

# Handleiding Annotation tool eHOST

Versie 0.8

Doel van dit document is om zoveel mogelijk informatie over het gebruik van eHOST te verzamelen en op begrijpelijke wijze uiteen te zetten. De documentatie waar het op is gebaseerd, is voor de volledigheid als bijlage toegevoegd aan het eind.

## Inhoud

1. eHOST opstarten	2
1.1. eHOST downloaden en opstarten	2
1.2. eHOST starten met dubbelklik	2
1.3. eHOST vastmaken aan de Launcher	2
2. Naam annotator aangeven	3
3. <i>Workspace</i> voorbereiden en gegevens laden	3
3.1. Workspace aanmaken	3
3.2. Gegevens laden	4
3.3. Classes, attributes en relationships	4
Annotation Schema Manager / Class Manager	5
Attribute Schema Editor	5
Relationship Schema Editor	6
4. Annoteren	6
4.1. Personen en beroepen markeren	6
4.2. Relationships aangeven	7
4.3. Attributes toekennen	7
4.4. Workflow: twee mogelijkheden	8
4.5. Annotaties verwijderen	9
4.6. Eerder gemaakte annotaties laden	9
5. Resultaten opslaan en exporteren	10
6. Annotaties vergelijken	10
Appendix: casus specifieke annotation guidelines	13

## 1. eHOST opstarten

### 1.1. eHOST downloaden en opstarten

Het programma eHOST is te downloaden op de Google Code pagina van de applicatie.<sup>1</sup> eHOST zal worden gedownload als .zip-bestand, dat uitgepakt dient te worden. In de uitgepakte map is het bestand "eHOST.jar" te vinden. Dit bestand kan gebruikt worden om het programma op te starten. Dit kan door middel van de linux-terminal, met het volgende commando:

```
> sudo java -jar "eHOST.jar"
```

Let op dat je de terminal start in de map waar het .jar-bestand te vinden is. Dit kan je doen door in deze map op de rechtermuisknop te klikken en "open in terminal" te kiezen. Een andere wijze waarop vanuit de terminal naar de juiste map te navigeren is, is door middel van het "cd" commando (hieronder volgt een voorbeeld):

```
> cd /home/ruben/software/ehost-bp1
```

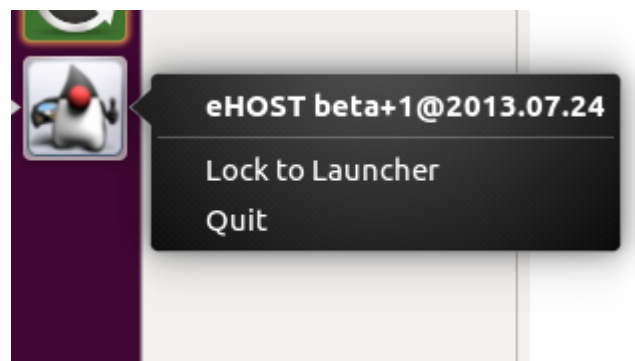
### 1.2. eHOST starten met dubbelklik

Makkelijker voor Ubuntu-gebruikers is de optie om het programma te starten door middel van een dubbelklik op het .jar-bestand. Dit kan door middel van de volgende stappen:

1. Rechtermuisknop op "eHost.jar"
2. Selecteer "Properties"
3. Navigeer naar tabblad "Permissions"
4. Kruis aan: "Allow executing file as program"
5. Nu kan je als het goed is eHOST.jar starten met een dubbelklik.

### 1.3. eHOST vastmaken aan de Launcher

Om het jezelf nog makkelijker te maken, kan je eHOST vastmaken aan de "Launcher" (de toolbar links). Dat doe je door als het programma draait op het eHOST-symbool in de Launcher te klikken met de rechtermuisknop en "Lock to Launcher" te selecteren. Nu kan je het programma daar standaard terugvinden.

