-prompt te vermijden, wordt de temperature op nul geplaatst en de *top_p* waarde wordt ingeschat op 80%. SpaCy wordt gebruikt om woordkenmerken zoals de PoS-tag op te halen, maar het systeem is vatbaar voor het niet kunnen vinden van afwisselende en meertalige woordenschat. Een mogelijke oplossing is om de taal te veranderen naar Engels of Frans, of een aangepast taalherkenningsmodel te gebruiken. Een andere optie is om de tekst voor te verwerken om de Nederlandse en Engelse woorden te scheiden voordat ze worden verwerkt met SpaCy. Adjectieven uit de tekst verwijderen is mogelijk zonder taalmodel. Aangezien alle woorden gekoppeld worden aan een PoS-tag, is het eenvoudig om de woorden gelinkt aan de span-tag van de adjectieven uit te filteren.

Annotaties van woordenschat

De eenduidige HTML-structuur van online woordenboeken maken het mogelijk om gratis en eenvoudig de definities van woorden op te halen. Zo is het mogelijk om annotaties op te halen zoals aangewezen in het onderzoek van Bulté e.a. (2018). Met behulp van Requests en BeautifulSoup is het mogelijk om lijsten met definities te scrapen van deze sites. De stam van het gemarkeerde woord wordt opgehaald en vervolgens meegegeven als zoekopdracht. De bron wordt samen met het resultaat aan de eindgebruiker getoond.

Tekstvereenvoudiging

Het prototype gebruikt een taalmodel van *HuggingFace* voor extraherende samenvattingen en zowel gratis taalmodellen van *HuggingFace* als het GPT-3 taalmodel voor abstraherende samenvattingen. Het model kan parameters, zoals maximale lengte van de gegenereerde tekst, ontvangen en biedt zowel gepersonaliseerde als niet-gepersonaliseerde vereenvoudiging. Het gebruik van *HuggingFace* vereist een internetverbinding en kan geen extra trainingsdata bevatten. De opstarttijd voor alle *HuggingFace*-taalmodellen wordt bij de start van de applicatie afgehandeld door middel van een extra parameter de request. Sleutels worden standaard bijgehouden in env-bestanden. Via de webtoepassing kan een gebruiker deze sleutel aanpassen. Binnen een lokale omgeving is dit in orde, al moeten ontwikkelaars rekening houden met beveiligingsmaatregelen wanneer een dergelijke tool wordt uitgerold. Het merendeel van de gebruikte taalmodellen is Engelstalig of is nadrukkelijk getraind op basis van Engelstalige datasets. De ingegeven tekst wordt eerst vertaald naar het Engels om zo de kans op een accurate vereenvoudiging te verhogen. Voor de vertaling wordt de Google Translate Python-package gebruikt. Deze is minder accuraat vergeleken met DeepL, maar biedt een gratis beschikbaar

en aanvaardbaar alternatief aan. Factoren zoals topic diversity en semantische redundantie moeten overwogen worden bij het kiezen van een taalmodel voor extraherend samenvatten. Lange documenten samenvatten kan zoals aangeduid in literatuurstudie door extraherende samenvatting, gevolgd door abstraherende samenvatting om de tekst coherent te doen blijken. Eerder werd er gekozen om de voltekst per paragraaf bij te houden. Uit iedere paragraaf wordt een ideaal aantal zinnen gemarkeerd om nadien geparafraseerd te worden door GPT-3 of een *HuggingFace* taalmodel, afhankelijk van de keuze van de eindgebruiker.

3.3.8. Personaliseerbaarheid aanreiken

Voor de webtoepassing worden de standaardparameters gebruikt die uitgewezen zijn in Rello en Baeza-Yates (2013) en Rello, Baeza-Yates, Dempere-Marco e.a. (2013). Met JavaScript is het mogelijk om deze parameters dynamisch en on-the-spot aan te passen. Deze gekozen parameters worden opgeslaan als sessievariabelen, zodat de eindgebruiker niet per pagina deze parameters moet instellen. Om het uitvoerbestand personaliseerbaar te maken, worden opties in een formulier opgevraagd en vervolgens meegegeven in de Pandoc YAML-header.

Taalmodellen

SC en BART-SC transformeren de tekst op lexicaal en syntactisch niveau. Zij bekijken enkel de gekregen zin. Andere taalmodellen zijn eerder geneigd om extra tekst toe te voegen. Er kan niet achterhaald worden waarom dat deze extra tekst wordt meegegeven. BART-SC kan bijzaak behouden, terwijl SC sneller de neiging heeft om enkel de kernzaak te behouden in de vereenvoudigde tekst. Bij de inference API's moet er expliciet worden aangegeven om welke transformatie dit gaat door kernwoorden zoals 'summarize:'.

3.3.9. Tekstinhoud uitschrijven naar PDF/DOCX

De zelfgebouwde Creator-klasse bouwt PDF's en docx-documenten op volgens de meegegeven personalisatie. Het prototype maakt gebruik van Pandoc, of PyPandoc via Python, om tekstinhoud naar een PDF of een docx-bestand uit te schrijven. Pandoc maakt gebruik van een tweestapsbeweging waarbij rauwe tekst eerst naar een Markdown-formaat wordt omgezet.

YAML-header

Met Python wordt eerst een YAML-header in het te-transformeren Markdown-bestand geschreven. De YAML-header omvat de titel, standaardlettertype en lettertype voor de titel, de datum, het type document dat moet worden gegenereerd, de marge-instelling, de standaardlettergrootte, woord-spatieëring en ten slotte de instelling voor de regeleinde. De meegekregen gepersonaliseerde instellingen wordt meegegeven in een LaTeX YAML-header.

Woordenlijst en vereenvoudigde tekst uitschrijven

De structuur om de woordenlijst op te bouwen, is identiek zoals dat van een Markdown-tabel. De woordenlijst wordt in dictionary-structuur meegegeven. De sleutels worden overlopen en vervolgens wordt ieder woord samen met de PoS-tag en de definitie uitgeprint. De vereenvoudigde tekst is eveneens in een dictionary-structuur opgeslaan. De keys stellen titels voor en worden uitgeprint voorafgegaan door twee hekje-symbolen. Tussen de titels worden breaklines toegevoegd, gevolgd door de tekst die bij de titel bijhoort. Indien gekozen werd voor een opsomming, dan wordt er gebruik gemaakt van een geneste for-lus waarbij iedere zin wordt voorafgegaan aan een asterisk-symbool. De woordenlijst en vereenvoudigde tekst worden naar hetzelfde Markdown-bestand uitgeschreven.

Documenten genereren

Als invoer wordt het pad naar opgevulde Markdown-bestand meegegeven. De uitvoer is het pad waarnaar het PDF- of DOCX-bestand moet worden opgeslaan. Vervolgens zet Pandoc het Markdown-bestand om naar een PDF-bestand gebouwd met de XeLaTeX engine of een Word-bestand op basis van meegekregen binaries. Pandoc Flask kan enkel één bestand aan de gebruiker teruggeven. Als oplossing comprimeert het prototype met *zipfile* de PDF- en Wordbestand tot één bestand.

3.3.10. Conclusie

Flowchart voor ontwikkeling 1. Requirementsanalyse + opzet moscow. 2. Te gebruiken taalmodellen achterhalen. 3. Functies schrijven in Jupyter notebooks of soortgelijke scripts.

Dit prototype wordt enkel binnen een lokale omgeving opgezet en is nog niet bruikbaar voor het grote publiek. Met PDF's of voltekst als invoer is het prototype in staat om teksten lexicaal en syntactisch te vereenvoudigen. Het prototype is functioneel voor zowel de doelgroep lectoren als leerlingen, twee doelgroepen die elk een andere functionaliteit prioriteren.

Het prototype gebruikt API's waaronder de *HuggingFace* Inference APIs en de GPT-3 API. Aanvullend hierop kunnen ontwikkelaars deze modellen extra trainen op basis van de gewenste casus.

Ontwikkelaars moeten rekening houden met het gebrek aan structuur bij het ophalen van tekstinhoud uit een PDF-bestand.

4

Discussie

In dit hoofdstuk worden de resultaten uit de requirementsanalyse, vergelijkende studie en de ontwikkeling van het prototype besproken.

4.1. Requirementsanalyse

De geteste toepassingen representeren de state-of-the-art toepassingen voor tekstvereenvoudiging. Proof-of-concepts Webtools presenteren op hun sites verschillende functionaliteiten die aan de eindgebruiker worden uitgeleend. Deze zijn terug te vinden bij het uittesten en er is geen sprake van false advertising.

Woorden- en synoniemenlijsten kunnen een ondersteunend middel aanbieden voor zowel scholieren met dyslexie als zonder bij het lezen van wetenschappelijke artikelen en wordt aangeboden in Kurzweil. Automatisch genereren is enkel prevalent binnen ChatGPT en de Bing chatbot, maar de tools houden geen rekening met de doelgroep, tenzij expliciet aangegeven met een one-shot summary. Andere tools houden helemaal geen rekening met de doelgroep en kunnen enkel woordenlijsten genereren op basis van gekozen woorden.

