

Handleiding extra controle

Welkom!

Fijn dat je mee wil werken aan het controleren van de gegevens voor het project WoordWaark: Geschreven Gronings! De boeken die je tegenkomt in dit programma hebben een extra controle nodig om ervoor te zorgen dat er zo min mogelijk fouten in ons corpus terechtkomen en dat de metadata helemaal klopt. In deze handleiding staan de stappen beschreven die je doorloopt om de controle uit te voeren.

1. Beginnen

Ga in je browser naar <https://woordwaark.housing.rug.nl/grunnegsextra/#login> en log daar in met de gebruikersnaam woordwaark en het wachtwoord w00rdwaark. Je komt nu in het menu 'Selecteer bron' terecht, waar je een tekst kan selecteren om mee te beginnen. Aan de teksten die onder 'te controleren' staan wordt nog niet door iemand anders gewerkt, dus je kan elke tekst selecteren. Aan de teksten die onder ' bezig met controle' staan wordt al wel gewerkt, dus selecteer alleen een tekst uit deze lijst wanneer je zelf degene was die de tekst eerder had geopend. Klik op de tekst die je wil bewerken en daarna op 'Voer controle uit' om deze te openen.

2. Het programma

Je ziet nu twee vensters: links de scan en rechts de tekst die er door automatische tekenherkenning uit is gehaald. Bovenaan in het rechter venster zie je een aantal regels die beginnen met een dollarteken (\$): dit is de metadata over het gehele werk. Als het goed is zie je ook een aantal regels die beginnen met een kruisje (#): dit is de metadata over de individuele artikelen in een werk. Soms ontbreekt deze informatie. In dat geval moet je die zelf toevoegen (zie sectie 3.3). Onder de metadata staat de echte tekst van het boek, waar je ook dingen in weg kan halen, toe kan voegen of kan wijzigen. Onder het linker venster staan de knoppen 'Bron', waarmee je de metadata over het gehele werk aanpast, en 'Artikel', waarmee je de metadata over individuele artikelen aanpast.

3. Controle van metadata

Klik om de metadata van het gehele werk te controleren op de knop 'Bron'. Controleer of de informatie in de velden klopt, en pas zo nodig wijzigingen toe of vul de informatie aan. Je kan deze informatie het beste vinden door in het linker venster naar de titelpagina en de eerste paar bladzijden te kijken. Het menu dat verschijnt is met de verschillende velden kan je verslepen zodat je beide dingen tegelijk kan zien.

3.1. Namen van auteurs/samenstellers

Schrijf de namen van de auteurs zoveel mogelijk in hun geheel op in het volgende format: voornaam/voornamen, tussenvoegsel(s), achternaam. Dus 'Fenna Slapsma-Tiessens', 'Kornelis ter Laan' etc. Als er meerdere auteurs zijn, scheid je hun namen met een komma.

3.2. Overige velden bij Bron

Vul de velden zo veel mogelijk in. De velden *deel/volume*, *editie/druk*, *website* en *opmerkingen* mogen leeg blijven, maar probeer voor de rest zo veel mogelijk informatie te

vinden (eventueel buiten het bestand door op het internet te zoeken). Als je iets niet kan vinden, vul dan ‘onbekend’ in. Als het werk een ondertitel heeft, neem die dan ook mee in het veld *titel*. Als je alle velden hebt ingevuld, druk je onderaan op OK. De informatie wordt nu automatisch ingevoegd in het rechter venster bij de regels die beginnen met \$.

3.3. Metadata per artikel

Voor elk boek moet er ook informatie over de individuele artikelen worden toegevoegd.

BELANGRIJK: Dit geldt zowel voor boeken van één auteur, bijvoorbeeld een roman, als voor boeken die bestaan uit teksten van verschillende schrijvers. Als de volledige tekst van een boek door één auteur is geschreven, hoeft je de metadata van het artikel maar één keer in te vullen en recht onder de metadata van het volledige werk te plakken. Als het boek bijvoorbeeld bestaat uit korte verhalen en gedichten van verschillende auteurs, vul je per tekst de metadata in en plak je die boven het deel van de tekst waar het om gaat.