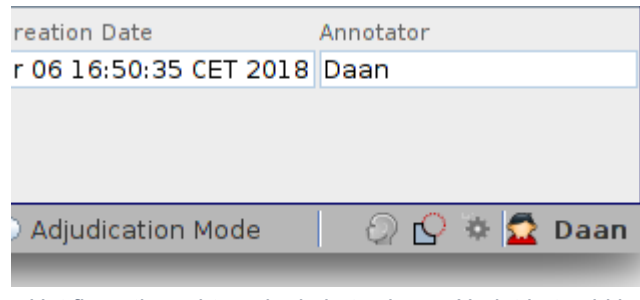


---

<sup>1</sup> <https://code.google.com/archive/p/ehost/downloads>

## 2. Naam annotator aangeven

Het is mogelijk om aan te geven wie annotateert. Dat doe je door rechtsonder in het scherm op het menselijk figuurtje te klikken (dubbelklik). Het annotator-scherm verschijnt, daar kun je je naam invullen. Die naam wordt vervolgens geassocieerd met de bestanden die je in jouw programma annotateert; dat is mogelijk handig voor Ruben en Daan om het overzicht te behouden.



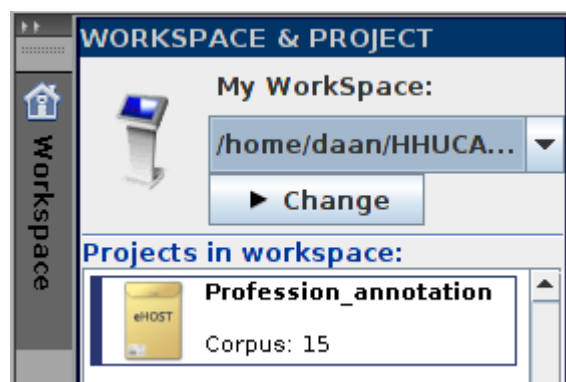
## 3. Workspace voorbereiden en gegevens laden

### 3.1. Workspace aanmaken

Om met eHOST aan de slag te gaan, moet je eerst een *Workspace* selecteren. Dat is simpelweg een map waarin je project wordt opgeslagen. Binnen een *Workspace* kunnen verschillende projecten bestaan, dat zullen in ons geval de wekelijks aangeleverde reeks teksten zijn. Om een tekst te bewerken, selecteer je eerst een project en vervolgens binnen dat project de tekst die je wil bewerken.

Binnen elk project moeten drie mappen aanwezig zijn: “config”, “corpus” en “saved”. In de map genaamd “config” bevindt zich “projectschema.xml”, dat eHOST vertelt welke schema's in het project worden gebruikt. Schema's zijn de *classes*, *attributes* en *relationships*. In de map “corpus” bevinden zich de tekstbestanden die bewerkt moeten worden en in de map “saved” worden de annotaties opgeslagen (maar let op: deze mappenstructuur wordt je toegestuurd, zie 3.2.).

Handig is om een map aan te maken die “hhucap” heet waarin je eHOST.jar hebt staan en in die map een nieuwe map genaamd “Projects” aan te maken. Vervolgens kan je die map als *Workspace* selecteren in eHOST en heb je voor de rest van het project één overzichtelijke plek waar alle data die je toegestuurd krijgt kan verzamelen.



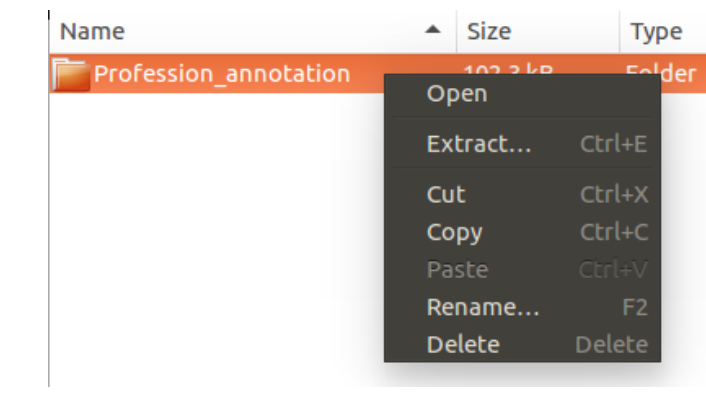
### 3.2. Gegevens laden

Het programma eHOST laat een gebruiker "classes" creëren om te markeren. Die zijn in dit geval al aangemaakt, namelijk "Person" en "Profession". Daarnaast kunnen met eHOST classes met elkaar in verband worden gebracht door "relationships" tussen twee markeringen te maken. Die relationships moeten in formele RegEx-taal aan het programma gevoed worden, maar ook dat is al voor ons gedaan.