PDF-upload steekt uit als de must-have optie om wetenschappelijke artikelen op te laden, al beschikken de geteste chatbots zoals ChatGPT en Bing Chat hier niet over. Tekstinhoud inlezen komt niet voor bij de erkende software in het onderwijs, in tegenstelling tot de webtoepassingen die hier wel over beschikken. Voornamelijk komt dit doordat de tools als snelle optie tot vereenvoudiging of samenvatting kunnen dienen, terwijl het opladen van een PDF meer tijd in beslag kan nemen. ChatGPT en Bing chatbot genereren teksten met een gelijkaardige schrijfstijl als die van de mens, maar deze toepassingen kunnen geen PDF's verwerken. Het kopiëren en plakken van tekst uit het originele document kan leiden tot weinig fouten, maar komt deze manier van werken is omslachtig en moet verbeterd worden. Be-

staande tools hebben moeite met oudere PDF's waarbij niet alle tekst kan worden geëxtraheerd en daarom moet het prototype een vangnet voorzien om dit mogelijk scenario tegen te gaan.

Tools bieden weinig inzicht op leesgraadsmetrieken van zowel het ingegeven document als het vereenvoudigde document. Simplish steekt boven de rest uit door aan de hand van kleurcodes criteria over de vereenvoudigde tekst mee te geven, waaronder niet-veranderende woorden, adequate vertalingen, uitleg naar de voetnoot, homoniemen of woorden die geen eenvoudigere synoniemen hebben. Zoals aangegeven in ?? duidt de vergelijkende weergave de verschillen aan tussen de oorspronkelijke en vereenvoudigde tekst en met behulp van kleurcodes worden de verschillende transformaties aangegeven.

Color code	
Black	Words in Black don't change between the two versions.
Green	Words in Green mean they have been translated adequately.
Purple	Words in Purple display a further explanation in foot notes.
Olive	Words in Olive contain two or more possible meanings (a tooltip is provided for these words, place the mouse cursor on top of olive words to see possible meanings).
Blue	Words in Blue are recognized in Wikipedia (normally names, places, people, organizations, etc.).
Orange	Words in Orange are not currently available in Basic English.
Red	Words in Red are names, special terms or not recognized by the translating tool.

Note : Double click on any word to add it to your personal dictionary.

Input Text

Artificial intelligence has been applied more into occupations by companies and individuals. However, the effects within the benefits are both imaginable and unpredictable. Sexual discrimination in jobs is also a debatable topic. The purpose of this paper is to combine the topics of both AI and sexual discrimination and discuss their effects in the job field in the future. Automation, big data and the algorithm applied in the job field would be some of the points to discover. To briefly summarize, automation is the use of machines and computers that reduces human intervention. Big data is a collection of data from various sources, it is related to AI because the more data input into AI the better it becomes. Since AI absorbs the information and learns from them. AI algorithm takes the data input and uses mathematics and logic to produce the output. [1] Gender discrimination in AI not only reflects the pre-existing biases in the society, but it could also reinforce them through automation, hiring system and decision making. This paper is not totally against the use of AI but advocates that artificial intelligence should be used in a more careful, gender responsible way to reduce sexual discrimination in the job field.

Simplified

artificial intelligence ¹ has been made a request more into work by companies and beings. however, the effects within the gets help are both idea forming and not able to say before hand, sex caused decision making in regular work is also an about which argument is possible thing talked of. The purpose of this paper is to trading group the interests of both AI and sex caused decision making and have a discussion about their effects in the regular work field in the future. automation ², greatly sized facts and the algorithm ³ applied in the regular work field would be some of the points to discover. To briefly give a short account of, automation ² is the use of machines and knowledge processing machines that gets changed to other form to do with man coming between groups. Big facts is a group of facts from different starting points, it is related to AI because the more facts input into AI the better it becomes. Since AI takes up the news given and learns from them. AI algorithm ³ takes the facts input and uses mathematics and tests, reasoning to produce the out put. [1] sex statement decision making in AI not only gives back (light, heat, sound) the in existence beforehand has a tendency in a certain direction in the society, but it could also make stronger them through automation ², getting use of person for money system and decision making. This paper is not totally against the use of AI but Advocates ⁴ that artificial intelligence ¹ should be used in a more careful, sex statement responsible way to get changed to other form sex caused decision making in the regular work field.

Menu ▾

artificial intelli... science that gives great weight to ways of making come into existence intelligent machines that work and have reactions like those of man. [Continue reading.](#)

automation² the technology of making machines, instruments, process, and the like go through a certain train of operations without further impulse or control from outside after being started. [Continue reading.](#)

algorithm³ a word coming from the name of the expert in mathematics /Al-Khwarizmi@who (780-850ac), used to give the way to work out or solve points to be answered. [Continue reading.](#)

Advocates⁴ A barrister or solicitor representing a party in a hearing before a Court. [Continue reading.](#)

MoSCoW-principe	Functionaliteit
Must-have	<p>Gepersonaliseerde vereenvoudiging aanbieden, waaronder lexicale en syntactische vereenvoudiging aanbieden, na het toevoegen van een respectievelijke API-sleutel.</p> <p>Wetenschappelijke artikelen in PDF-vorm opladen.</p> <p>Personaliseerbare site: lettertype -en grootte aanpassen, tekst-formaat aanpassen, achtergrondkleur aanpassen</p> <p>Lokale opzet</p>
Should-have	<p>Glossary genereren na handmatige selectie van moeilijke woorden</p> <p>Personaliseerbare PDF- of Worddocumentlay-out</p> <p>Uitvoer als PDF of Word-bestand teruggeven.</p> <p>Tekstanalyse voor en na de vereenvoudiging aanbieden.</p>
Could-have	Glossary genereren na automatische selectie van moeilijke woorden
Wont-have	<p>Beschikbaarheid tot de tool zonder Docker Desktop, in de vorm van online webtoepassing of browserextensie.</p> <p>Beschikbaarheid tot de standaard- en gepersonaliseerde opties zonder API-sleutels</p>

De requirementsanalyse neemt geen functionaliteiten op van obscure proof-of-concepten. Deze ontbreken de nodige testen en zekerheid dat deze effectieve gepersonaliseerde en geautomatiseerde tekstvereenvoudiging aanbieden. Twee van de online tools, namelijk Resoomer en Scispace, bieden uitsluitend samenvattingsfunctionaliteiten aan. Resoomer is sneller geneigd om de belangrijkste zinnen te markeren en vervolgens deze in een kortere tekst terug te geven. Dit kan een effect hebben op de betrouwbaarheid van de functionaliteiten, alsook op de leesgraadmetriecken.

4.2. Vergelijkende studie

De huidige softwaretools die worden gebruikt in het middelbaar onderwijs zijn niet in staat om de oorspronkelijke tekst te transformeren, wat betekent dat syntactische vereenvoudiging momenteel niet haalbaar is. Hoewel er online webtoepassingen beschikbaar zijn, bieden ze minder functionaliteiten om de moeilijkheidsgraad van zinsyntaxis te verlagen en zijn ze voornamelijk gericht op het verkorten van de oorspronkelijke tekst ofwel samenvatting. Het aanpassen van tangconstructies, verwijswaarden, voorzetseluitdrukkingen, samengestelde werkwoorden en onregelmatige werkwoorden blijft daarom een uitdaging voor deze toepassingen. Zelfs het schrijven in de actieve stem kan problematisch zijn, en er zijn alleen vooraf ge-

definieerde prompts beschikbaar om deze transformaties uit te voeren.

HuggingFace-taalmodellen zijn zoals verwezen in de bijlage, minder in staat om syntactische vereenvoudiging op een tekst toe te passen. De twee uitgeteste taalmodellen zijn voornamelijk gebouwd om lexicale vereenvoudiging te realiseren. Hoewel taalmodellen zoals T4 in staat zijn om zinsyntaxtransformaties uit te voeren, kunnen ze problemen ondervinden bij het verwerken van alle meegegeven transformaties. Het taalmodel kan transformaties ontbreken als alle transformaties in één prompt zijn betrokken. De voorgestelde aanpak hiervoor bestaat uit een pipeline van minstens drie transformaties ofwel drie prompts.

Ontwikkelaars kunnen voor lexicale vereenvoudigingstaken gebruik maken van algemene taalmodellen die vrij beschikbaar staan op *HuggingFace*. Deze taalmodellen komen te kort bij gepersonaliseerde tekstvereenvoudigingstaken en daarom kunnen ontwikkelaars GPT-3 inschakelen als taalmodel voor gepersonaliseerde tekstvereenvoudiging van wetenschappelijke artikelen. Het uitgeteste GPT-3-model blinkt uit in gepersonaliseerde vereenvoudigingstaken. Engelstalige prompts die expliciet de uitvoertaal vermelden zijn nauwkeuriger dan Nederlandstalige prompts. Geen enkel taalmodel kan gegarandeerd rekening houden met de doelgroep. GPT-3 komt dichtbij, maar maakt enkele inschattingfouten.

