

Verslag bijeenkomst BiographyNed 12-11-12, 9:45-11:45

Aanwezig: Serge ter Braake, Jane Darnell (Wikipedia), Antske Fokkens, Jelle Gerbrandy (via Skype), Rik Hoekstra, Els Kloek (tot 10:45) Piek Vossen

Afwezig met bericht: Susan Legêne, Walter Ravenek

Voorzitter: Piek

Verslag: Serge

Stand van zaken:

Serge is bezig geweest met het handmatig invullen van case 5 over de gouverneurs-generaal van Nederlands-Indië, waarvan de resultaten als streefdoel moeten gaan dienen voor de demonstrator. De benodigde data zijn nadrukkelijk alleen uit het BP gehaald, uit die woordenboeken waar ook de demonstrator uiteindelijk mee aan de slag kan. Serge is nu nog bezig met de analyse van die data en het genereren van cijfers en verwacht daar snel mee klaar te zijn. De verwachting is dat deze case ook kan worden uitgediept voor moeilijkere onderzoeksvragen, met betrekking op bijvoorbeeld events en waardering van personen.

Antske is vooral bezig met het kijken naar beschikbare tools voor dataherkenning om personen, plaatsnamen en chronologische data in de teksten te herkennen. De uit de teksten gegenereerde data worden teruggekoppeld naar de metadata van het BP. De identificatie verloopt voorspoedig, alleen data voor het jaar 1000 worden niet herkend. Dit is echter geen onoverkomelijk of heel urgent probleem.

Antske en Serge hebben overlegd over de mogelijke problemen van een bevooroordeelde onderzoeksmethode bij de ontwikkeling van de demonstrator, op het moment dat er al hypothesen zijn over de uitkomst van een bepaalde onderzoeksvraag. Zo lang Niels nog niet begonnen is - en hij als hij begonnen is vermoedelijk vooral zal werken aan de opzet van de infrastructuur - is dit mogelijke probleem niet direct urgent en kan over twee weken in de groep worden voorgelegd.

Els heeft een lijst van betrouwbaarheid van de verschillende collecties uit het BP rondgestuurd en contact gelegd met Jane Darnell van Wikipedia, die nu ook aanwezig is.

Wikipedia:

Jelle (aangevuld door Els en Rik) doet uit de doeken hoe de koppeling tussen personen in het BP en Wikipedia tot stand is gekomen. De koppeling is onvolledig omdat de links alleen automatisch gegenereerd zijn naar personen op Wikipedia die met een hoge mate van zekerheid te identificeren waren met de mensen in het BP. Bovendien is de koppeling circa anderhalf jaar geleden gemaakt en is er in de tussentijd veel bijgekomen op Wikipedia. De vraag is in hoeverre de koppeling nog een keer, eventueel geholpen door hetgeen er voor BiographyNed wordt ontwikkeld, gemaakt moet worden. Er is behoefte aan duurzame, linking software.

Jane geeft aan welke projecten er binnen Wikipedia en zusterpagina's (Wikidata, Wiki Commons etc ...) lopen en waar er eventueel mogelijkheden zijn voor een samenwerking met BiographyNed. Ze benadrukt dat de Engelse en Duitse Wikipedia veel enthousiaster zijn over "data" en dat een samenwerking daar wellicht vruchtbaarder is dan bij de Nederlandse variant, waar men het nut van het BP niet inziet. Bijkomend voordeel is dat in de Duitse Wikipedia de personen goed geïdentificeerd zijn door de toekenning van een uniek PND((Personennamendatei). De Nederlanders bij Wikipedia zijn minder overtuigd van de meerwaarde van een toekenning van een nummer. Piek en Antske zeggen dat het geen probleem zal zijn om ook Engelse en Duitse teksten mee te nemen in het onderzoek en dat dit zelfs wenselijk is.

De mogelijkheden van het inzetten van Wikipedisten of andere vrijwilligers (crowdsourcing) bij BiographyNed wordt besproken. Mogelijkerwijs kunnen zij in de ontwikkelingsfase bepaalde data op gestructureerde wijze ordenen (een standaard zetten zoals nu bij case 5 wordt gedaan), waar de resultaten van de demonstrator tegen kunnen worden afgezet. Ook kunnen de vrijwilligers helpen bij het verder invullen van in de toekomst door de demonstrator automatisch gegenereerde stambomen.

Afspraken:

- Het is wenselijk dat Huygens en Wikipedia verder praten over mogelijke samenwerking in de toekomst. Voor BiographyNed kan dat een rijkere te gebruiken dataset opleveren, maar is het op dit moment niet van direct belang voor de verdere ontwikkeling van de demonstrator. Via Jane houden we de ontwikkelingen bij Wikipedia graag verder in de gaten.
- De medewerkers van BiographyNed concentreren zich de komende drie jaar in de eerste plaats op het creëren en gebruiken van een eigen database in RDF die gekoppeld is aan het BP, zoals besproken in de vorige vergadering.
- Antske en Jelle gaan via de chat verder overleg plegen over de te gebruiken tools.
- Serge voltooit case 5 handmatig en reflecteert verder over de bruikbaarheid van het BP voor het beantwoorden van historische vraagstellingen.
- Antske en Serge presenteren in de volgende vergadering iets over de mogelijke methodologische problemen bij de ontwikkeling van de demonstrator.