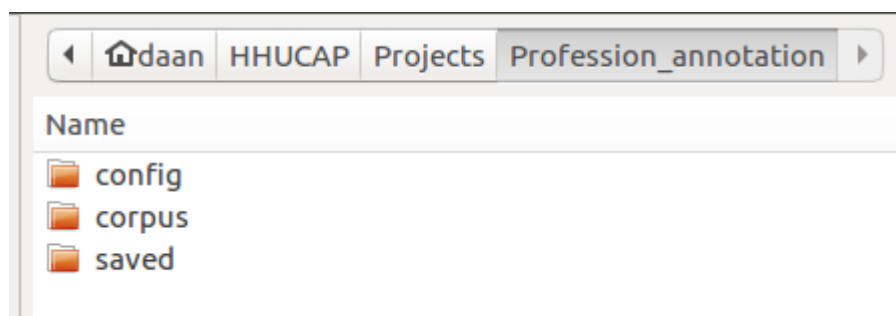
Relationship
<code>profession_holder &lt;((Profession))((Person))((Profession))&gt;</code>

Daar hoeven wij ons dus verder niet mee bezig te houden. Het enige dat voor annotators van belang is, is om de voorgeprogrammeerde classes en relationships op correcte wijze in het programma te laden om daar vervolgens mee aan de slag te kunnen.

Om de testgegevens te laden, moet het zip-bestand (professionannotation.zip) in de e-mail die Isa Maks op 27 februari heeft verstuurd, worden gedownload en uitgepakt (of: extracted). Door bij het uitpakken te kiezen voor de workspace ("Projects") die bij de vorige stap is aangemaakt, of voor een workspace met een zelfgekozen naam, komen de bestanden direct op de juiste plek terecht.



Als de bestanden juist zijn uitgepakt en zich op de juiste manier op de harde schijf/virtuele schijf



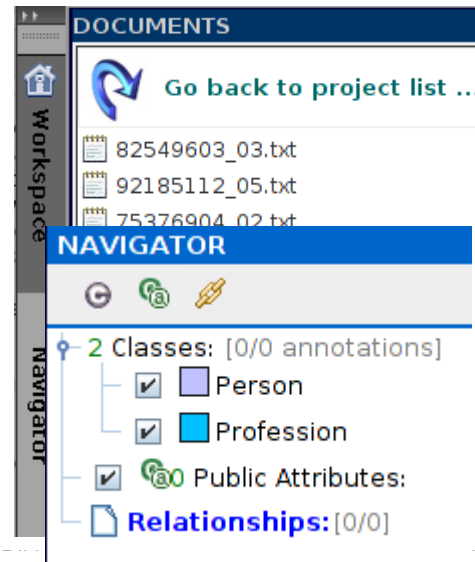
bevinden, zoals in de illustratie hierboven, zouden zowel de tekstbestanden als de *classes*, *attributes* en *relationships* zich in de workspace van eHOST moeten bevinden.

### 3.3. Classes, attributes en relationships

Als we nu eHOST openen, geeft eHOST als het goed is in de workspace de map weer waarbinnen zich de tekstbestanden bevinden die bewerkt moeten worden, zoals op de afbeelding op **pagina**

2. Die map is dus een project. Zodra je dat project selecteert, verschijnt op de verticale taakbalk aan de linkerkant van eHOST een nieuw tabblad: "Navigator".

Door op het tabblad Navigator te klikken, verschijnen de classes die gemarkeerd kunnen worden in een project. In ons geval zou hier nu dus "Person" en "Profession" moeten staan.