4.3. Opbouw van het prototype

Deze ontwerpkeuze bespaart geheugenruimte voor ontwikkelaars en vermindert de benodigde rekenkracht voor een prototype. Eenmaal ontwikkelaars de toepassing willen uitrollen naar het grote publiek, wordt er net zoals bij (...) aangeraden om de taalmodellen zelf te hosten.

Intuïtieve handelingen kunnen vrij snel en eenduidig worden ontwikkeld door middel van HTML en JavaScript. Er is geen complex framework nodig om deze handelingen te kunnen ontwikkelen, alsook is de kennis sporadisch online beschikbaar. Sites en webontwikkelaars houden zelden rekening met deze personalisatiemogelijkheden, al wijst dit onderzoeksdeel uit dat deze methoden

5

Conclusie

Verschillen tussen de handmatige tekst, het origineel, uitvoer van beschikbare tools en het prototype.

Uit een vergelijkende studie blijkt dat er lichte verschillen zijn tussen de verschillende tools wat betreft de scores op het gebied van FRE, SMOG en Kincaid. Het prototype en het GPT-3-model dat wordt gebruikt voor gepersonaliseerde samenvattingen scoren beter op leesbaarheid. Een niet-gepersonaliseerde samenvatting scoort echter lager en heeft problemen in de doorlopende tekst, waarbij het taalmodel soms woorden plaatst die niet nuttig zijn in de context. Resoomer en Scispace verkorten effeHierdoor wordt de bruikbaarheid van deze scores en het gebruik van de Python-bibliotheek in twijfel getrokken. Hoewel deze scores een goed alternatief zijn om de leesbaarheid te meten, houden ze geen rekening met verkeerd geïnterpreteerde resultaten, zoals letterlijk overgenomen bronvermeldingen of verkeerd gegenereerde woordenschat door het taalmodel. Deze fouten zijn minder vaak aanwezig bij het vereenvoudigen van teksten met GPT-3. Er is meer onderzoek nodig om de bruikbaarheid van deze scores te bepalen en om te begrijpen hoe de scores zich verhouden tot de kwaliteit van de vereenvoudigde tekst.

Het vergelijken met een referentietekst blijft nog steeds een handmatige vergelijking en biedt een inkijk in hoe lectoren teksten kunnen vereenvoudigen. De vergelijkende studie hield geen rekening met het hoofdstuk waarin een zin werd beoordeeld. Vragen naar het verschil na een tekstvereenvoudiging per hoofdstuk in een wetenschappelijk artikel kan daarmee niet beantwoord worden en moet opgevolgd worden in een verder onderzoek.

Het gebruik van HuggingFace-taalmodellen en LLM's voor een toepassing van tekstvereenvoudiging.

Bestaande taalmodellen maken de ontwikkeling van toepassingen op het gebied van semantische analyse, kernwoordenidentificatie en het extraheren van samenvattingen veel eenvoudiger. Een prototype heeft aangetoond dat vrij beschikbare taalmodellen, zoals BERT en gerelateerde BERT-varianten, een oplossing bieden voor deze taken. Voor meer complexe vereenvoudigingstechnieken, zoals lexicale en syntactische vereenvoudiging en abstraherende samenvatting, zijn complexere taalmodellen zoals GPT-3 geschikt. Ontwikkelaars moeten echter rekening houden met de schaal van de modellen bij het maken van deze keuze. Hoewel vrij beschikbare modellen op HuggingFace in staat zijn om abstraherende samenvattingen of lexicale vereenvoudiging mogelijk te maken, staan ze in de schaduw van GPT-3, dat voor ontwikkelaars vrij beschikbaar is.

Omgekeerd is GPT-3 echter een overkill voor taken zoals het identificeren van kernwoorden of het aanduiden van belangrijke zinnen, die kosteneffectief kunnen worden aangepakt zonder het gebruik van GPT-3. GPT-3 moet niet voor iedere functionaliteit worden ingezet om zo kostenbesparend te werken. Het GPT-3 model maakt complexe en granulaire NLP-transformaties op lexicaal en syntactisch niveau mogelijk voor geautomatiseerde tekstvereenvoudiging. Echter houdt het model geen rekening met referenties buiten de getrainde data, wat tot problemen bij de credibiliteit van data kan leiden. Bing AI daarentegen doet dit wel en vormt een goede fundering voor ontwikkelaars om meer referentiemateriaal aan te bieden in ondersteunende software binnen het onderwijs. Verder onderzoek op de toepassing van deze AI via een mogelijke API is zeker nodig en kan baanbrekend zijn voor de onderwijssector.

Verdere finetuning en pre-training van taalmodellen.

HuggingFace-taalmodellen bieden een gratis alternatief voor ontwikkelaars bij het creëren van prototypes of volledig functionerende webtoepassingen. Hoewel deze modellen in staat zijn om tekst te vereenvoudigen zonder hinder van idiomen en ambiguïteit, zijn ze niet bestand tegen ontbrekende woorden en wordt er geen rekening gehouden met de doelgroep, wat bij geautomatiseerde tekstvereenvoudiging van wetenschappelijke artikelen een belemmering kan vormen. Het huidige prototype maakt geen gebruik van verdere finetuning op HuggingFace-taalmodellen.

Het toegepaste GPT-3 model is enkel gefinetuned per API-parameters en bevat geen vooraf getrainde extra data van wetenschappelijke papers. Er is echter wetenschappelijke data beschikbaar die kan worden gebruikt om het GPT-3 taalmodel accurater te maken op interpretatie van complexiteit bij wetenschappelijke artikelen. Er is een licht effect waargenomen op de verschillen in lexicale complexiteit tussen de HuggingFace-taalmodellen die wel getraind zijn op wetenschappelijke

papers in vergelijking met taalmodellen die getraind zijn op algemene data, maar meer onderzoek is nodig om deze verschillen beter te begrijpen binnen de context van wetenschappelijke papers. Het is belangrijk op te merken dat de taalmodellen van OpenAI voortdurend evolueren en dat er overwogen wordt om GPT-2 achterwege te laten in het licht van verdere edities van de GPT-modellen. Op dit moment worden GPT-4 en Bing AI uitgerold, maar deze zijn nog niet klaar voor gebruik in productie, dus verder onderzoek is nodig naar het gebruik van deze modellen in het onderwijs. GPT-3 kan een baanbrekende oplossing aanbieden voor geautomatiseerde tekstvereenvoudiging van wetenschappelijke artikelen.

- Het gebruik van GPT-3 maakt het mogelijk om moeilijke woorden snel en efficiënt te identificeren binnen een doorlopende tekst. Toekomstig onderzoek zou zich kunnen richten op het potentieel van de combinatie van GPT-3 en full-text-search technologieën, waarbij specifieke zoektermen en thema's worden gebruikt om woordenlijsten te genereren. Dit zou kunnen bijdragen aan een nog meer geoptimaliseerde ondersteuning van het leerproces.

Met alsmaar grotere taalmodellen, zoals het opkomende GPT-4 en LLaMa, is er ook meer onderzoek nodig naar de verschillen op taalvlak ten opzichte van de toename in parameters. Het GPT-3 model dat in dit onderzoek werd gebruikt, maakte enkel gebruik van aangepaste parameters zoals de *temperature* en *top_p*. Hoewel de overstap naar andere taalmodellen kostelijk kan zijn voor ontwikkelaars, is het belangrijk om te onderzoeken of en hoe deze nieuwe modellen kunnen bijdragen tot betere resultaten. Het is echter eerder uitgewezen dat de grootte van taalmodellen alsmaar minder relevant wordt.

Software ontwikkelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs

De erkende software uitgeleend aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs voldoet niet aan de noden. De software biedt ondersteunende functionaliteiten aan zoals het aanmaken van een woordenlijst, alsook het markeren van zinnen om deze later om te vormen naar een tekst. Syntactische vereenvoudiging of abstraherende samenvatting zijn niet tot de orde. Online toepassingen staan verder en reiken functionaliteiten aan die hoogstaand zijn en reproduceerbaar zijn voor mensen met informaticakennis. Echter is er geen manusje-van-alles en er is daarmee nood aan een toepassing die alle functionaliteiten kan combineren. De erkende softwarepakketten zoals Kurzweil kunnen opgeschaald worden, zodat deze de functionaliteiten hebben om verbeterde tekstvereenvoudigingstechnieken aan te reiken aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs.

Functionaliteiten combineren is een haalbare zaak voor zowel klein- als grootschalige softwareondernemingen. Het prototype is opgebouwd uit kennis en tools die

aangeleerd worden in alle richtingen Toegepaste Informatica bij alle Vlaamse hogescholen. Met gebruik van kant-en-klare taalmodellen, API's en gekende programmeertalen zijn ontwikkelaars ertoe in staat om een webtoepassing te ontwikkelen die ondersteuning kan bieden aan scholieren met dyslexie in de derde graad van het middelbaar onderwijs. Dit prototype is gebouwd op eerder onderzochte visuele kenmerken en de impact van een vereenvoudigde tekst op de leessnelheid en -begrip bij een scholier met dyslexie tijdens het intensief lezen van een tekst. Het prototype werd niet uitgetest bij het doelpubliek en fungeert enkel als meting van de haalbaarheid voor een dergelijke tool.

