

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

Dylan Cluyse

dylan.cluyse@student.hogent.be

Promotor: Lena De Mol

Co-promotor: Johan Decorte (Hogeschool Gent); Jana Van Damme (Hogeschool Gent)

Hogeschool Gent, Valentin Vaerwyckweg 1, 9000 Gent

Samenvatting

Ingewikkelde woordenschat en zinsbouw hinderen scholieren met dyslexie in de derde graad van het middelbaar onderwijs bij het begrijpend lezen van wetenschappelijke artikelen. Gepersonaliseerde manual text simplification (ATS) 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 geautomatiseerde en gepersonaliseerde tekstvereenvoudiging. Hiervoor stelt het onderzoek de volgende onderzoeksvraag op: "Hoe kan een wetenschappelijk artikel automatisch worden vereenvoudigd, gericht op de unieke noden van scholieren met dyslexie in de derde graad middelbaar onderwijs?". Een requirementsanalyse achterhaalt de benodigde functionaliteiten om gepersonaliseerde en geautomatiseerde tekstvereenvoudiging mogelijk te maken. Vervolgens wijst de vergelijkende studie uit welk taalmodel ontwikkelaars kunnen inzetten om de taak van gepersonaliseerde en geautomatiseerde tekstvereenvoudiging mogelijk te maken. De requirementsanalyse wijst uit 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 de derde graad middelbaar onderwijs. Toepassingen voor gepersonaliseerde automated text simplification zijn mogelijk, maar ontwikkelaars moeten meer inzetten op de unieke noden van deze scholieren.

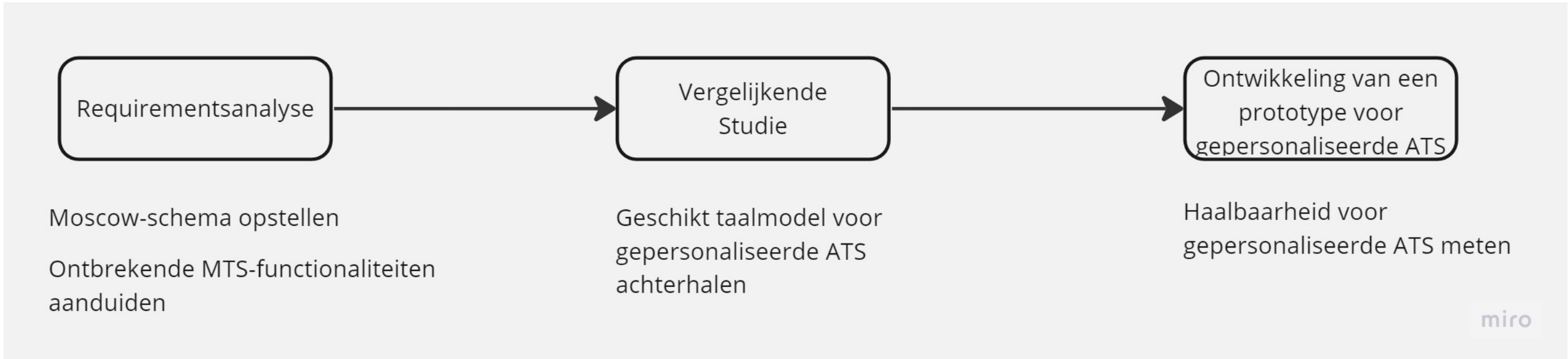
Keuzerichting: AI en data engineering

Sleutelwoorden: Gepersonaliseerde tekstvereenvoudiging; dyslexie; natuurlijke taalverwerking

Broncode: <https://github.com/dylancluyse/bachelorproef-nlp-tekstvereenvoudiging>

1. Introductie

Vindt u wetenschappelijke artikelen vermoeiend om te lezen? Storen het vakjargon, het compacte formaat u bij het begrijpend lezen van nieuw wetenschappelijk onderzoek? Dit is niet abnormaal, want het begrijpend lezen van wetenschappelijke artikelen vraagt een alsmaar grotere geletterdheid van de lezer. Alle doelgroepen krijgen het moeilijk te verduren, maar scholieren met dyslexie ervaren hier een nog grotere hindernis mee. Toch kunnen leerkrachten deze hindernis voor de studenten verwerpen, door teksten met *manual text simplification* (MTS) te vereenvoudigen. Om de werkdruk in het onderwijs tegen te gaan, moet dit proces ook geautomatiseerd worden met *automated text simplification* (ATS). Om een werkwijze te achterhalen voor het ontwikkelen van dergelijk toepassing, voert het onderzoek drie fasen uit weergegeven in figuur 1.