Onder "Navigator" staan drie symbooltjes. Van links naar rechts staan die voor "Annotation Schema Editor", "Attribute Schema Editor" en "Relationship Schema Editor". In principe hebben we hier als annotators niets te zoeken, maar het kan geen kwaad om even een kijkje onder de motorkap te nemen.

#### *Annotation Schema Manager / Class Manager*

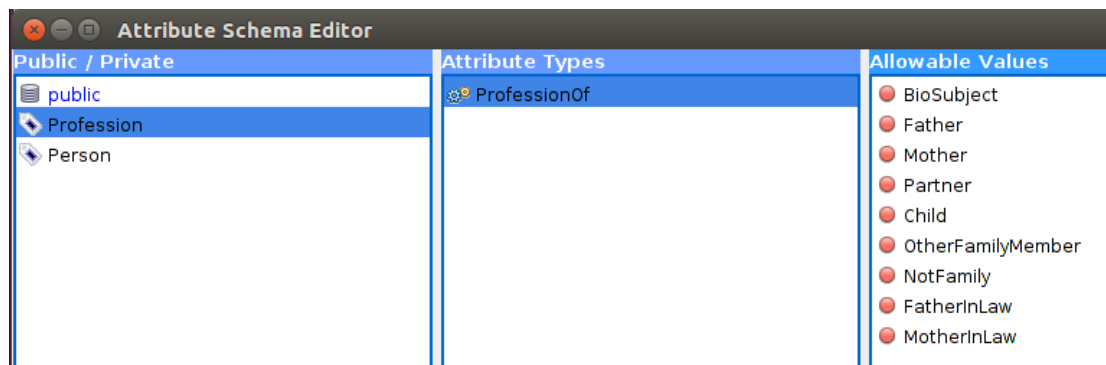
Als we op het eerste symbool klikken, opent de Annotation Schema Manager, ook wel de Class Manager. In dit scherm kunnen de namen en kleuren van bestaande classes aan worden gepast en ook nieuwe classes worden toegevoegd. Dat gaan we niet doen, dus we kunnen dit scherm direct weer sluiten.

#### *Attribute Schema Editor*

Het tweede symbool staat voor de Attribute Schema Editor. Hier zien we drie kolommen, met in de meest linkse de twee classes "Profession" en "Person". Met deze editor kunnen attributen worden toegekend aan classes. Ook in dit geval is dat al voor ons gedaan. Binnen dit project heeft alleen de class "Profession" een attribuut, namelijk "ProfessionOf".

Door Profession te selecteren verschijnt dit attribuut in de kolom "Attribute Types" en door vervolgens dat weer te selecteren verschijnen de "Allowable Values": BioSubject, Father, Mother, Partner, Child, OtherFamilyMember, NotFamily, FatherInLaw, MotherInLaw.

Dit zijn de waarden die aan elk gemarkeerd beroep moeten worden toegekend om aan te geven wat de relatie tussen de hoofdpersoon van de biografische tekst en de uitoefenaar van het beroep is. Hoe dat moet, komt zo aan bod.



### Relationship Schema Editor

Tot slot de Relationship Schema Editor. Wederom is dit schema volledig voorgeprogrammeerd. De eerder besproken (onder 3.2) formele invulling van relationships is hier ingevoerd. Deze formule vertelt eHOST welke relationships mogelijk zijn binnen het betreffende project.

Relationship Name:	
profession_holder	
Start From Class:	To Class:
Profession	Profession
Person	Person

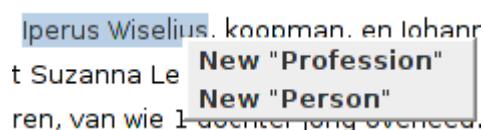
Door op Edit te klikken, verschijnt bovenstaand scherm. Let erop dat je hier niets aan verandert, omdat het beïnvloedt wat voor verbanden je kunt leggen. Voor onze doeleinden is het alleen nodig om relaties van Profession naar Person te leggen. Dat betekent dat je een relationship alleen kunt creëren door éérst op een gemarkeerd beroep te klikken en daarna op een persoon, niet andersom. Hierover dadelijk meer.