Het prototype is ontwikkeld met een snelle en intuïtieve deployment in gedachten. Door het meegegeven script-bestand is enkel de installatie van Docker Desktop vereist. Al is een online deployment optimaler, voor experimentdoeleinden is deze opzet nog steeds ideaal en met een intuïtieve handeling kunnen computerleeks ook van dit prototype gebruikmaken inclusief met de gegeven instructies.

- Onderzoekers binnen het vakdomein logopedie kunnen dit prototype gebruiken om experimenten af te nemen die het effect op leesbegrip achterhaalt bij scholieren met dyslexie in de derde graad van het middelbaar onderwijs.
- Onderzoekers binnen het vakdomein secundair onderwijs kunnen de effecten bij leerkrachten in het middelbaar onderwijs waarnemen en concluderen of deze tool al dan niet van pas kan komen.
- Er is echter meer onderzoek nodig naar hoe de inzet van webtoepassingen en browserextensies voor tekstvereenvoudiging in het onderwijs kan worden verbeterd.

Onderzoek naar het verschil tussen het laten schrijven van prompts en vooraf gedefinieerde prompts is schaars, maar deze keuze kan een effect hebben op het gedrag of ervaring van de eindgebruiker. De doelgroep wordt expliciet aangeduid in de prompts en is daarmee parameteriseerbaar. Er is verder onderzoek nodig naar de effecten op het meegeven van doelgroepen via prompts en of deze al dan niet rekening houden met de doelgroep.

Einde

Wetenschappelijke artikelen worden alsmaar complexer en de nood aan een ondersteunende tool zal hierdoor enkel maar toenemen. De ontwikkeling van het prototype moet ontwikkelaars de aanzet geven om deze API's en middelen in handen te nemen, om zo een kwalitatieve (web)toepassing te ontwikkelen met oog op scholieren voor dyslexie in de derde graad van het middelbaar onderwijs.



Onderzoeksvoorstel

Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in het derde graad middelbaar onderwijs bij het lezen van wetenschappelijke artikelen. Gepersonaliseerde tekstvereenvoudiging helpt deze scholieren bij hun leesbegrip. Daarnaast kan artificiële intelligentie (AI) dit proces automatiseren om de werkdruk bij leraren en scholieren te verminderen. Dit onderzoek achterhaalt met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een AI-toepassing voor adaptieve en geautomatiseerde tekstvereenvoudiging. Hiervoor is de volgende onderzoeksvraag opgesteld: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in het derde graad middelbaar onderwijs?". Een vergelijkende studie beantwoordt deze onderzoeksvraag en is uitgevoerd met bestaande toepassingen en een prototype voor adaptieve en geautomatiseerde tekstvereenvoudiging. Uit de vergelijkende studie blijkt dat toepassingen om wetenschappelijke artikelen te vereenvoudigen, gemaakt zijn voor een centrale doelgroep en geen rekening houden met de unieke noden van een scholier met dyslexie in het derde graad middelbaar onderwijs. Adaptieve software voor geautomatiseerde tekstvereenvoudiging is mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

A.1. Introductie

Het Vlaams middelbaar onderwijs staat op barsten. Leraren en scholieren worden overspoeld door werkdruk en stress. Bovendien is de derde graad van het middelbaar onderwijs een belangrijke mijlpaal voor de verdere loopbaan van scholieren, al hebben zij volgens Dapaah en Maenhout (2022) dan moeite om grip te krijgen

op de vakliteratuur bij STEM-vakken. Het STEM-agenda¹ van de Vlaamse Overheid moet het STEM-onderwijs tegen 2030 aantrekkelijker te maken, door de ondersteuning voor zowel leerkrachten als scholieren te verbeteren. Toch wordt het aanpakken van de steeds complexere wetenschappelijke taal, zoals beschreven in Barnett en Doubleday (2020), niet opgenomen in het STEM-agenda. Wetenschappelijke artikelen vereenvoudigen, op maat van de noden voor een scholier met dyslexie in het middelbaar onderwijs, is tijds- en energie-intensief voor leerkrachten en scholieren. Automatische en adaptieve tekstvereenvoudiging biedt hier een baanbrekende oplossing om de werkdruk in het middelbaar onderwijs te verminderen. Het doel van dit onderzoek is om te achterhalen met welke technologische en logopedische aspecten AI-ontwikkelaars rekening moeten houden bij de ontwikkeling van een adaptieve AI-toepassing voor geautomatiseerde tekstvereenvoudiging. De volgende onderzoeksvraag is opgesteld: "Hoe kan een wetenschappelijke artikel automatisch vereenvoudigd worden, gericht op de verschillende behoeften van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Het doel wordt bereikt door een antwoord op de volgende deelvragen te formuleren. Eerst geeft de literatuurstudie een antwoord op de eerste vier deelvragen. Daarna vormt het veldonderzoek een antwoord op de vijfde deelvraag. Ten slotte beantwoordt de vergelijkende studie de zesde en laatste deelvraag. De resultaten van dit onderzoek zetten AI-ontwikkelaars aan om een toepassing te maken om scholieren met dyslexie te kunnen ondersteunen in de derde graad middelbaar onderwijs.

1. Welke aanpakken zijn er voor geautomatiseerde tekstvereenvoudiging? Aansluitende vraag: "Hoe worden teksten handmatig vereenvoudigd voor scholieren met dyslexie?"
2. Welke specifieke noden hebben scholieren van de derde graad middelbaar onderwijs bij het begrijpen van complexere teksten?
3. Wat zijn de specifieke kenmerken van wetenschappelijke artikelen?
4. Met welke valkuilen bij taalverwerking met AI moeten ontwikkelaars rekening houden?
5. Welke toepassingen, tools en modellen zijn er beschikbaar om Nederlandstalige geautomatiseerde tekstvereenvoudiging met AI mogelijk te maken?
6. Welke functies ontbreken AI-toepassingen om geautomatiseerde én adaptieve tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs? Aansluitende vraag: "Welke manuele methoden voor tekstvereenvoudiging komen niet in deze tools voor?"

¹<https://www.vlaanderen.be/publicaties/stem-agenda-2030-stem-competenties-voor-een-toekomst-en-missiegericht-beleid>

A.2. State-of-the-art

A.2.1. Tekstvereenvoudiging

De voorbije tien jaar is artificiële intelligentie (AI) sterk verder ontwikkeld. Vasista (2022) benadrukt dat de toename in kennis voor nieuwe toepassingen zorgde. Tekstvereenvoudiging vloeide hier uit voort. Momenteel bestaan er al robuuste toepassingen die teksten kunnen vereenvoudigen, zoals Resoomer², Paraphraser³ en Prepostseo⁴. Binnen het kader van tekstvereenvoudiging is er bestaande documentatie beschikbaar waar onderzoekers het voordeel van toegankelijkheid aanhalen, maar volgens Gooding (2022) ontbreken deze toepassingen de extra noden die scholieren met dyslexie in de derde graad middelbaar onderwijs vereisen.

Shardlow (2014) haalt aan dat het algemene doel van tekstvereenvoudiging is om ingewikkelde bronnen toegankelijker te maken. Het zorgt voor verkorte teksten zonder de kernboodschap te verliezen. Siddharthan (2014) haalt verder aan dat tekstvereenvoudiging op één van drie manieren gebeurt. Er is conceptuele vereenvoudiging waarbij documenten naar een compacter formaat worden getransformeerd. Daarnaast is er uitgebreide modificatie die kernwoorden aanduidt door gebruik van redundantie. Als laatste is er samenvatting die documenten verandert in kortere teksten met alleen de topische zinnen. Met deze concepten zijn ontwikkelaars volgens Siddharthan (2014) in staat om ingewikkelde woorden te vervangen door eenvoudigere synoniemen of zinnen te verkorten zodat ze sneller leesbaar zijn.

Tekstvereenvoudiging behoort tot de zijtak van *Natural Language Processing* (NLP) in AI. NLP omvat methodes om menselijke teksten om te zetten in tekst voor machines. Documenten vereenvoudigen met NLP kan volgens Chowdhary (2020) op twee manieren: extract of abstract. Bij extractieve vereenvoudiging worden zinnen gelezen zoals ze zijn neergeschreven. Vervolgens bewaart een document de belangrijkste taalelementen om de tekst te kunnen hervormen. Deze vorm van tekstvereenvoudiging komt volgens (Sciforce, 2020) het meeste voor. Daarnaast is er abstracte vereenvoudiging waarbij de kernboodschap wordt bewaard. Met de kernboodschap wordt er een nieuwe zin opgebouwd. Volgens het onderzoek van Chowdhary (2020) heeft deze vorm potentieel, maar het zit nog in de kinderschoenen.

