**Woordenbank van de Nederlandse Dialecten - Import-handleiding**

**Stap 1:**

Installeer de macro's in word (via Bart De La Ruelle).

**Stap 2:**

Open het word-document en kijk dit na:

1. Allereerst krijgt het word-document een titel. Dit is doorgaans gewoon de letter. Zorg dat de titel niet vet en niet cursief is, anders zullen er flink wat <lemma>-foutmeldingen optreden bij validatie van het txt-bestand.
2. Daarna moeten de woordenboekartikelen volgens de volgende opmaak aangepast worden:

**Trefwoord: vet**

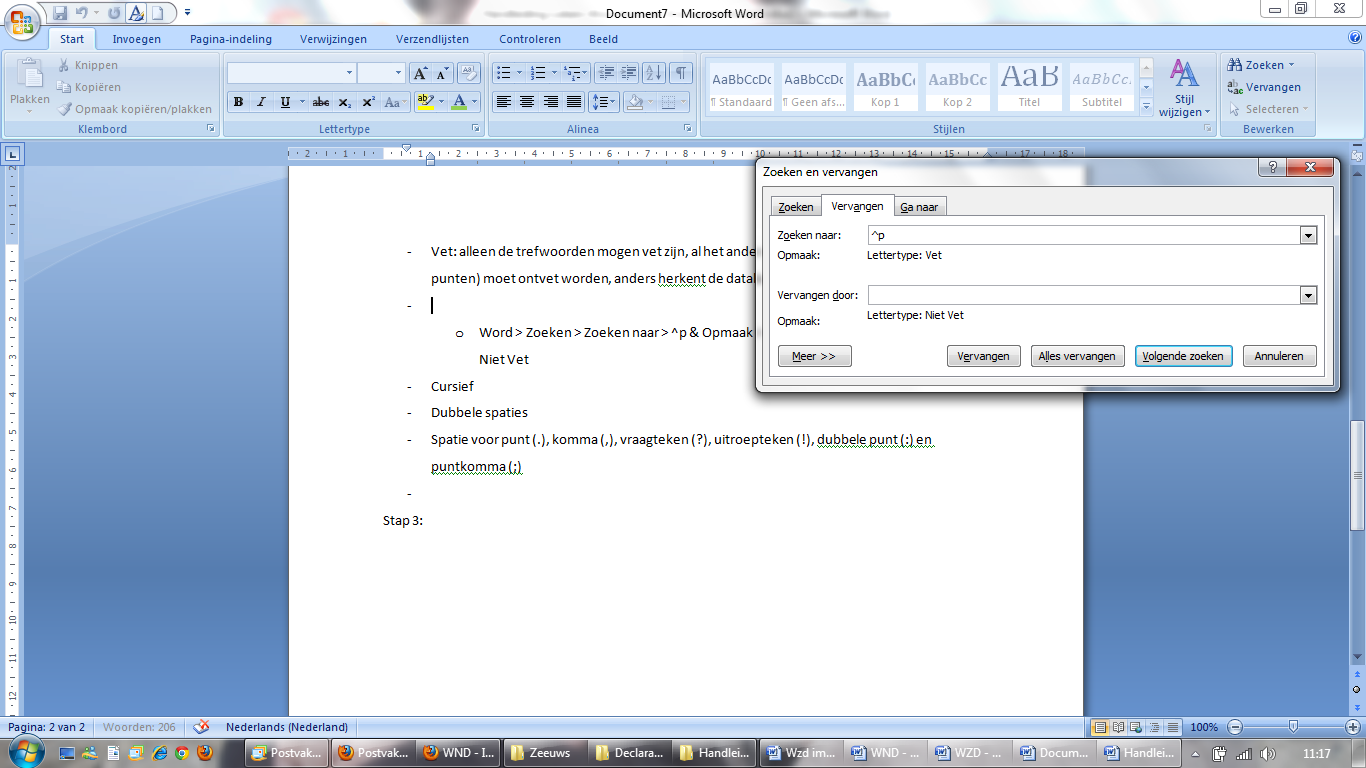
Uitleg: gewoon

*Voorbeelden: cursief*

Na elk woordenboekartikel/lemma moeten twee enters (harde returns) staan, zodat er 1 regel witruimte tussen de verschillende woordenboekartikelen is. Als een woordenboekartikel de allerlaatste regel van een pagina vult, moet de allereerste regel op de volgende nieuwe pagina wit gelaten worden. Anders denkt de database bij het importeren dat de twee lemma’s aan elkaar horen. Helemaal op het einde van het document moeten achter het laatste lemma 2 regels witruimte toegevoegd worden, anders wordt het laatste lemma niet geïmporteerd.

Per woordenboek zijn er andere probleemgevallen. Hieronder een aantal aandachtspunten:

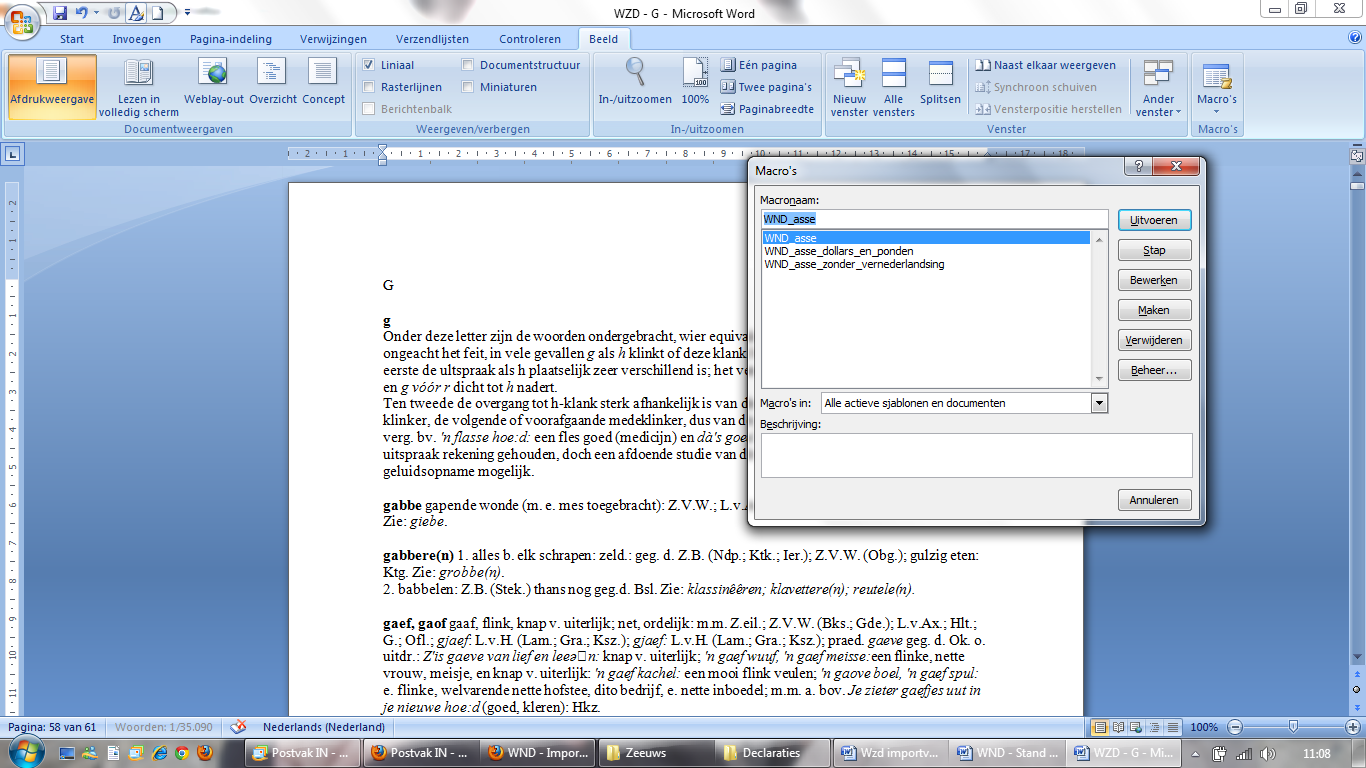
* Vet: controle op vette returns via Zoeken en vervangen (Ctrl+H): Alineamarkering (= ^p). Ook checken of er zich geen vette spaties, komma's, punten, etc. bevinden in de rest van het woordenboekartikel (enkel het lemma zelf mag vet staan). Deze moeten anders ontvet worden, anders wordt dit door de database als aparte lemmata herkend.