## 4. Annoteren

Als alle bovenstaande stappen correct zijn doorlopen (en de instellingen in de Relationship Schema Editor niet aangepast zijn), kan het markeren beginnen. Eerst leggen we de verschillende stappen uit, waarna twee suggesties voor een "workflow" worden gedaan.

### 4.1. Personen en beroepen markeren

Een persoon of beroep markeren is simpelweg een kwestie van het selecteren van de naam van een persoon of beroep. Zodra de muis wordt losgelaten, verschijnt een mini-menuutje met "New Profession" en "New Person". Selecteer wat van toepassing is.



## 4.2. Relationships aangeven

Deze stap is lastiger, niet zozeer omdat het een ingewikkelde taak is, maar vooral omdat eHOST hier nog weleens dwars wil liggen.

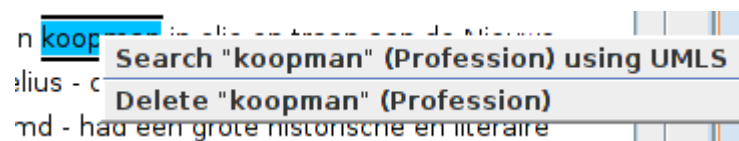
In theorie is het eenvoudig: markeer een persoon, markeer een beroep en klik op het groen-witte plus-tekentje in de Annotation Editor aan de rechterzijde van het programma. Klik nu met de rechtermuisknop op het beroep en vervolgens op de bijbehorende persoon: een geel lijntje verschijnt en de connectie is gemaakt.

In de praktijk komt het soms voor dat de relationship om een of andere reden *niet* tot stand komt. Één verklaring zou kunnen zijn dat in de Relationship Schema Editor alleen het verband Profession -> Person als mogelijkheid is gegeven. Dat betekent **denk ik** dat het niet mogelijk is om een relationship te creëren door eerst op de persoon te klikken; je *moet* eerst op het beroep klikken en daarna pas op de persoon.

Je moet bovendien eerst met de linkermuisknop het beroep al selecteren. Dus: met links op *beroep*, met rechts op *beroep*, met rechts op *persoon*. Je kunt zien wat je hebt geselecteerd aan de zwarte lijntjes boven en onder de tekst. Zie de afbeelding hieronder.

Na de Latijnse School met glans te hebben doorlopen en na nog een jaar de colleges van de **filoloog** en filosoof **Daniël Wyttenbach** op het Athenæum Illustre te hebben gevolgd vertrok Samuel in 1786 naar Franeker. Tot de beoogde studie in

Let er daarnaast op dat het groen-witte plusje is aangeklikt voordat je een verband wil leggen. Dat wil zeggen: je kan alleen een verband leggen als het groen-witte kruisje *niet* zichtbaar is. Als het kruisje niet is geselecteerd, mislukt de poging om een relationship aan te maken en krijg je dit menuutje te zien:



**Wiselius, Samuel Iperuszoon, bestuurder**  
4 2 1769 - Amsterdam 15 5 1845). Zoon

## 4.3. Attributes toekennen

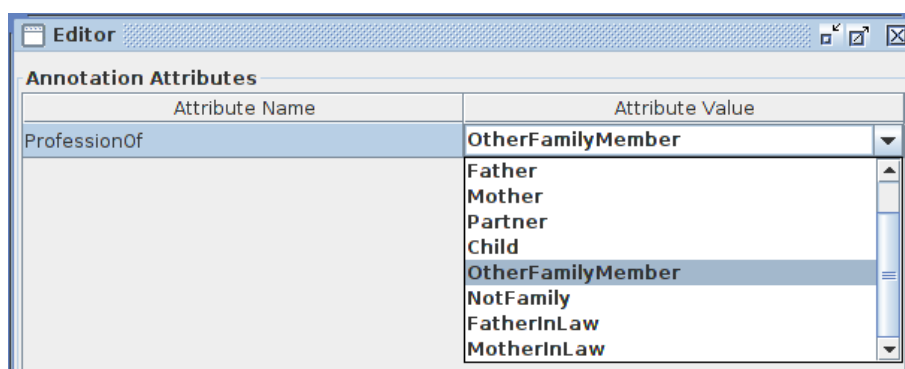
De volgende stap in het annotatieproces is om de *attributes* toe te kennen die te zien zijn in de Attribute Schema Editor, maar we hoeven gelukkig niet helemaal naar die editor te navigeren om ze aan de tekstbestanden toe te voegen. Het kan door simpelweg een gemarkeerd beroep te