A.2.2. Noden van scholieren met dyslexie

Het experiment van Franse wetenschappers

Gala en Ziegler (2016) illustreert dat manuele tekstvereenvoudiging schoolteksten toegankelijker

maakt voor kinderen met dyslexie. Dit deden ze door simpelere synoniemen en

²<https://resoomer.com/nl/>

³<https://www.paraphraser.io/nl/tekst-samenvatting>

⁴<https://www.prepostseo.com/tool/nl/text-summarizer>

zinsstructuren te gebruiken. Tien kinderen werden opgenomen in het experiment, variërend van 8 tot 12 jaar oud. Verwijswoorden werden vermeden en woorden kort gehouden. De resultaten waren veelbelovend. Het leestempo lag hoger en de kinderen maakten minder leesfouten. Ook bleek er geen verlies van begrip in de tekst bij geteste kinderen. Resultaten van de studie werden gebundeld voor de mogelijke ontwikkeling van een AI-tool.

De visuele weergave van tekst beïnvloedt de leessnelheid bij scholieren met dyslexie. Zo haalt het onderzoek van Rello e.a. (2012) tips aan waarmee teksten en documenten rekening moeten houden bij scholieren met dyslexie in de derde graad middelbaar onderwijs. Het gaat over speciale lettertypes, spreiding tussen woorden en het gebruik van inzoomen op aparte zinnen. Het onderzoek haalt verder aan dat teksten voor deze unieke noden aanpassen tijdrovend is, dus tekstvereenvoudiging door AI kan een revolutionaire oplossing bieden. De Universiteit van Copenhagen is met bovenstaande idee aan de slag gegaan. Onderzoekers Bingel e.a. (2018) hebben gratis software ontwikkeld, genaamd Hero⁵, om tekstvereenvoudiging voor scholieren in het middelbaar onderwijs met dyslexie te automatiseren. De software bestudeert met welke woorden de gebruiker moeite heeft, en vervangt die door simpelere alternatieven. Hero bevindt zich nu in beta-vorm en wordt enkel in het Engels en Deens ondersteund. Als alternatief is er Readable⁶. Dit is een Engelstalige AI-toepassing dat zinnen beoordeeld met leesbaarheidsformules.

Roldós (2020) haalt aan dat NLP in de laatste decennia volop in ontwikkeling is, maar ontwikkelaars botsen nog op uitdagingen. Het gaat om zowel interpretatie- als dataproblemen bij AI-modellen. Het onderzoek haalt twee punten aan. Allereerst is het voor een machine moeilijk om de context van homoniemen te achterhalen. Bijvoorbeeld bij het woord 'bank' is het niet duidelijk voor de machine of het gaat over de geldinstelling of het meubel. Daarnaast zijn synoniemen een probleem voor tekstverwerking.

Het onderzoek van Sciforce (2020) haalt aan dat het merendeel van NLP-toepassingen Engelstalige invoer gebruikt. Niet-Engelstalige toepassingen zijn zeldzaam. De opkomst van AI technologieën die twee datasets gebruiken, biedt een oplossing voor dit probleem. De software vertaalt eerst de oorspronkelijke tekst naar de gewenste taal, voordat de tekst wordt herwerkt. Hetzelfde onderzoek bewijst dat het vertalen van gelijkaardige talen, zoals Duits en Nederlands, een minimaal verschil opleverd. Volgens Plavén-Sigray e.a. (2017) houden onderzoekers zich vaak in hun eigen taalbubbel, wat negatieve gevolgen heeft voor de leesbaarheid van een wetenschappelijk artikel. Bovendien vormt de stijgende trend van het gebruik aan acroniemen Barnett en Doubleday (2020) een extra hindernis. Donato e.a. (2022) haalt aan dat onbegrijpelijke literatuur, waaronder studiemateriaal geschreven door de docent en online wetenschappelijke artikelen, één van de redenen is waarom scholieren

⁵<https://beta.heroapp.ai/>

⁶<https://readable.com/>

met dyslexie in het middelbaar onderwijs van richting veranderen.

A.2.3. Huidige toepassingen

Vlaanderen heeft weinig zicht op de geïmplementeerde AI software in scholen. Dit werd vastgesteld door (Martens e.a., 2021a), een samenwerking tussen de Vlaamse universiteiten en overheid voor AI. Vergeleken met andere Europese landen, maakt België het minst gebruik van leerling-georiënteerde hulpmiddelen. Degenen die wel gebruikt worden, zijn vooral online leerplatformen voor zelfstandig werken. Ook maakt België amper gebruik van beschikbare software die de leermethoden en -noden van leerlingen evalueert (Martens e.a., 2021b).

Verhoeven (2023) haalt aan dat AI-toepassingen zoals ChatGPT, Google Bard en Bing AI kunnen helpen om routinematig werk te verminderen in het onderwijs. Echter haalt Deckmyn (2021) aan dat GPT-3, het model van ChatGPT, sterker staat in het maken van Engelstalige teksten vergeleken met Nederlandstalige teksten. De databank waar het GPT-3 model mee is getraind, bestaat uit 92% Engelstalige woorden, terwijl er 0,35% Nederlandse woorden aanwezig zijn in dezelfde databank. Ontwikkelaars moeten de evolutie van deze modellen opvolgen, voordat er Nederlandstalige toepassingen mee worden gemaakt.

A.2.4. Ontwikkelen met AI

Python staat bovenaan de lijst van programmeertalen voor NLP-toepassingen. Volgens het onderzoek van Thangarajah (2019) is dit te wijten aan de eenvoudige syntax, kleine leercurve en grote beschikbaarheid van kant-en-klare bibliotheken. Wetenschappelijke berekeningen of statistische analyses kunnen worden uitgevoerd met één lijn code. Malik (2022) haalt de twee meest voorkomende aan, namelijk NLTK⁷ en Spacy⁸. *Deep Martin*⁹ bouwt verder op het onderzoek van Shardlow (2014) naar een pipeline voor lexicale vereenvoudiging. *Deep Martin* maakt gebruik van *custom transformers* om invoertekst te converteren naar een vereenvoudigde versie van de tekstinhoud.

Voor Germaanse talen zijn er enkele datasets en word embeddings beschikbaar die de complexiteit van woorden bijhouden. Zo zijn er in de Duitse taal Klexikon¹⁰ en TextComplexityDE¹¹. Een onderzoek van Suter e.a. (2016) bouwde een rule-based NLP-model met 'Leichte Sprache', wat een dataset is met eenvoudige Duitstalige zinsconstructies. Nederlandstalige datasets zijn in schaarse hoeveelheden beschikbaar, dus het vertalen uit een Germaanse taal is hier een optie.

Volgens Garbacea e.a. (2021) is het belangrijk dat AI-ontwikkelaars niet alleen aandacht besteden aan het aanpassen van woorden en zinnen, maar ook aan de ge-

⁷<https://www.nltk.org/>

⁸<https://spacy.io/>

⁹<https://github.com/chrislemke/deep-martin>

¹⁰<https://github.com/dennlinger/klexikon>

¹¹<https://github.com/babaknaderi/TextComplexityDE>

bruiker meegeven waarom iets is aangepast. De onderzoekers wijzen op twee ethische aspecten. Eerst moet de toepassing duidelijk aangeven waarom een woord of zin is aangepast. Het model moet de moeilijkheidsgraad van de woorden of zinnen bewijzen. Iavarone e.a. (2021) beschrijft een methode met regressiemodellen om de moeilijkheidsgraad te bepalen door een gemiddeld moeilijkheidspercentage per zin te berekenen. Daarnaast benadrukt Garbacea e.a. (2021) het belang van het markeren van de complexere delen van een tekst. Hiervoor haalt hetzelfde onderzoek methoden aan zoals *lexical* of *deep learning*.

Er is een tactvolle aanpak nodig om een vereenvoudigde tekst met AI te beoordelen. De studie van Swayamdipta (2019) haalt aan dat er extra nood is aan NLP-modellen waarbij de tekst zijn kernboodschap behoudt. Samen met Microsoft Research bouwden ze NLP-modellen die gericht waren op de bewaring van zinsstructuur en -context door *scaffolded learning*. Hiervoor maakten de onderzoekers gebruik van een voorspellingsmethode die de positie van woorden en zinnen in een document beoordeelde. De Flesch-Kincaid leesbaarheidstest is volgens

Readable (2021) een alternatieve manier om vereenvoudigde tekstinhoud te beoordelen, zonder de nood aan *pre-trained* modellen. Deze score kan eenvoudig worden berekend met de *Python-library textstat*¹².