Klik om de metadata in te vullen op ‘Artikel’ en vul de velden in. Hier gelden dezelfde conventies voor namen als in 3.1. Soms is het niet duidelijk wie bijvoorbeeld een voorwoord heeft geschreven bij een boek. Als er wel iets genoemd wordt als ‘jury van wedstrijd X’ of de naam van een commissie/vereniging kan je die invullen. Als er niks bekend is, vul je ‘onbekend’ in. Als een voorwoord of inleiding geen titel heeft, vul dan in [voorwoord]; als er een nawoord is zonder titel vul je in [nawoord]; als het gaat om de achterkant van een boek vul je in [achterkant boek] (allemaal inclusief de vierkante haken). Als er gedichten zonder titel zijn, vul je de eerste regel van het gedicht in tussen vierkante haken, bijvoorbeeld [t Schemert en dreum het mie verloaten]. Als je alle velden in hebt gevuld, klik je weer op OK (hierdoor worden de velden gekopieerd), en navigeer je in het rechter venster naar het artikel waar de metadata bij hoort. Plak de metadata voor de tekst waar het om gaat door op ctrl+v te drukken op het toetsenbord of door met de rechter muisknop in het venster te klikken en op ‘plakken’ te drukken.

3.4. Opmerkingen

Zowel bij Bron als bij Artikel is er ruimte voor opmerkingen. Gebruik dit veld voor extra informatie waarvan je denkt dat die relevant kan zijn, maar die niet in één van de andere velden hoort. Een voorbeeld is de leeftijd en/of locatie van deelnemers aan een schrijfwedstrijd als die vermeld wordt. Je kan dit veld ook gebruiken wanneer er iets raars aan de tekst opvalt, maar als je een vraag hebt waar je antwoord op wil kan je beter contact opnemen met Hedwig.

Het kan zijn dat alle velden al goed ingevuld zijn, maar dit zal nogal verschillen per werk.

4. Tekstcontrole

Voor het corpus hebben we zinnen nodig die (1) Groningstalig en (2) volledig zijn. Dat betekent dus dat alle tekst die niet Groningstalig is, maar bijvoorbeeld Nederlands, Duits of een andere variant van het Nedersaksisch, weggehaald moet worden. Daarnaast willen we geen zinsfragmenten hebben, zoals de titels van werken, inhoudsopgaven, paginanummers, woordenlijsten en andere informatie die niet in volledige zinnen staat.

4.1. Tekst weghalen

Scrol door het rechter venster om te controleren of er nog tekst staat die weg moet. Houd hierbij dus in gedachte dat we alleen volledige Groningstalige zinnen willen hebben. Soms herkent het programma voor tekstherkenning iets wat geen tekst is (maar bijvoorbeeld een afbeelding of een vuiltje op het papier) als tekst, wat resulteert in onzin. Dit soort tekst, die er vaak uitziet als @#\$^#\$&\$% moet ook weg. Let vooral goed op dat er geen titels van verhalen of gedichten blijven staan: die horen in de metadata en niet in de tekst zelf.

4.2. Extra controle

Sommige boeken bevatten veel kleine fouten die tijdens het herkennen van de tekst gemaakt zijn, zoals de verwisseling van de letters 'l' en 't' of de letters 'e' en 'c' omdat deze in sommige lettertypen veel op elkaar lijken. Scan of lees een paar stukjes van de tekst om te kijken of dat bij dit boek het geval is. Je mag deze fouten corrigeren als je ze tegenkomt, maar dit kan erg veel tijd in beslag nemen. Als een boek erg veel van dit soort fouten bevat, neem dan contact op door bovenin de pagina te klikken op 'Contact' en het formulier verder in te vullen. Wij kijken dan of het de moeite waard is dat iemand het volledige boek controleert of dat dit teveel werk wordt en de uren beter ergens anders in kunnen worden gestoken.

BELANGRIJK: we nemen alle tekst precies zo over als die in het boek geschreven staat, dus als je een woord tegenkomt met een spelfout die ook in de scan aanwezig is, moet je deze spelfout laten staan. We corrigeren dus alleen tekst als die niet correspondeert met de tekst in de scan. Met ctrl+f kan je in beide vensters zoeken, wat het eenvoudig maakt om de juiste plek in de scan te vinden wanneer je de teksten wil vergelijken.

5. Opslaan en afsluiten

In principe worden al je wijzigingen automatisch opgeslagen en komt de tekst in ' bezig met controle' te staan in het Selecteer bron menu zodra je een wijziging gedaan hebt. Als je helemaal klaar bent met de tekst ga je bovenin naar het menu Controle afronden en druk je op Sla resultaten definitief op. Doe dit alleen wanneer je zeker weet dat je helemaal klaar bent, want daarna komt de tekst bij 'goedgekeurd' te staan in het Selecteer bron menu en kan je er niet meer bij.

Als je vragen hebt, aarzel dan niet om een mailtje te sturen naar Hedwig (h.g.sekeres@rug.nl) of te bellen/Whatsappen (06-23123458).

Veel succes en plezier met het nakijken!