selecteren en vervolgens in de Annotation Editor (het rechter zijdscherm) op het vergrootglasje rechts van *Attributes* te selecteren.

Er verschijnt een pop-up scherm. Hier kan je uit de lijst met verschillende *attribute values* kiezen wat van toepassing is: welke sociale positie heeft de uitoefenaar van dit beroep ten opzichte van

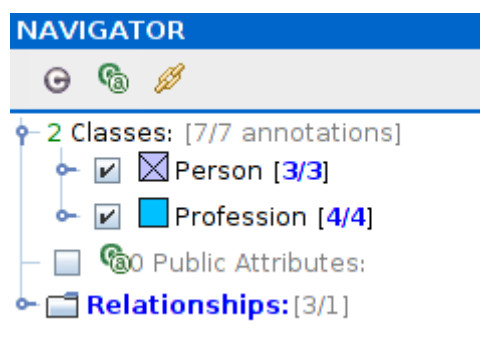


de hoofdpersoon van de biografie? Als het de persoon zelf is, kies je *BioSubject*. Is de persoon in kwestie geen familie van de hoofdpersoon, kies dan *NotFamily*, etc. Het wijst zichzelf.



#### 4.4. Workflow: twee mogelijkheden

Hoewel je zelf moet bepalen hoe je te werk gaat, maakt eHOST ruwweg twee manieren mogelijk. De ene wijze ligt voor de hand, namelijk het om en om markeren van persoon en beroep en dan de relaties ertussen nog aangeven. Die manier van werken is mogelijk doordat eHOST bij het selecteren van een stuk tekst automatisch de mogelijkheid geeft om een nieuw beroep of een nieuwe persoon toe te voegen, zoals we bij 4.1 hebben gezien.



Het kan ook anders: eHOST maakt het heel eenvoudig om een van beide opties tijdelijk als standaard in te stellen, waardoor elk stuk tekst dat je markeert automatisch wordt geclassificeerd als, bijvoorbeeld, beroep.

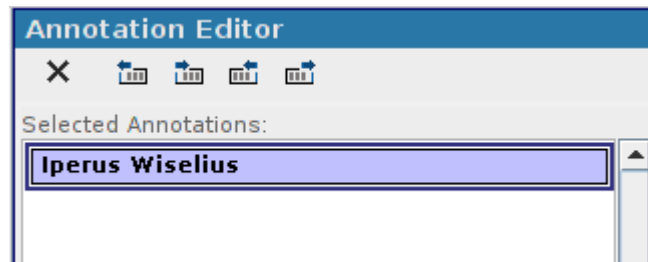
Je stelt dit in door in het "Navigator"-tabblad met de rechtermuisknop op de *class* te klikken die je tijdelijk als standaardoptie wil instellen. Je maakt het ongedaan door er nog eens met rechts op te klikken.



Dit maakt het mogelijk om in één keer rap door de tekst te werken op zoek naar enkel persoonsnamen, om pas daarna over beroepen en relaties na te denken. Misschien is dit niet handiger dan de andere optie, maar het is de moeite waard om het te proberen.