A.3. Methodologie

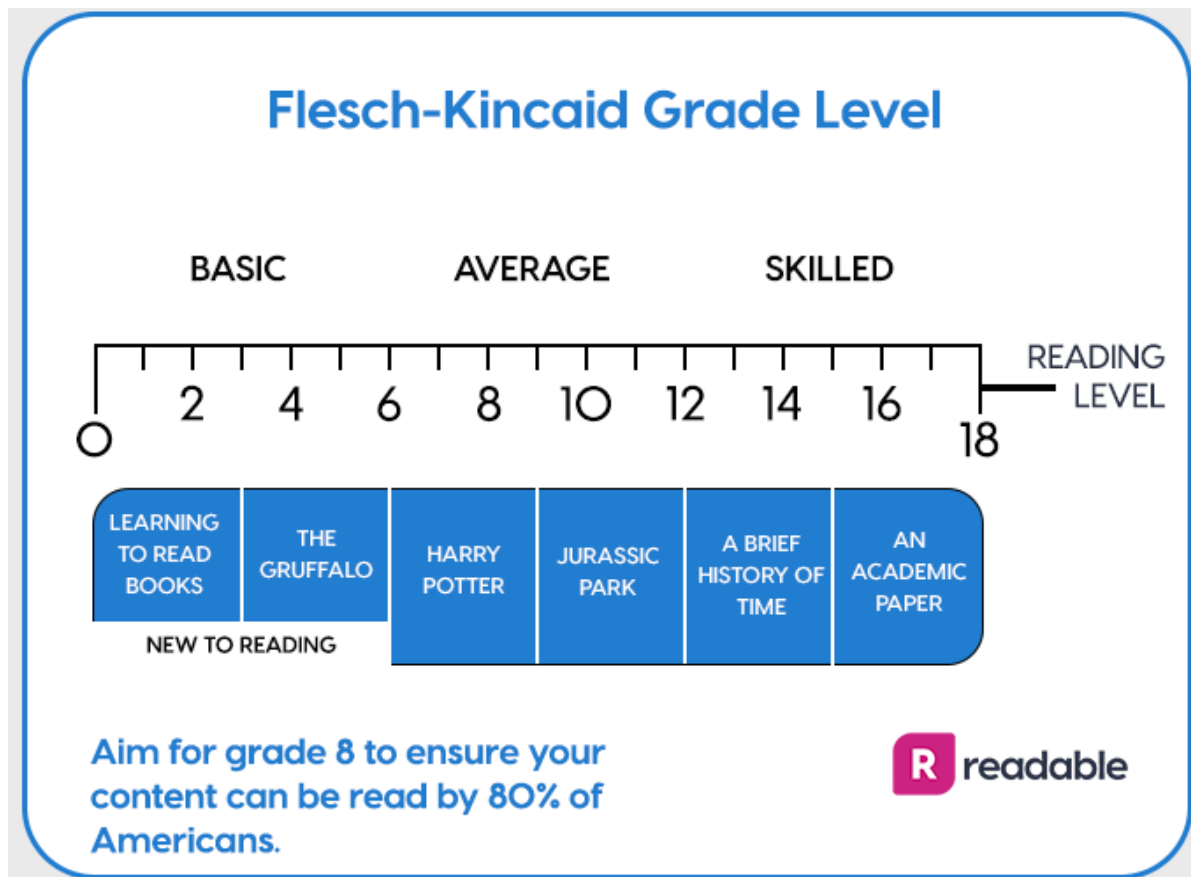
Een *mixed-methods* onderzoek toont aan hoe toepassingen automatisch een wetenschappelijke artikel kunnen vereenvoudigen, gericht op scholieren met dyslexie in de derde graad middelbaar onderwijs. Het onderzoek houdt vijf grote fases in. De eerste fase is het proces van geautomatiseerde tekstvereenvoudiging beschrijven. Dit gebeurt via een grondige studie van vakliteratuur en wetenschappelijke teksten. Ook blogs van experts komen hier aan bod. Na het verwerven van de nodige inzichten wordt er een verklarende tekst opgesteld.

De tweede fase bestaat uit het analyseren van wetenschappelijke werken over de bewezen voordelen van tekstvereenvoudiging bij scholieren met dyslexie van de derde graad middelbaar onderwijs. Hiervoor zijn geringe thesissen beschikbaar, die zorgvuldigheid vragen tijdens interpretatie. De resulterende tekst bevat de voordelen samen met hun wetenschappelijke onderbouwing.

De derde fase is opnieuw een beschrijving. Hier worden de valkuilen bij taalverwerking met AI-software nagegaan. Deze fase van het onderzoek brengt mogelijke nadelen en tekortkomingen van AI-software bij tekstvereenvoudiging aan het licht. Dit gebeurt aan de hand van een technische uitleg.

De vierde fase omvat een toelichting over beschikbare AI toepassingen voor tekstvereenvoudiging. Aan de hand van een veldonderzoek op het internet en bij bedrijven wordt een longlist opgesteld van beschikbare toepassingen voor tekstvereenvoudiging in het middelbaar onderwijs. Met een requirementsanalyse wordt

¹²<https://pypi.org/project/textstat/>



Figuur (A.1)
(Readable, 2021)

er een shortlist opgesteld van software. Het toetsen van verschillende tools wordt ook betrokken in deze fase. De shortlist vormt de basis voor de ontwikkeling van een prototype voor geautomatiseerde en adaptieve tekstvereenvoudiging.

De vijfde en laatste fase van het onderzoek bestaat uit het testen en beoordelen van gekozen AI-toepassingen voor tekstvereenvoudiging. In dit experiment proberen scholieren met dyslexie in de derde graad middelbaar onderwijs de shortlisted AI toepassingen en het prototype uit. Het doel van het experiment is om de effectiviteit en gebruikersvriendelijkheid van deze toepassingen te beoordelen. Na een grondige analyse wordt er met de resultaten bepaalt of de toepassingen aan de unieke noden van een scholier met dyslexie in de derde graad middelbaar onderwijs voldoen om wetenschappelijke artikelen te vereenvoudigen voor scholieren in het middelbaar onderwijs.

A.4. Verwacht resultaat, conclusie

Er wordt verwacht dat de huidige softwareoplossingen voor tekstvereenvoudiging onvoldoende aansluiten bij de noden van scholieren met dyslexie in de derde graad middelbaar onderwijs. Het prototype is moeilijk af te stemmen op de specifieke noden van deze doelgroep. Ontwikkelaars die werken met bestaande modellen moeten *custom transformers* inzetten om bevredigende resultaten te krijgen. Bovendien ontbreken er Nederlandstalige word embeddings die de complexiteit van elk woord bijhouden en aan kant-en-klare modellen die de inhoud van wetenschappelijke artikelen kunnen vereenvoudigen. Word embeddings uit een Germaanse taal gebruiken, gevolgd door vertaling naar het Nederlands is wel een aanvaardbaar alternatief.

Bibliografie

- Althunayyan, S. & Azmi, A. (2021). Automated Text Simplification: A Survey. *ACM Computing Surveys*, 54, Article no. 43. <https://doi.org/10.1145/3442695>
- Ball, P. (2017). It's not just you: science papers are getting harder to read. *Nature*.
- Barnett, A. & Doubleday, Z. (2020). Meta-Research: The growth of acronyms in the scientific literature (P. Rodgers, Red.). *eLife*, 9, e60080.
- Bezem, A. & Lugthart, M. (2016). Visuele Disfunctie een onzichtbare belemmering bij lezen, spelling en concentratie. <https://beeldenbrein.nl/>
- Bilici, Ş. (2021). Sequence labeling.
- Bingel, J., Paetzold, G. & Søgaard, A. (2018). Lexi: A tool for adaptive, personalized text simplification. *Proceedings of the 27th International Conference on Computational Linguistics*, 245–258.
- Binz, M. & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, 120(6).
- Bonte, M. (2020). *Bestaat Dyslexie?: En is het een relevante vraag?* uitgeverij SWP.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners.
- Bulté, B., Sevens, L. & Vandeghinste, V. (2018). Automating lexical simplification in Dutch. *Computational Linguistics in the Netherlands Journal*, 8, 24–48. <https://clinjournal.org/clinj/article/view/78>
- Canning, Y., Tait, J., Archibald, J. & Crawley, R. (2000). Cohesive Generation of Syntactically Simplified Newspaper Text. In P. Sojka, I. Kopeček & K. Pala (Red.), *Text, Speech and Dialogue* (pp. 145–150). Springer Berlin Heidelberg.
- Cao, M. (2022). A Survey on Neural Abstractive Summarization Methods and Factual Consistency of Summarization.
- Chowdhary, K. (2020). *Fundamentals of Artificial Intelligence*. Springer, New Delhi.
- Coster, W. & Kauchak, D. (2011). Learning to Simplify Sentences Using Wikipedia. *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, 1–9. <https://aclanthology.org/W11-1601>
- Crevits, H. (2022, maart 13). *Kwart van bedrijven gebruikt artificiële intelligentie: Vlaanderen bij beste leerlingen van de klas* (Persbericht). Vlaamse Overheid Departement Economie, Wetenschap en Innovatie.