* Cursief: Zie bovenstaande figuur en verwijder ook alle cursieve alineamarkeringen, om te vermijden dat regeleindes cursief staan (dit zorgt nogal voor importproblemen).
* Dubbele spaties weghalen.
* Spaties voor punt (.), komma (,), vraagteken (?), uitroepteken (!), dubbele punt (:) en puntkomma (;) weghalen.
* Geen spaties rondom de dollar- en pondtekens ($$$$ ££££)
  + FOUT: $$ aaaa $$ ££ bbbb ££
  + GOED: $$aaaa$$ ££bbbb££
  + Na het laatste pondteken mogen geen spaties staan, hierna moet direct een enter komen.
  + Het toevoegen van de dollar- en pondtekens kan automatisch (indien deze nog niet in het document aangebracht zijn): Zoeken > Vervangen > Zoeken naar: ^p^p > Vervangen door: ^p$$$$ ££££^p^p > Opmaak: Lettertype: Niet Cursief door 2 maal op Ctrl + i te klikken > Alles vervangen.

**Stap 3:**

Laat de macro los over het word-bestand: Beeld > Macro's > Uitvoeren:

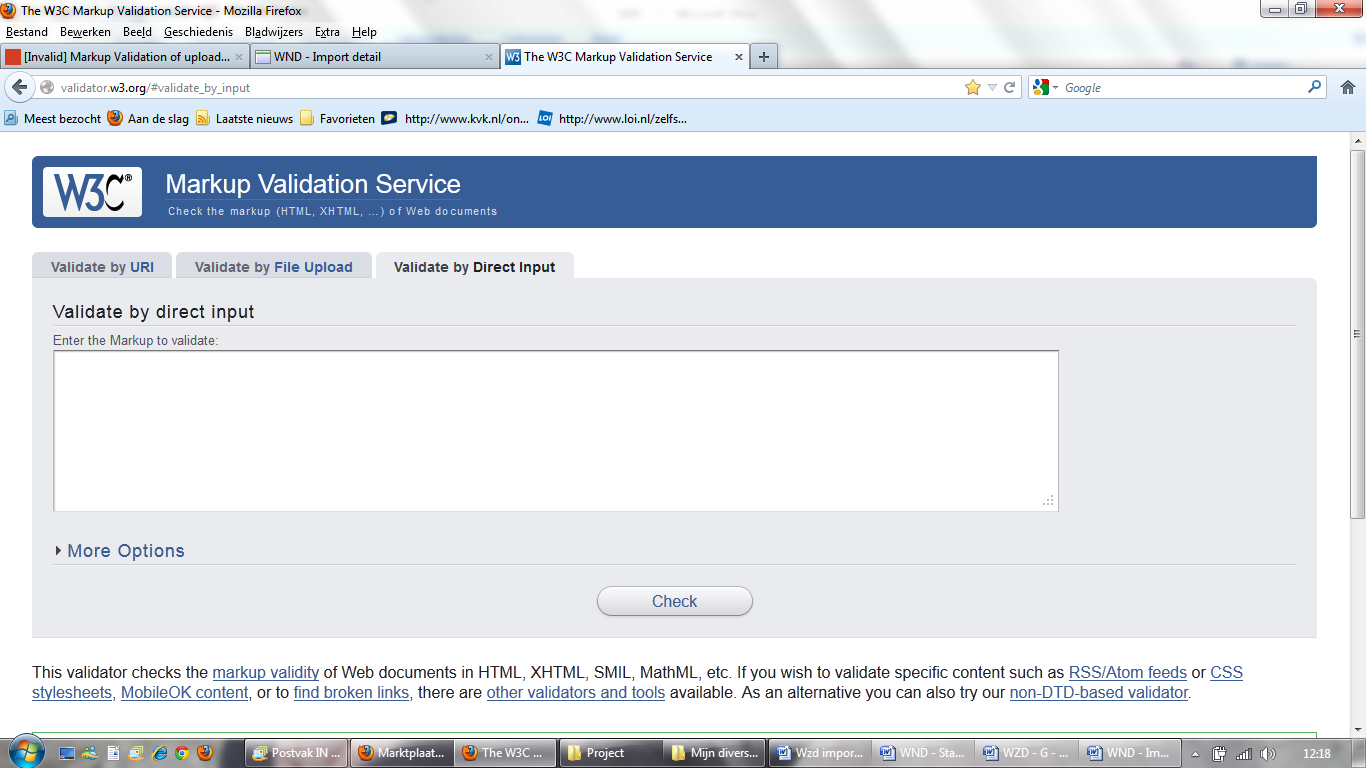


* WND\_asse is voor documenten zonder toegevoegde vernederlandsing en vertaling (zonder $$$$ ££££).
* WND\_asse\_dollars\_en\_ponden is voor documenten met toegevoegde vernederlandsing en vertaling (met $$$$ ££££).

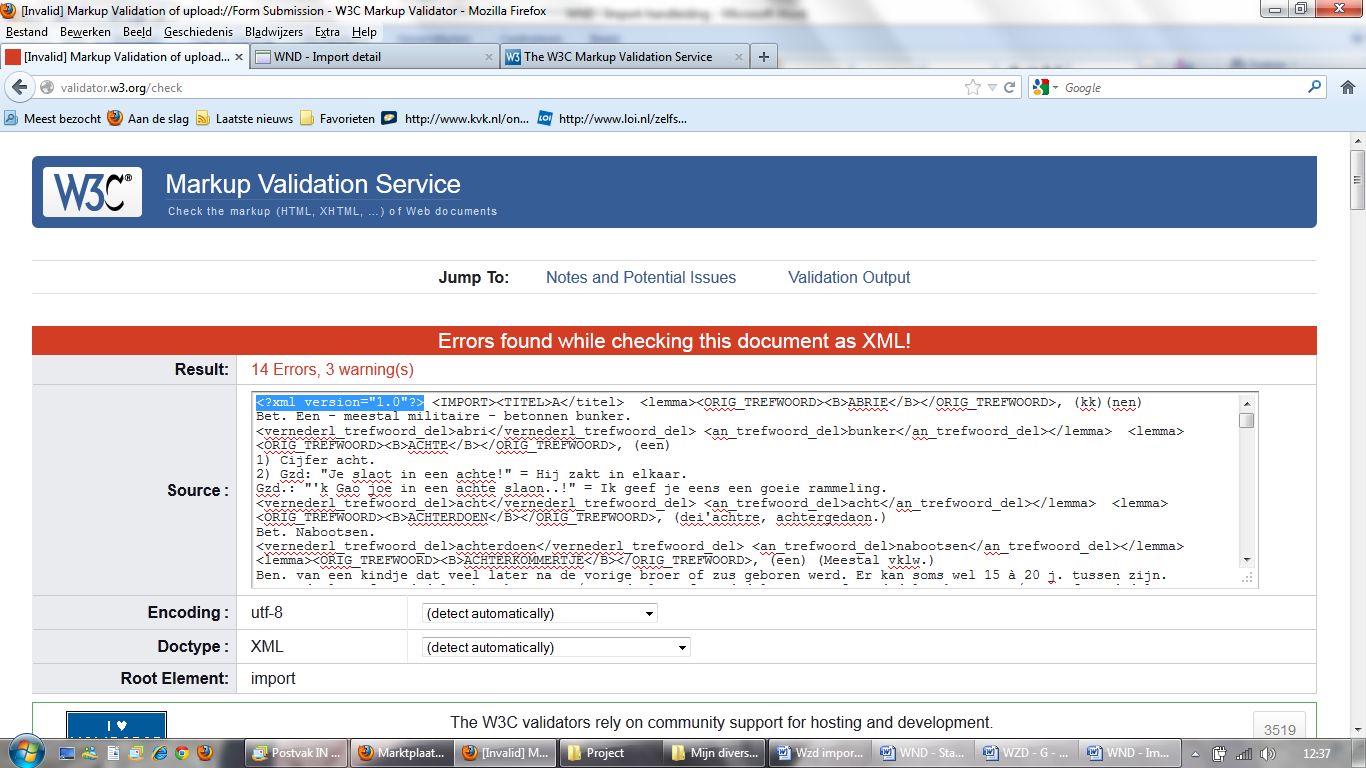
**Stap 4:**

Laat het txt-bestand controleren door de validator:

* Ga naar <http://validator.w3.org/#validate_by_input>.



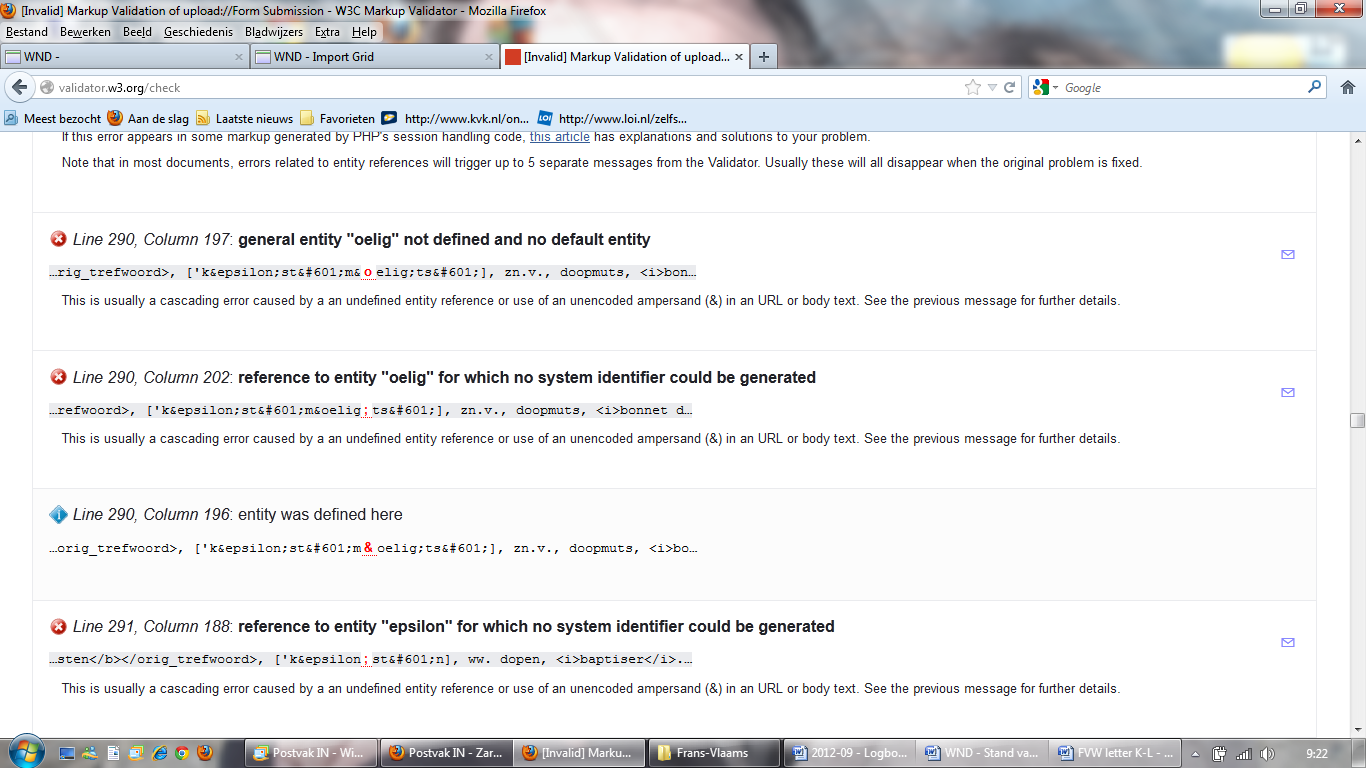
* De meeste (eventuele) fouten treden op in verband met cursivering (“end tag for element “I” which is not open”), maar zouden achterwege moeten blijven indien de cursieve returns zijn weggewerkt via de Zoeken-en-vervangenfunctie van hierboven. Ook het gebruik van diakritische tekens kan voor ongeregeldheden zorgen. Deze worden dan als fout aangegeven, maar kunnen genegeerd worden.
* **TIP**: veel foutmeldingen (die vaak niet direct te zien zijn en eerder op de achtergrond van word lopen) zijn weg te werken door simpelweg alle lemmata te selecteren en in een nieuw word-bestand te plakken.
* Plak eerst de regel **<?xml version="1.0"?>** in de validator. Kopieer daarna de tekst van het txt-bestand en plak deze in de validator achter de vorige code. Druk op 'Check' en dan verschijnen de eventuele 'Errors'.



* Scroll naar beneden om de foutmeldingen te bekijken. Werk deze foutmeldingen weg in het txt-bestand. De rode > geven de fouten weer.



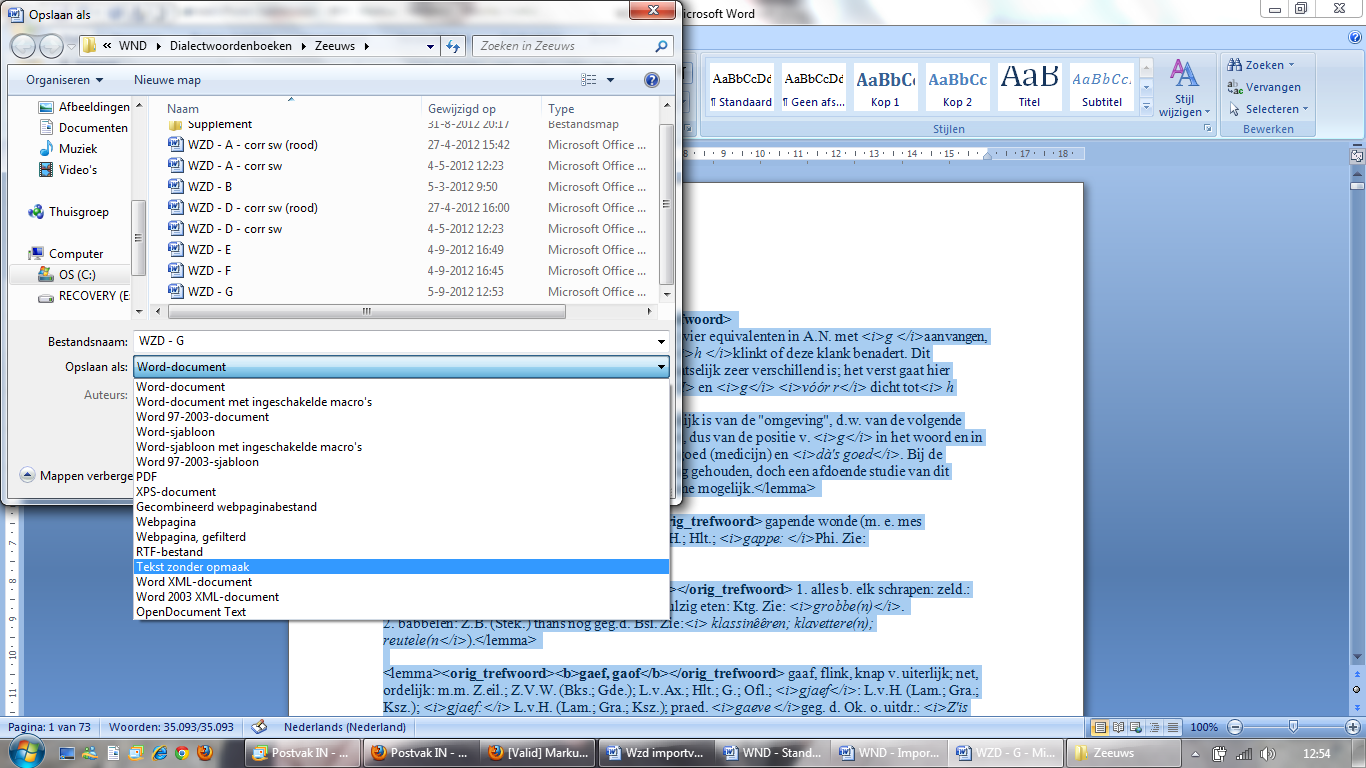
* Een aantal fouten kan niet weggewerkt worden. Deze fouten hebben meestal te maken met diakritische tekens. Ze worden hieronder in het rood met een o, ; en & weergegeven (andere tekens, zoals e of c, zijn ook mogelijk). Negeer deze foutmeldingen.