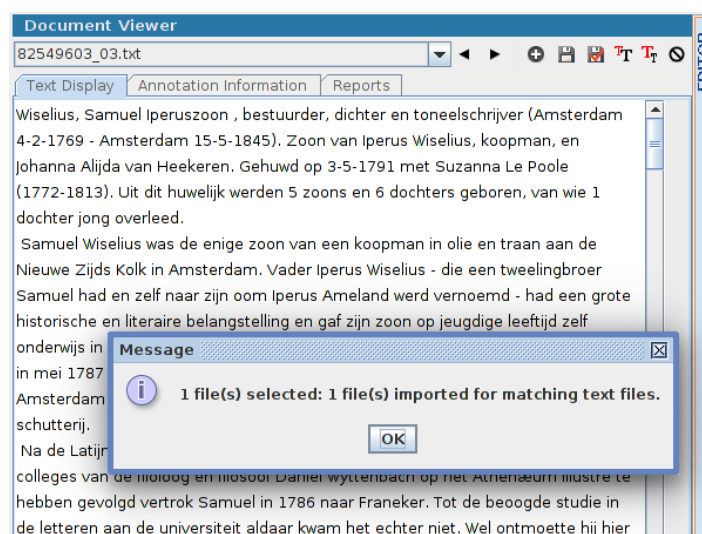
#### 4.5. Annotaties verwijderen

Annotaties kunnen eenvoudig verwijderd worden. Selecteer de annotatie die je wil verwijderen en klik dan op het kruisje bovenin de Annotation Editor aan de rechterzijde van het scherm.



#### 4.6. Eerder gemaakte annotaties laden

Als je eerder gemaakte annotaties wil laden in een nieuw project, moet je als eerste stap de tekst openen waar je de annotaties in tevoorschijn wil toveren. Heb je dat gedaan, dan kan je op het plusje bovenin het Document Viewer-scherm klikken. Deze knop laat je namelijk annotaties importeren uit "Knowtator XML"-bestanden – dat is het format waar annotaties in worden opgeslagen. Om bestaande annotaties te importeren, hoef je dus alleen te weten waar je ze bewaart.



Let er wel op dat het importeren alleen werkt als de bestandsnaam en het bestand dat je importeert overeenkomt met het tekstbestand dat je aan het bewerken bent. Is dat het geval, dan verschijnen de eerder gemaakte annotaties direct in je tekst.

## 5. Resultaten opslaan en exporteren

Zoals eerder aangegeven, bestaat een project uit drie mappen: config, corpus en saved. In die laatste map worden de annotaties die je maakt automatisch opgeslagen. Zodra je klaar bent, kan je de inhoud van de saved-map dus integraal bij ons inleveren in de xml-map op Google Drive.

## 6. Annotaties vergelijken

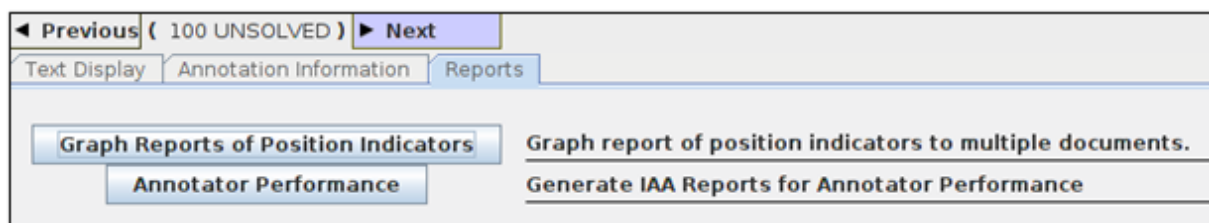
1. Laad de annotaties van de eerste annotator.

2. Controleer of de naam van de annotator bij de annotaties is vermeld. Is dat niet het geval, voeg dan de juiste naam aan de annotaties toe door rechtsonder in eHOST op het menselijk icoontje te klikken, de naam van de annotator in te voeren en op 'apply this user to all existing annotations' te klikken.

3. Laad de annotaties van de tweede annotator. Controleer ook hier of de naam van de annotator is vermeld. Zo niet, dan is het volgens mij nodig om eerst alle annotaties te verwijderen (knop rechtsboven in het middelste scherm), dan opnieuw de tweede (niet de eerste) annotator en pas hierop stap twee toe.

4. Doe hetzelfde voor de derde annotator.

5. Om rapporten op te stellen, selecteer je 'Reports' bovenin het middelste scherm. Klik op 'Annotator Performance'.



6. In dit scherm selecteer je op basis van welke kenmerken je de verschillende annotatoren wil vergelijken. Volgens mij moeten we classes (profession/person), relationships en attributes (Father, Son, NotFamily, etc.). Klik vervolgens op 'Generate Reports'.