- Dandekar, N. (2016). How to use machine learning to find synonyms. <https://medium.com/@nikhilbd/how-to-use-machine-learning-to-find-synonyms-6380c0c6106b>
- Dapaah, J. & Maenhout, K. (2022, juli 8). *Iedereen heeft boter op zijn hoofd* (D. Standaard, Red.). https://www.standaard.be/cnt/dmf20220607_97763592
- De Craemer, J., Van Beeumen, L., Cooreman, A., Moonen, A., Rottier, J., Wagemakers, I. & Mardulier, T. (2018). Aan de slag met voorleessoftware op school. Een gids met 8 vragen en antwoorden. <https://onderwijs.vlaanderen.be/nl/onderwijspersoneel/van-basis-tot-volwassenenonderwijs/lespraktijk/ict-in-de-klas/voorleessoftware-voor-leerlingen-met-leesbeperkingen/aan-de-slag-met-voorleessoftware-op-school>
- De Meyer, I., Janssens, R. & Warlop, N. (2019). Leesvaardigheid van 15- jarigen in Vlaanderen: Overzicht van de eerste resultaten van PISA2018. <https://data-onderwijs.vlaanderen.be/documenten/bestand.ashx?id=12265>
- Deckmyn, D. (2021, maart 19). *Robot schrijft mee De Standaard* (D. Standaard, Red.). https://www.standaard.be/cnt/dmf20210319_05008561
- Donato, A., Muscolo, M., Arias Romero, M., Caprì, T., Calarese, T. & Olmedo Moreno, E. M. (2022). Students with dyslexia between school and university: Post-diploma choices and the reasons that determine them. An Italian study. *Dyslexia*, 28(1), 110–127.
- DuBay, W. H. (2004). The principles of readability. *Online Submission*.
- Eisenstein, J. (2019). *Introduction to Natural Language Processing*. MIT Press. <https://books.google.be/books?id=72yuDwAAQBAJ>
- Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R. & Radev, D. (2020). SummEval: Re-evaluating Summarization Evaluation.
- Gala, N. & Ziegler, J. (2016). Reducing lexical complexity as a tool to increase text accessibility for children with dyslexia. *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, 59–66.
- Garbacea, C., Guo, M., Carton, S. & Mei, Q. (2021). Explainable Prediction of Text Complexity: The Missing Preliminaries for Text Simplification. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1086–1097. <https://doi.org/10.18653/v1/2021.acl-long.88>
- Garg, H. (2022). Using GPT-3 for education: Use cases. <https://indiaai.gov.in/article/using-gpt-3-for-education-use-cases>
- Ghesquière, P. (2018). *Als leren pijn doet: Kinderen met een leerstoornis opvoeden en begeleiden*. Acco.
- Gooding, S. (2022). On the Ethical Considerations of Text Simplification. *Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022)*, 50–57. <https://doi.org/10.18653/v1/2022.slpac-1.7>

- Gooding, S. & Kochmar, E. (2019). Complex word identification as a sequence labelling task. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 1148–1153.
- Greg, B., Atty, E., Elie, G., Joane, J., Logan, K., Lim, R., Luke, M. & Michelle, P. (2023). Introducing chatgpt and Whisper Apis. <https://openai.com/blog/introducing-chatgpt-and-whisper-apis>
- Hahn, U. & Mani, I. (2000). The Challenges of Automatic Summarization. *Computer*, 33, 29–36. <https://doi.org/10.1109/2.881692>
- Hartley, J. (1999). From Structured Abstracts to Structured Articles: A Modest Proposal. *Journal of Technical Writing and Communication*, 29(3), 255–270. <https://doi.org/10.2190/3RWW-A579-HC8W-6866>
- Harwell, D. (2023). Tech's hottest new job: Ai whisperer. no coding required. <https://www.washingtonpost.com/technology/2023/02/25/prompt-engineers-techs-next-big-job/>
- Hayes, D. P. (1992). The growing inaccessibility of science. <https://www.nature.com/articles/356739a0>
- Hern, A. (2023). TechScape: Will meta's massive leak democratise AI – and at what cost? <https://www.theguardian.com/technology/2023/mar/07/techscape-meta-leak-llama-chatgpt-ai-crossroads>
- Hollenkamp, J. (2020). Summary and analysis of Scientific Research Articles - San Jose State ... <https://www.sjsu.edu/writingcenter/docs/handouts/Summary%20and%20Analysis%20of%20Scientific%20Research%20Articles.pdf>
- Hsu, W.-T., Lin, C.-K., Lee, M.-Y., Min, K., Tang, J. & Sun, M. (2018). A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss.
- Huang, S., Wang, R., Xie, Q., Li, L. & Liu, Y. (2019). An Extraction-Abstraction Hybrid Approach for Long Document Summarization. *2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, 1–6.
- Hubbard, K. E. & Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PLOS ONE*, 12(12), 1–16.
- Iavarone, B., Brunato, D. & Dell'Orletta, F. (2021). Sentence Complexity in Context. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 186–199. <https://doi.org/10.18653/v1/2021.cmcl-1.23>
- IBM. (2022). IBM Global AI Adoption Index 2022. <https://www.ibm.com/downloads/cas/GVAGA3JP>
- Iredale, G. (2022). An overview of tokenization algorithms in NLP. <https://101blockchains.com/tokenization-nlp/>
- Iskender, N., Polzehl, T. & Möller, S. (2021). Reliability of Human Evaluation for Text Summarization: Lessons Learned and Challenges Ahead. *Proceedings of the*

- Workshop on Human Evaluation of NLP Systems (HumEval)*, 86–96. <https://aclanthology.org/2021.humeval-1.10>
- Jiang, R. K. (2023). Prompt engineering : Deconstructing and managing intention. <https://www.linkedin.com/pulse/prompt-engineering-deconstructing-managing-intention-jiang/>
- Jones, R., Colusso, L., Reinecke, K. & Hsieh, G. (2019). r/science: Challenges and Opportunities in Online Science Communication. *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3290605.3300383>
- Jurafsky, D., Martin, J., Norvig, P. & Russell, S. (2014). *Speech and Language Processing*. Pearson Education. <https://books.google.be/books?id=Cq2gBwAAQBAJ>
- Kandula, S., Curtis, D. & Zeng-Treitler, Q. (2010). A semantic and syntactic text simplification tool for health content. *AMIA annual symposium proceedings, 2010*, 366.
- Khan, A. (2014). A Review on Abstractive Summarization Methods. *Journal of Theoretical and Applied Information Technology*, 59, 64–72.
- Khurana, D., Koli, A., Khatter, K. & Singh, S. (2022). Natural Language Processing: State of The Art, Current Trends and Challenges. *Multimedia Tools and Applications*, 82, 25–27.
- Li, C. (2022). OpenAI's GPT-3 language model: A technical overview. <https://lambdalabs.com/blog/demystifying-gpt-3>
- Li, J., Sun, A., Han, J. & Li, C. (2018). A Survey on Deep Learning for Named Entity Recognition.
- Lin, H. & Bilmes, J. (2010). Multi-document summarization via budgeted maximization of submodular functions. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 912–920.
- Linderholm, T., Everson, M. G., van den Broek, P., Mischinski, M., Crittenden, A. & Samuels, J. (2000). Effects of Causal Text Revisions on More- and Less-Skilled Readers' Comprehension of Easy and Difficult Texts. *Cognition and Instruction*, 18(4), 525–556.
- Lisowski, E. (2023). GPT-3 vs. CHATGPT. the key differences. <https://addepto.com/blog/gpt-3-vs-chat-gpt-the-key-differences/>
- Lissens, F., Asmar, M., Willems, D., Van Damme, J., De Coster, S., Demeestere, E., Maes, R., Baccarne, B., Robaeyst, B., Duthoo, W. & Desoete, A. (2020). Het stopt nooit...De impact van dyslexie en/of dyscalculie op het welbevinden en studeren van (jong)volwassenen en op de transitie naar de arbeidsmarkt: een bundeling van Vlaamse pilootstudies.
- Liu, Q., Kusner, M. J. & Blunsom, P. (2020). A Survey on Contextual Embeddings.

- Malik, R. S. (2022, juli 4). *Top 5 NLP Libraries To Use in Your Projects* (T. Al, Red.). <https://towardsai.net/p/l/top-5-nlp-libraries-to-use-in-your-projects>
- Martens, M., De Wolf, R. & Evens, T. (2021a). *Algoritmes en AI in de onderwijscontext: Een studie naar de perceptie, mening en houding van leerlingen en ouders in Vlaanderen*. Kenniscentrum Data en Maatschappij. Verkregen 30 maart 2022, van <https://data-en-maatschappij.ai/publicaties/survey-onderwijs-2021>
- Martens, M., De Wolf, R. & Evens, T. (2021b, juni 28). *School innovation forum 2021*. Kenniscentrum Data en Maatschappij. Verkregen 1 april 2022, van <https://data-en-maatschappij.ai/nieuws/school-innovation-forum-2021>
- McCombes, S. (2022). How to write A summary: Guide amp; examples. <https://www.scribbr.com/working-with-sources/how-to-summarize/>
- McDonald, R. (2007). A study of global inference algorithms in multi-document summarization. *Advances in Information Retrieval: 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007. Proceedings* 29, 557–564.
- McFarland, A. (2023). What is prompt engineering in AI amp; Why It Matters. <https://www.unite.ai/what-is-prompt-engineering-in-ai-why-it-matters/>
- McKeown, K., Klavans, J. L., Hatzivassiloglou, V., Barzilay, R. & Eskin, E. (1999). Towards multidocument summarization by reformulation: Progress and prospects.
- McNutt, M. (2014). Reproducibility. *Science*, 343(6168), 229–229. <https://doi.org/10.1126/science.1250475>
- Menzli, A. (2023). Tokenization in NLP: Types, challenges, examples, tools. <https://neptune.ai/blog/tokenization-in-nlp>
- Miszczak, P. (2023). Prompt engineering: The ultimate guide 2023 [GPT-3 amp; chat-gpt]. <https://businessolution.org/prompt-engineering/>
- Mottes, C. (2023). GPT-3 vs. Bert: Comparing the two most popular language models. <https://blog.invgate.com/gpt-3-vs-bert>
- Nallapati, R., Zhai, F. & Zhou, B. (2017). SummaRuNNer: A Recurrent Neural Network Based Sequence Model for Extractive Summarization of Documents. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10958>
- Nandhini, K. & Balasundaram, S. (2013). Improving readability through extractive summarization for learners with reading difficulties. *Egyptian Informatics Journal*, 14(3), 195–204.
- Nenkova, A. & Passonneau, R. (2004). Evaluating Content Selection in Summarization: The Pyramid Method. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, 145–152.