* Laat na het aanpassen het txt-bestand nog een keer controleren door de validator, totdat er geen foutmeldingen meer verschijnen ('3 warnings' blijft altijd staan, dat kan genegeerd worden).

**Stap 5:**

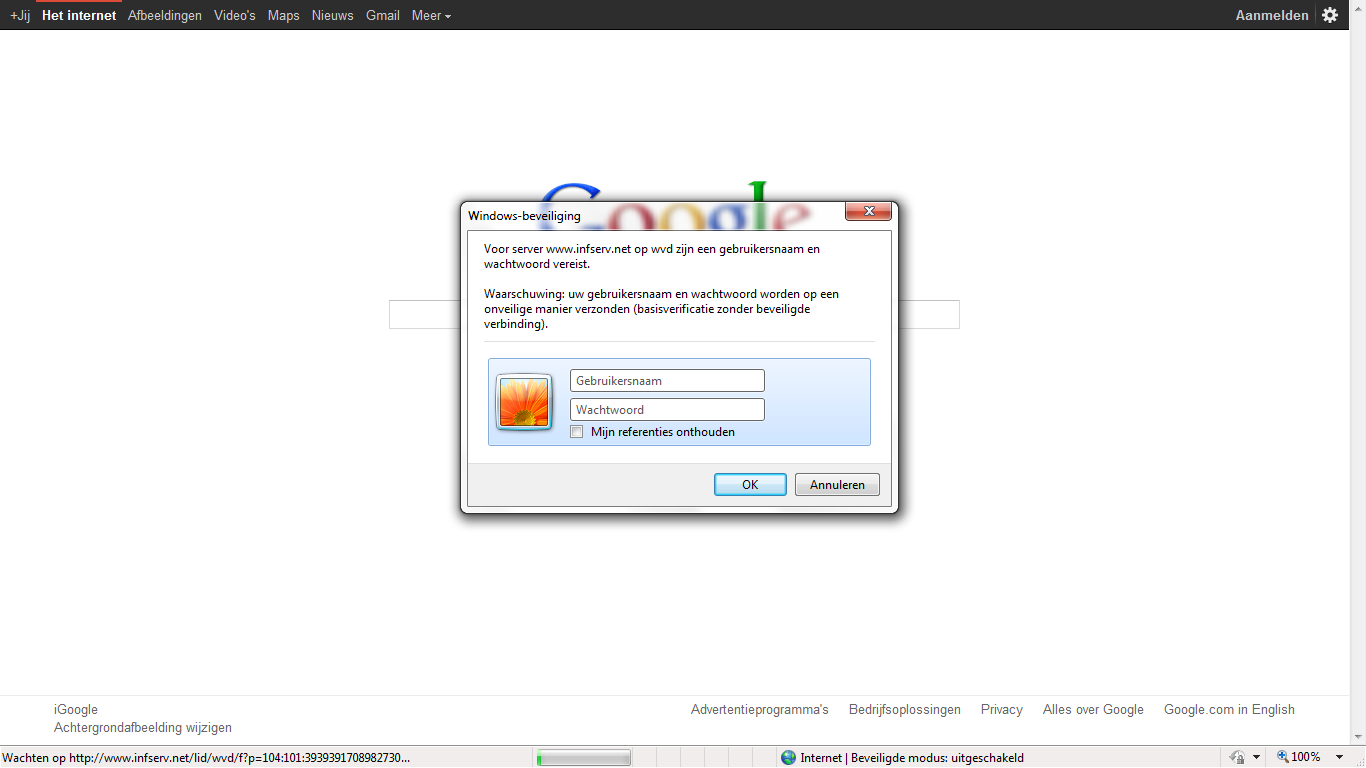
Sla het word-bestand met de macro op als een txt-file (Tekst zonder opmaak). Zorg ervoor dat het originele word-bestand zonder macro niet overgeschreven wordt!



**Stap 6:**

Importeer de letter in de database:

* Ga naar [www.infserv.net/lid/wvd/f?p=104](http://www.infserv.net/lid/wvd/f?p=104) en vul twee keer de inloggegevens in (Gebruikersnaam: vrijwilliger, Wachtwoord/Paswoord: vrijwilliger).

