Interview Biography Net

Hun team bestaat uit een historicus (Serge), een computer-linguist (Antske) en een informaticus (Niels). Hun begeleider is Victor.  
Ze willen door middel van een NLP pipeline een netwerk van naar elkaar verwijzende RDF data maken, waardoor opeens de inhoud van de biografieën ook toegankelijk is voor de computer, en dus doorzoekbaar.

NLP PipelineAntske heeft ons haar NLP pipeline laten zien, waarin ze de tekst van een biografie via verschillende stappen omzet in andere formaten.  
Af:   
- tokenizer + morfologische data  
- syntactische structuur  
- elk woord gematcht met een ‘named entity’ via DBpedia

Mee bezig (hopelijk over 3 weken af):  
- triples maken van verschillende entities: dus isDaughterOf(Maria, Pieter)  
- …. ? (vergeten)

Compleetheid van data  
Gemiddelde lengte biografieën: 200 woorden  
Naam en categorie: 100%  
Sterfdatum: 80%  
Geboortedatum: 60% ongeveer  
Maar een paar duizend hebben verwijzingen naar ouders. Heel weinig verwijzingen naar namen van kinderen, dus dat dubbel koppelen wordt moeilijk.

Verschillende weergavemogelijkheden  
Niels zorgt ervoor dat je op verschillende manieren een overzicht kan zien de data van meerdere biografieën / mensen tegelijkertijd.  
- Tijdlijn met balkjes van levens  
- Wereldkaartje met eventueel een te verschuiven tijdlijn (zodat je kan scrollen door de tijd)

Ook maakt hij een tool voor als bijvoorbeeld een geboortedatum onzeker is, dan kan je als historicus zien welke bronnen welke datum ondersteunen, en ook waar dat in de tekst staat.

Praktisch  
We mogen best een dag / een aantal dagen onofficiëel op de VU werken, als we iets moeilijks aan het doen zijn. Zodat zij onze vragen kunnen beantwoorden / ons helpen.

Onze projectmogelijkheden  
1. **Gerelateerd**  
Gerelateerde mensen. Dus op het moment dat je zoekresultaten van een persoon krijgt, dat je suggesties krijgt over wie je misschien ook interessant vindt (een familielid, ook een wiskundige, etc).

Antwoord: Dit is een goed project, ook goed mogelijk. Dan willen we waarschijnlijk de info gebruiken uit het deel van de NLP pipeline die Antske nu aan het ontwikkelen is. Dan kan je uit een biografie halen wie de ouder is. Dan moet je dus nog bedenk wie dat is en die erbij zoeken en ook tonen.  
Soort functie: Op 1 persoon  
Probleem: in een biografie staat vaak niet wie de kinderen zijn.  
Probleem: kan voor het grootste deel pas over 3 weken (als het dan af is).  
Probleem: misschien best moeilijk

**1. Percentages**  
Percentages laten zien: bijvoorbeeld dat je kan zoeken: ‘hoeveel wiskundigen waren vrouw?’ ‘hoeveel wiskundigen waren moslim?’ etc  
 Dus dat je bij een bepaald attribuut ‘percentage’ aan kan vinken, dat je dan combineert met de andere attributen. Dus als je aanvinkt: rubriek: wiskundige, geboortedatum: 1900, percentage:vrouw, dat je zoekt: welk percentage wiskundigen, geboren in 1900 waren vrouw?

Antwoord: (Niet zo enthousiast, wss omdat hun tool niet per se voor dit soort dingen geschikt is). Kunnen we natuurlijk nog steeds doen. Is denk ik ook redelijk makkelijk.

Eventueel:

3. **Kaartje**  
a. Kaartje (bijv. google maps) laten zien als je bijvoorbeeld zoekt op ‘rubriek: wiskundigen, levenstijd: ….’ En dat je dan oftewel de geboorteplaats oftewel de sterfplaats kan zien van alle wiskundigen in die tijd. Dat je bijvoorbeeld die kan aanvinken, of later kan aanklikken onder in een pagina van resultaten.  
b. Kaartje met pijlen van geboorteplaats naar sterfplaats voor een bepaalde combinatie van attributen: bijvoorbeeld: ‘wat is de sterfplaats van de wiskundigen die zijn geboren in Amsterdam?’ Zo kan je stromen zien (bijvoorbeeld dat nu alle goede acteurs naar de Verenigde Staten gaan (oid).  
c. Een kaartje met een tijdlijn waar je doorheen kan scrollen (migratie door de tijd zichtbaar)

Antwoord: Is ook een goed project. Niels is hier al deels mee bezig. Er bestaat al ergens (bij een andere biografiëen site?) Waar dan?  
Soort functie: Op meerdere personen  
Probleem: Best moeilijk, Niels al deels mee bezig?