- Niemeijer, A., Frederiks, B., Riphagen, I., Legemaate, J., Eefsting, J. & Hertogh, C. (2010). Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. *Psychogeriatrics*, 22(7), 1129–1142. <https://doi.org/10.1017/S1041610210000037>
- OnderwijsVlaanderen. (2023). Voorleessoftware voor Leerlingen met Leesbeperkingen. <https://onderwijs.vlaanderen.be/voorleessoftware-voor-leerlingen-met-leesbeperkingen>
- Paetzold, G. & Specia, L. (2016). SemEval 2016 Task 11: Complex Word Identification. *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 560–569. <https://doi.org/10.18653/v1/S16-1085>
- Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C. & Thompson, W. H. (2017). Research: The readability of scientific texts is decreasing over time (S. King, Red.). *eLife*, 6, e27725.
- Poel, M., Boschman, E. & op den Akker, R. (2008). A Neural Network Based Dutch Part of Speech Tagger [<http://eprints.ewi.utwente.nl/14662>; 20th Benelux Conference on Artificial Intelligence, BNAIC 2008, BNAIC ; Conference date: 30-10-2008 Through 31-10-2008]. In A. Nijholt, M. Pantic, M. Poel & H. Hondorp (Red.), *BNAIC 2008* (pp. 217–224). Twente University Press (TUP).
- Premjith, P., John, A. & Wilscy, M. (2015). Metaheuristic Optimization Using Sentence Level Semantics for Extractive Document Summarization, 347–358. https://doi.org/10.1007/978-3-319-26832-3_33
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. e.a. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Rani, R. & Kaur, B. (2021). The TEXT SUMMARIZATION AND ITS EVALUATION TECHNIQUE. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(1), 745–752.
- Readable. (2021). *Flesch Reading Ease and the Flesch Kincaid Grade Level*. <https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/>
- Rello, L. & Baeza-Yates, R. (2013). Good fonts for dyslexia. *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS 2013*.
- Rello, L., Baeza-Yates, R., Dempere-Marco, L. & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia.
- Rello, L., Baeza-Yates, R. & Saggion, H. (2013). The Impact of Lexical Simplification by Verbal Paraphrases for People with and without Dyslexia. 7817, 501–512.
- Rello, L. & Baeza-Yates, R. A. (2015). How to present more readable text for people with dyslexia. *Universal Access in the Information Society*, 16, 29–49.

- Rello, L. & Bigham, J. (2017). Good Background Colors for Readers: A Study of People with and without Dyslexia, 72–80.
- Rello, L., Kanvinde, G. & Baeza-Yates, R. (2012). Layout Guidelines for Web Text and a Web Service to Improve Accessibility for Dyslexics. *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*.
- Ribas, J. (2023). Building the new bing. <https://www.linkedin.com/pulse/building-new-bing-jordi-ribas/>
- Ribeiro, E., Ribeiro, R. & de Matos, D. M. (2018). A Study on Dialog Act Recognition using Character-Level Tokenization.
- Rijkhoff, J. (2022). Tekst Inkorten?: 9 tips om Je Teksten korter Te Maken. <https://dialoogtrainers.nl/tekst-inkorten-tips/>
- Rivero-Contreras, M., Engelhardt, P. E. & Saldaña, D. (2021). An experimental eye-tracking study of text adaptation for readers with dyslexia: effects of visual support and word frequency. *Annals of Dyslexia*, 71, 170–187.
- Roldós, I. (2020, december 22). *Major Challenges of Natural Language Processing (NLP)*. MonkeyLearn. Verkregen 1 april 2022, van <https://monkeylearn.com/blog/natural-language-processing-challenges/>
- Roose, K. (2023). Don't ban chatgpt in schools. teach with it. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>
- Sciforce. (2020, februari 4). *Biggest Open Problems in Natural Language Processing*. Verkregen 1 april 2022, van <https://medium.com/sciforce/biggest-open-problems-in-natural-language-processing-7eb101ccfc9>
- Shardlow, M. (2013). A Comparison of Techniques to Automatically Identify Complex Words. *51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop*, 103–109. <https://aclanthology.org/P13-3015>
- Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications(IJACSA), Special Issue on Natural Language Processing 2014*, 4(1). <https://doi.org/10.14569/SpecialIssue.2014.040109>
- Siddharthan, A. (2006). Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4(1), 77–109. <http://oro.open.ac.uk/58888/>
- Siddharthan, A. (2014). A survey of research on text simplification. *ITL - International Journal of Applied Linguistics*, 165, 259–298.
- Sikka, P. & Mago, V. (2020). A Survey on Text Simplification. *CoRR*, abs/2008.08612. <https://arxiv.org/abs/2008.08612>
- Simon, J. (2021). Large language models: A new moore's law? <https://huggingface.co/blog/large-language-models>

- Snow, C. (2010). Academic Language and the Challenge of Reading for Learning About Science. *Science (New York, N.Y.)*, 328, 450–2.
- Sohom, G., Ghosh; Dwight. (2019). *Natural Language Processing Fundamentals*. Packt Publishing. <https://medium.com/analytics-vidhya/natural-language-processing-basic-concepts-a3c7f50bf5d3>
- Strubell, E., Ganesh, A. & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP.
- Suleiman, D. & Awajan, A. (2020). Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges. *Mathematical Problems in Engineering*, 2020.
- Suter, J., Ebling, S. & Volk, M. (2016). Rule-based Automatic Text Simplification for German.
- Swayamdipta, S. (2019, januari 22). *Learning Challenges in Natural Language Processing*. Verkregen 1 april 2022, van <https://www.microsoft.com/en-us/research/video/learning-challenges-in-natural-language-processing/>
- Tanya Goyal, G. D., Junyi Jessy Li. (2022). News Summarization and Evaluation in the Era of GPT-3. *arXiv preprint*.
- Thangarajah, V. (2019). Python current trend applications-an overview.
- Tops, W., Callens, M., Brysbaert, M. & Schouten, E. L. (2018). *Slagen met Dyslexie in Het Hoger Onderwijs*. Owl Press.
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E. & Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models.
- van der Meer, C. (2022). Dyslexie hebben is Niet Zo Raar: Lezen is iets heel onnatuurlijks. <https://www.demorgen.be/beter-leven/dyslexie-hebben-is-niet-zo-raar-lezen-is-iets-heel-onnatuurlijks~bc608101/>
- Vasista, K. (2022). Evolution of AI Design Models. *Central Asian Journal of Theoretical and Applied Science*, 3(3), 1–4.
- Verhoeven, W. (2023, februari 8). *Applaus voor de studenten die ChatGPT gebruiken* (Trends, Red.). https://trends.knack.be/economie/bedrijven/applaus-voor-de-studenten-die-chatgpt-gebruiken/article-opinion-1934277.html?cookie_check=1676034368
- Verma, P. & Verma, A. (2020). A review on text summarization techniques. *Journal of scientific research*, 64(1), 251–257.
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. & Schmidt, D. C. (2023). A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT.

- Xu, W., Callison-Burch, C. & Napoles, C. (2015). Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3, 283–297.
- Zeng, Q., Kim, E., Crowell, J. & Tse, T. (2005). A Text Corpora-Based Estimation of the Familiarity of Health Terminology. In J. L. "Oliveira, V. Maojo, F. Martín-Sánchez & A. S. Pereira (Red.), *Biological and Medical Data Analysis* (pp. 184–192). Springer Berlin Heidelberg.
- Zhang, J., Zhao, Y., Saleh, M. & Liu, P. (2020). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *International Conference on Machine Learning*, 11328–11339.
- Zhang, M., Riecke, L. & Bonte, M. (2021). Neurophysiological tracking of speech-structure learning in typical and dyslexic readers. *Neuropsychologia*, 158, 107889.