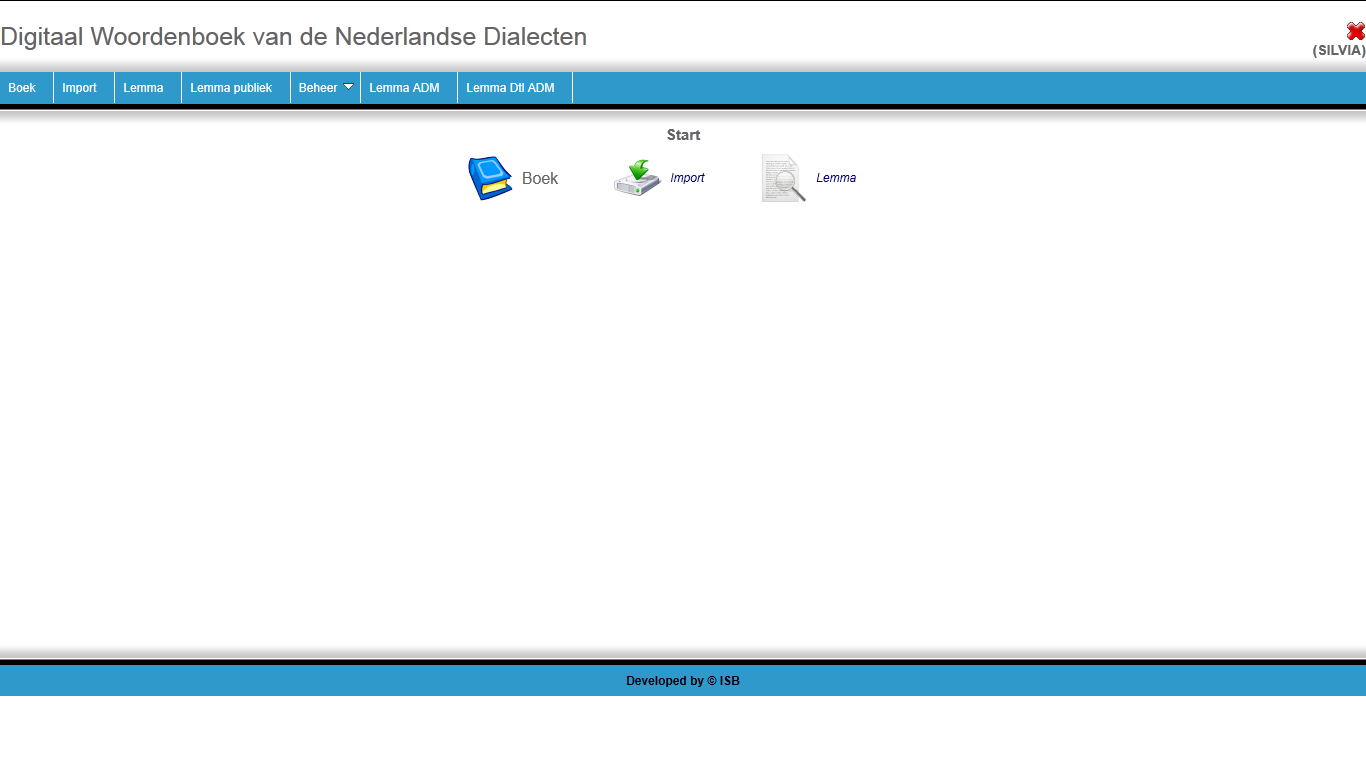
****

**1e inlog**

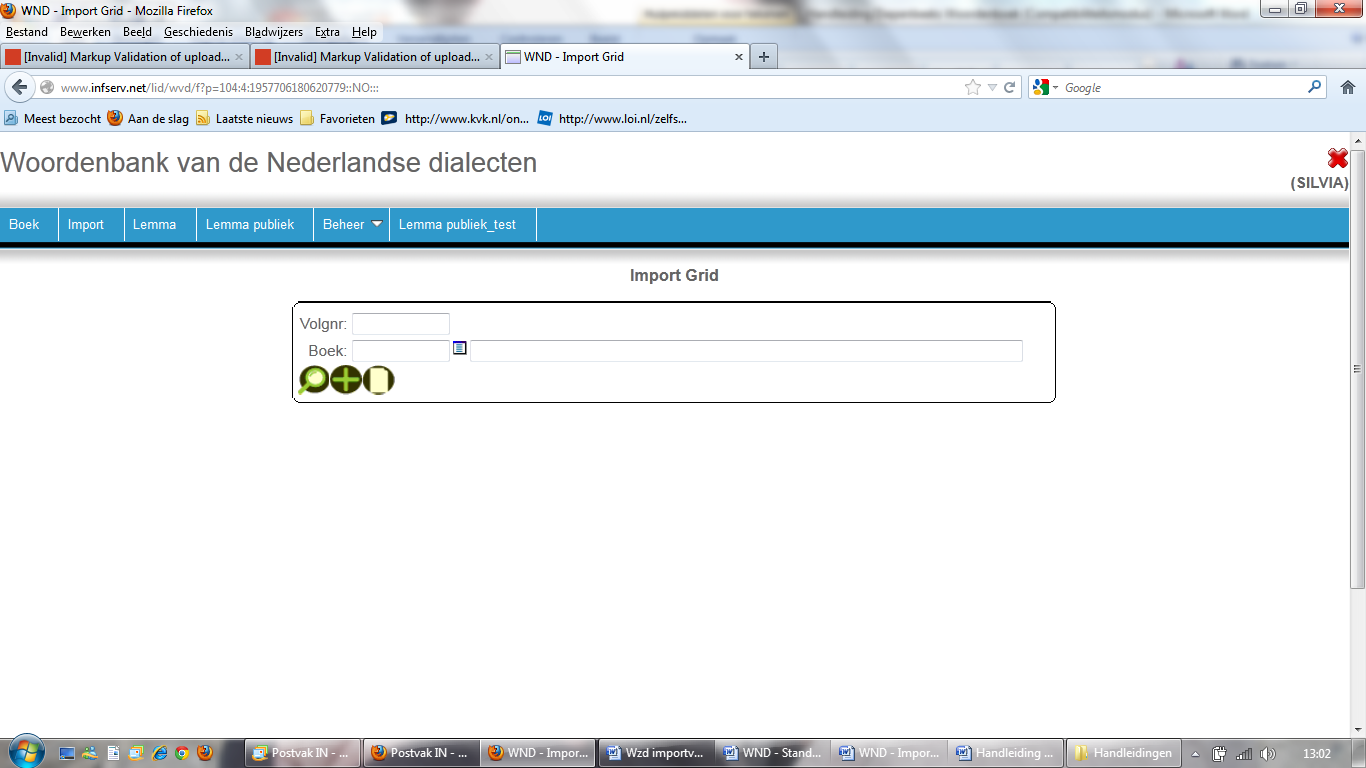


**2e inlog**

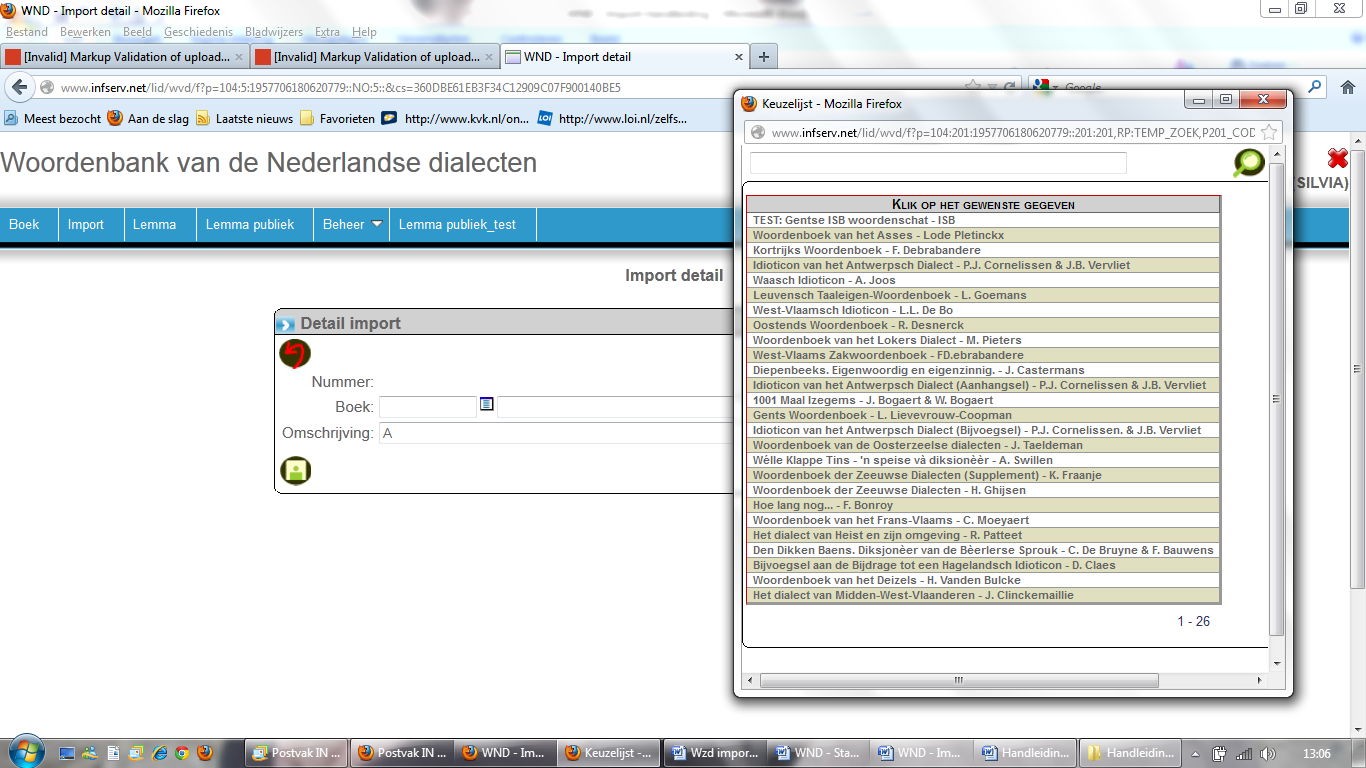
* Ga naar 'Import'