Figuur 1

2. Onderzoeksmethoden

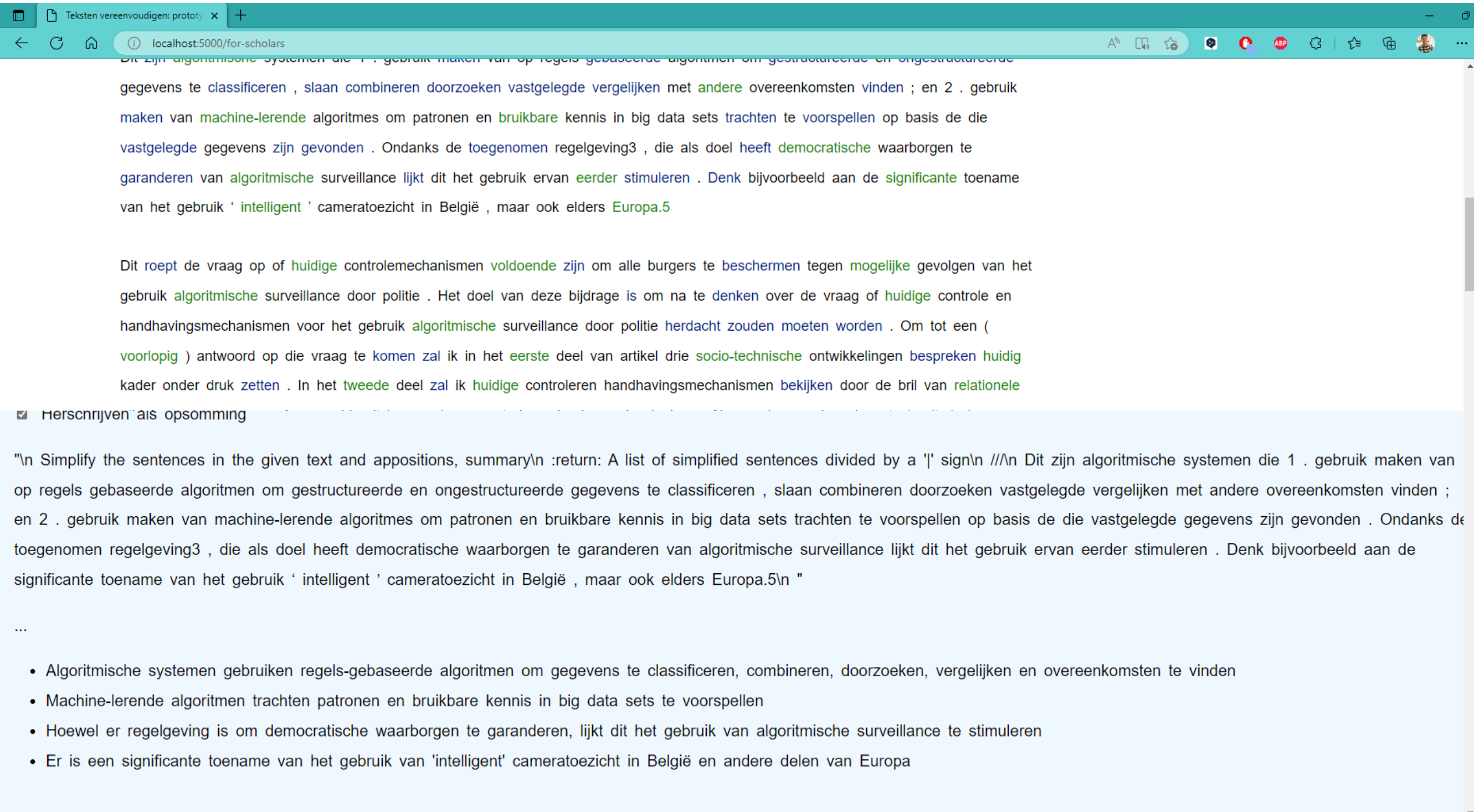
Het onderzoek biedt een antwoord op de volgende drie onderzoeksvragen:

Allereerst achterhaalt het onderzoek met een requirementsanalyse de functionaliteiten voor gepersonaliseerde ATS bij huidige tools en toepassingen. Het resultaat van deze fase is een Moscow-schema met alle *must-have* functionaliteiten tot alle *wont-have* functionaliteiten. Dit beantwoordt de volgende onderzoeksvraag: Welke functies ontbreken AI-toepassingen om geautomatiseerde tekstvereenvoudiging mogelijk te maken voor scholieren met dyslexie in de derde graad middelbaar onderwijs?

Vervolgens moet het prototype beschikken over een geschikt taalmodel voor gepersonaliseerde ATS. Na de requirementsanalyse volgt een vergelijkende studie rond beschikbare taalmodellen om gepersonaliseerde ATS te kunnen realiseren. Het doel van deze onderzoeksmethode is om de beschikbare taalmodellen op HuggingFace en het GPT-3 taalmodel tegenover elkaar te plaatsen. Deze fase beantwoordt de volgende onderzoeksvraag: Welk taalmodel of LLM is geschikt voor de ATS van wetenschappelijke artikelen voor scholieren met dyslexie in de derde graad van het middelbaar onderwijs, met dezelfde of gelijkaardige kwaliteiten als gepersonaliseerde MTS?

Tot slot volgt de ontwikkeling van het prototype. Het prototype maakt gebruik van vrij beschikbare technologieën en programmeertalen, zoals HTML/CSS en Python. Het prototype bestaat uit twee componenten. Enerzijds moet het prototype een ondersteunend middel bieden aan scholieren met dyslexie, door een wetenschappelijk artikel in een aanpasbaar formaat aan te bieden. Daarnaast kunnen

scholieren aanpassingen aan de tekst maken, zoals weergegeven in figuur 2. Dit beantwoordt de volgende onderzoeksvraag: Wat zijn de nodige stappen bij de ontwikkeling van een intuïtieve lokale webtoepassing die zowel scholieren met dyslexie als leerkrachten helpt?



Figuur 2: Voorbeeldweergave van het scholierencomponent. De scholier selecteert een tekst en vraagt voor een opsomming (met eenvoudige woordenschat) van de gemarkeerde tekst.

3. Conclusies

Uit de requirementsanalyse blijkt dat bestaande online tools en softwaretoepassingen voor ATS onvoldoende gepersonaliseerde functionaliteiten bieden. Ze missen bijvoorbeeld opmaakoptyes voor scholieren met dyslexie en de mogelijkheid om wetenschappelijke artikelen in te laden. Recente technologieën bieden wel tekstvereenvoudigingsopties, maar vereisen informaticakennis die de meeste scholieren en leraren niet hebben. Een vergelijkende studie toont aan dat het GPT-3 taalmodel geschikt is voor LS en SS-technieken, terwijl andere taalmodellen minder coherente tekst genereren. Het prototype voor gepersonaliseerde ATS maakt gebruik van AI en NLP-technologieën zoals PDFMiner, de GPT-3 API en Pandoc. Hoewel het prototype nog niet aan alle functionaliteiten voldoet, kunnen ontwikkelaars met de gebruikte softwarepakketten een volledig afgewerkte toepassing ontwikkelen.

4. Toekomstig onderzoek

Verder onderzoek naar de toepassing van AI via een API, zoals het GPT-4 taalmodel, is noodzakelijk en kan baanbrekend zijn voor het onderwijs. Dit onderzoek kan zich richten op doelgroepsinschattingen en onderzoek naar het potentieel van de combinatie van GPT-3 en full-text-simplification. Daarnaast is er

behoefte aan onderzoek naar de verschillen tussen taalmodellen getraind op wetenschappelijke artikelen en taalmodellen getraind op algemene data. Tot slot kunnen logopedisten en studenten in deze richting het prototype gebruiken om

verder onderzoek uit te voeren naar de effecten van gepersonaliseerde ATS bij scholieren met dyslexie in de derde graad van het middelbaar onderwijs.