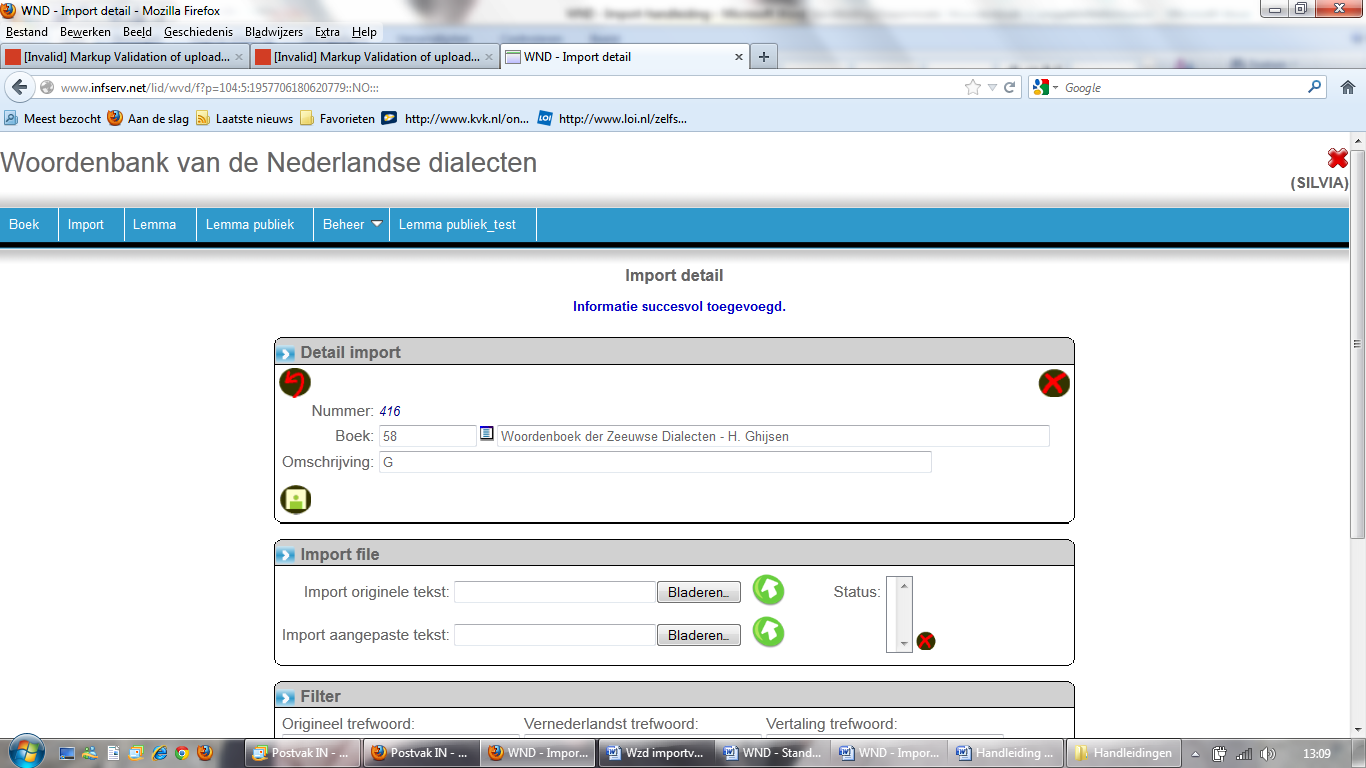
* Klik op het groene 'Plus-teken'.



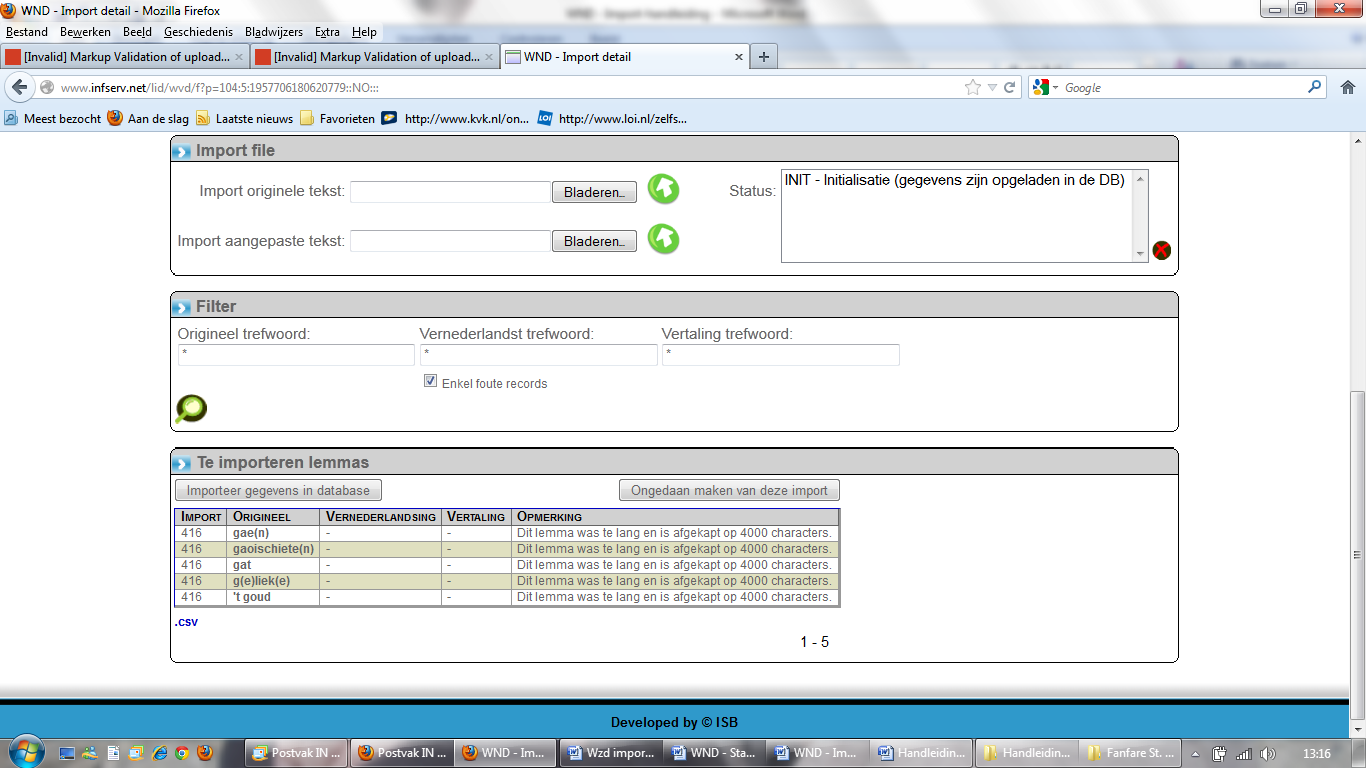
* Kies het juiste woordenboek in de keuzelijst en geef als omschrijving de titel van het bestand (d.h. de letter/letters die geïmporteerd wordt/worden). Druk op het 'Save-icoontje'.



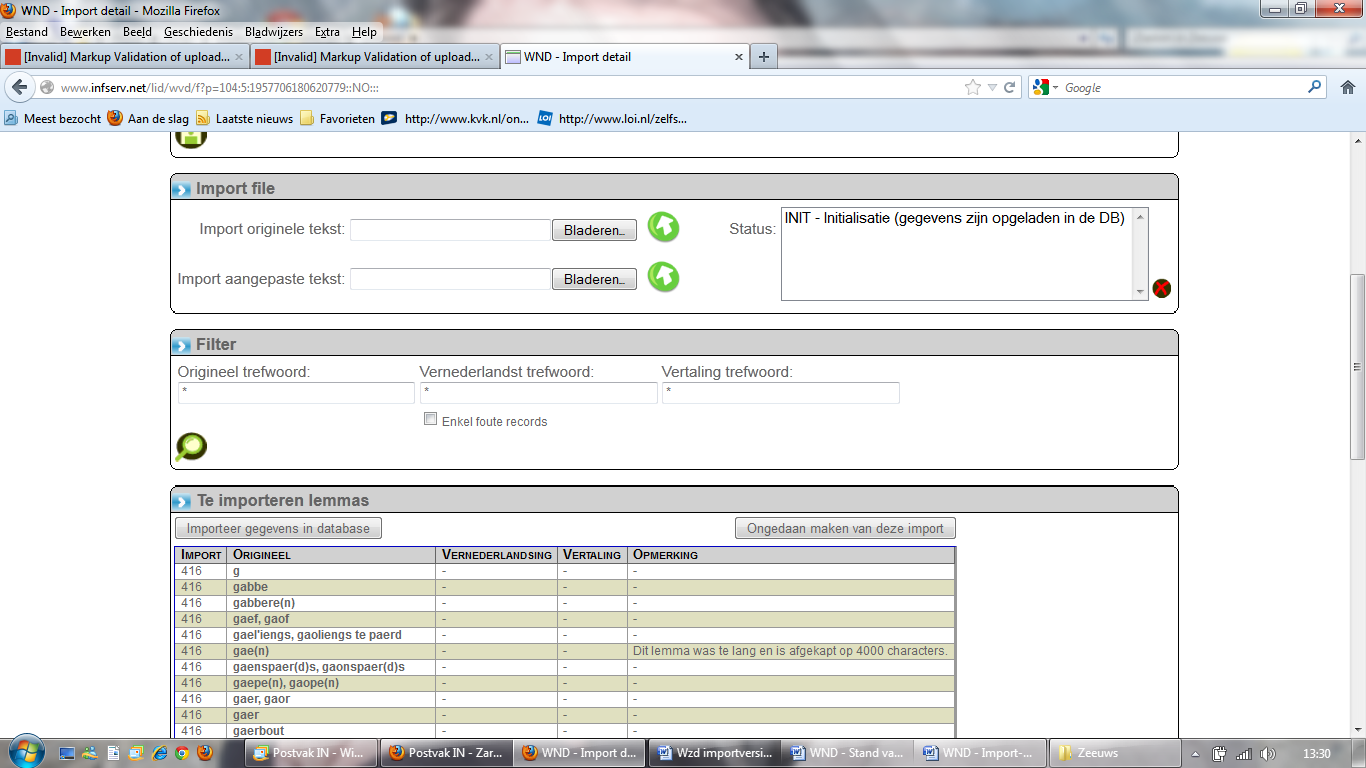
* Voeg de bestanden toe.



* Druk op 'Bladeren' bij 'Import originele tekst', kies het word-bestand van de te importeren letter en druk daarna op de wit/groene pijl. Klik daarna op 'Bladeren' bij 'Import aangepaste tekst' en kies hier het txt-bestand van dezelfde letter. Ook hierna weer op de wit/groene pijl klikken.
* Hierna verschijnt de lijst met alle lemmata. Vink het hokje 'Enkel foute records' aan, zodat de foutmeldingen verschijnen nadat je op het groene vergrootglas hebt geklikt.

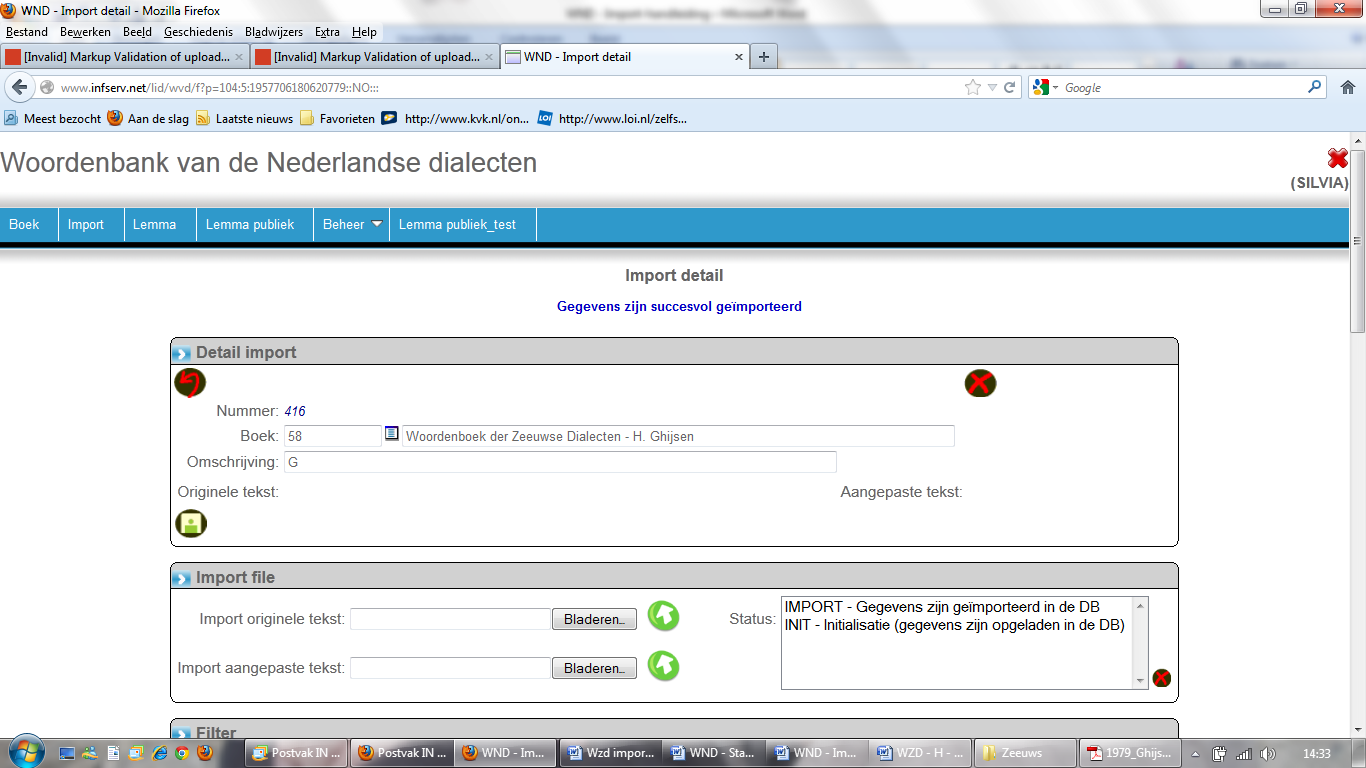


* 99% van alle foutmeldingen heeft verband met de twee zaken hieronder:
  + 'Meerdere originele trefwoorden gevonden': naast het vette lemma staat er nog iets van het artikel in het vet. Dit moet ontvet worden.
  + 'Dit lemma was te lang en is afgekapt op 4000 characters': hier is niets aan te doen en kan genegeerd worden. Dit geeft geen problemen voor de verdere import.
* Werk de foutmeldingen in het word-document weg, voer de macro opnieuw uit en sla het bestand opnieuw als txt-bestand op. Voer het bovenstaande import-proces nogmaals uit, totdat er geen foutmeldingen meer te zien zijn (behalve de '4000 characters').
* Vink de 'Enkel foute records' uit, klik op het groene vergrootglas en klik dan op 'Importeer gegevens in database' om de lemmata definitief te importeren.

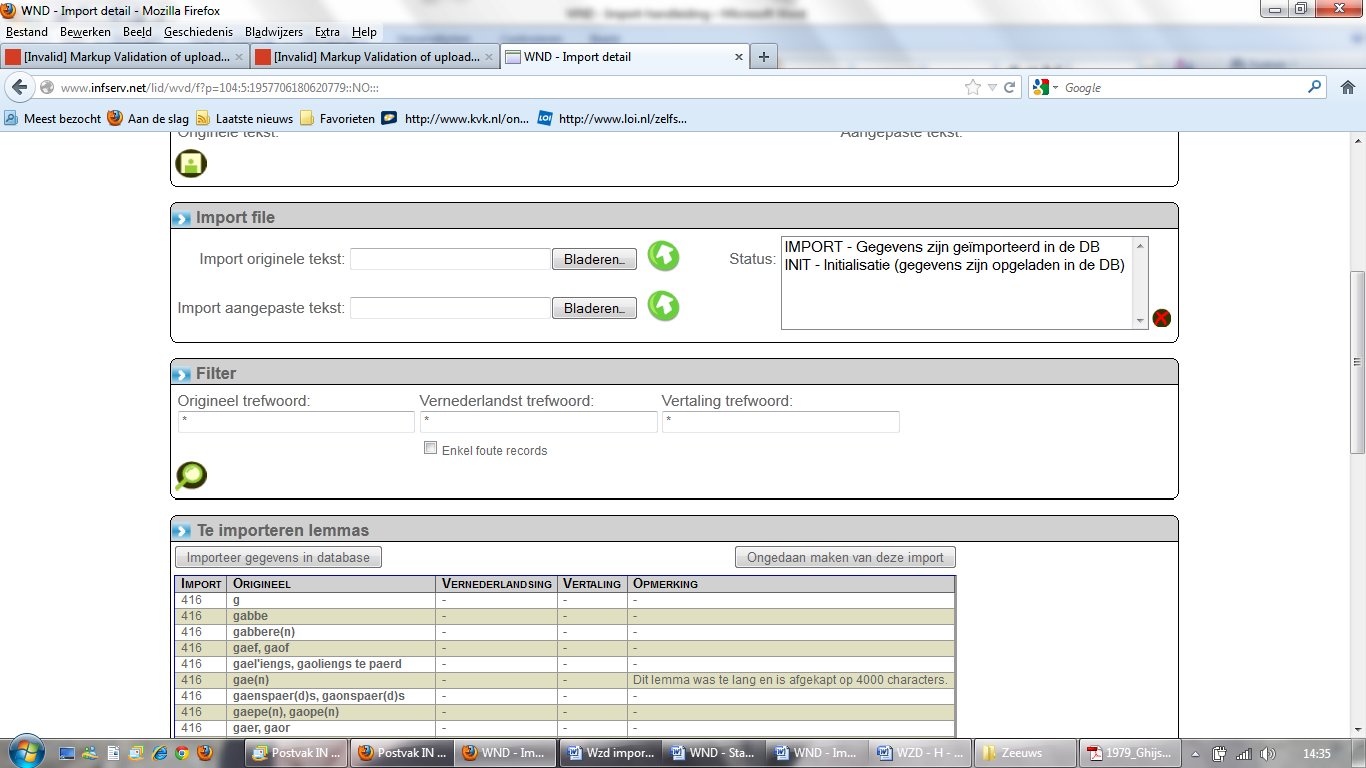


**Extra:**

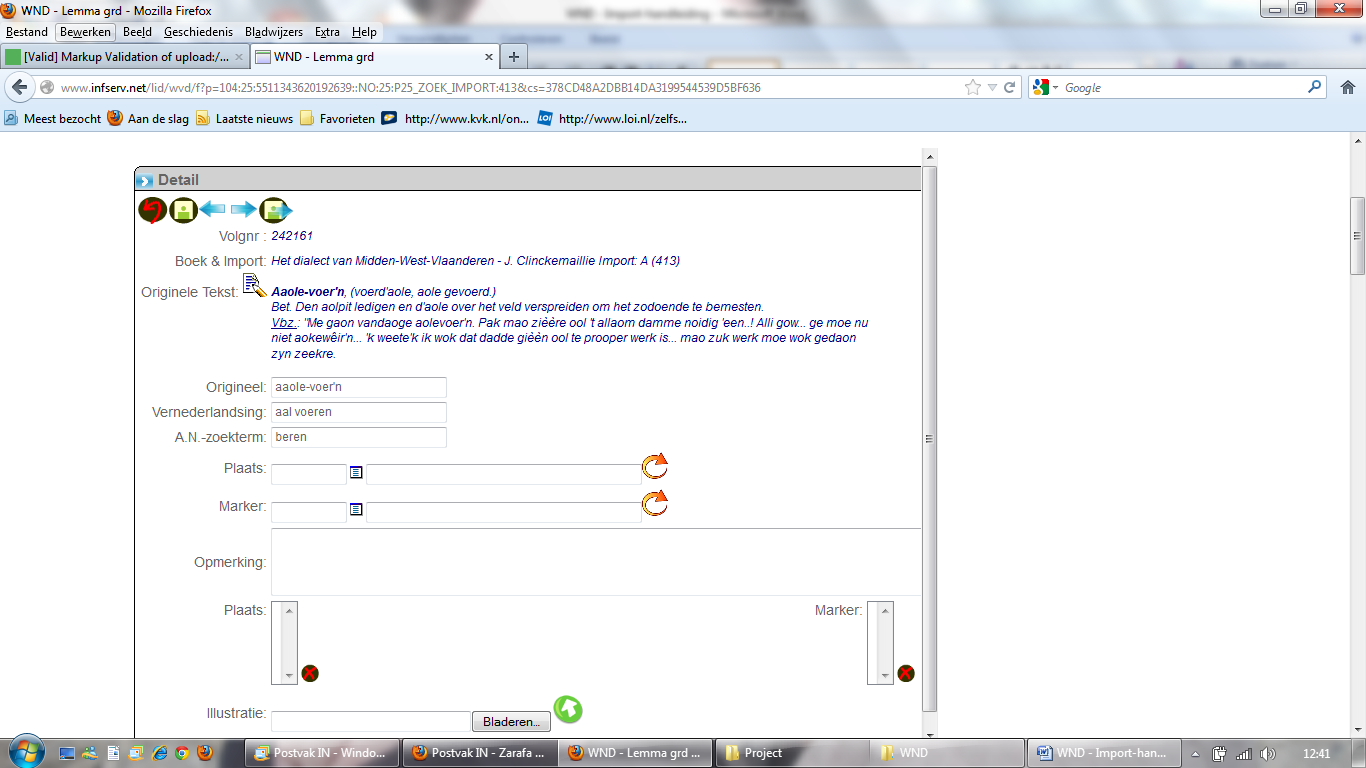
Om de aangemaakte letter/letters te verwijderen, klik op het rode kruisje, rechtsboven in de hoek.



Om de import ongedaan te maken, klik eerst op 'Ongedaan maken van deze import', selecteer daarna 'IMPORT – Gegevens zijn geïmporteerd in de DB' om vervolgens op het rode kruisje te klikken. Selecteer daarna 'INT – Initialisatie (gegevens zijn opgeladen in de DB)' en dan nog een keer op het rode kruisje. Hierna kunnen de documenten opnieuw geupload worden.



Om afbeeldingen toe te voegen aan de afzonderlijke lemmata moet je in het 'Detail'-veld van het lemma bij 'Illustratie' op 'Bladeren' klikken en de gewenste afbeeldingen kiezen. Druk daarna op de wit/groene pijl om de afbeelding te importeren.



Om de kaft, het titelblad, voorwerk en nawerk toe te voegen, moet je bij 'Detail Boek' de extra bestanden toevoegen door op 'Bladeren' te klikken en de gewenste documenten te kiezen. Druk daarna op de wit/groene pijl om de afbeelding te importeren.