Please Select Annotators to Compare ...

To generate an IAA report at least 2 annotators need to be selected from the following list:

**\* You need to select at least 2 annotators ...**

☒ Marjolein  
☒ Vincent

**Define criteria for differences:**  
 enOST will check annotations fitting the following conditions:  
☒ they have exact same spans ☐ or they also have their spans overlapped  
 and  
☒ they have same classes  
☒ they have same attributes  
☒ they have same relationships  
☐ they have same comments

**Select Classes:**  
☒ Profession  
☒ Person

Generate Reports Cancel

7. Er verschijnt een lijst met verschillende rapporten.

- [Class and span matcher](#)
- [Matched Details - SUMMARY](#)
- [Class\( Profession \) only](#)
- [Class\( Person \) only](#)
- [Unmatched Details for Annotator \( Marjolein \) SUMMARY](#)
- [Unmatched Details for Annotator \( Marjolein \) to class\( Profession \) only](#)
- [Unmatched Details for Annotator \( Marjolein \) to class\( Person \) only](#)
- [Unmatched Details for Annotator \( Vincent \) SUMMARY](#)
- [Unmatched Details for Annotator \( Vincent \) to class\( Profession \) only](#)
- [Unmatched Details for Annotator \( Vincent \) to class\( Person \) only](#)

8. De Class and span matcher geeft de percentages en absolute getallen voor overeenkomsten tussen de verschillende annotatoren. De precieze uiteenzetting van overeenkomsten vind je in 'Matched Details – SUMMARY' – dit is een opsomming van de annotaties die *precies* overeenkomen voor alle geselecteerde annotatoren. Dat zijn dus de juiste personen, de juiste beroepen, met de juiste relaties en de juiste attributen (dat percentage is op dit moment **2%** tussen twee gebruikers). De 'Unmatched Details' geeft een lijst met annotaties die wel door de ene annotator, maar niet door de andere(n) zijn aangemerkt.

## 2-way IAA Results

IAA calculated on 15 documents.  
all annotations = matches + non-matches  
IAA = matches / all annotations

For annotations between Annotator[Marjolein] and Annotator[Vincent]:

Type	IAA	matches	non-matches
All selected classes	2.0%	14	686
Profession	3.0%	12	384
Person	0.7%	2	302

## Pair-wise agreement

Gold standard set	compared set	true positives	false positives	false negatives	recall	precision	F-measure
Marjolein	Vincent	7	385	301	2.3%	1.8%	2.0%
Vincent	Marjolein	7	301	385	1.8%	2.3%	2.0%

Precision and recall are given equal weight for the F-score.

[\[Back to the index.html\]](#)

## Appendix: casus specifieke annotation guidelines

Hieronder volgen een serie vragen van annoteerders tijdens het HHuCap project. Hoewel deze vragen specifiek zijn voor het project, voegen we ze toe, omdat het inzicht geeft over het niveau van detail waarop je een annotation project moet voorbereiden.

1. Bijvoorbeeld wanneer mensen lid zijn van één of meerdere commissies of verenigingen, moet dat dan geannoteerd worden als beroep of niet? Of alleen als er duidelijk staat: bestuurslid of secretaris of voorzitter?

Antwoord: hoewel de tijdsinvestering voor activiteiten in verenigingen en commissies waarschijnlijk niet onderdeel voor die van een ‘officieel’ beroep zou ik zeggen dat ze niet aangemerkt hoeven te worden als beroepen, gezien die in mijn optiek omschreven kunnen worden als activiteiten die mensen primair ondernamen om in hun levensonderhoud te voorzien.

2. Aansluitend hierop, soms staat er niet direct achter het woord "lid" of "voorzitter" van welke commissie of vereniging de persoon lid of voorzitter is, maar iets verderop. Moet ik dan alles wat daartussen staat ook annoteren of alleen de woorden "lid" of "voorzitter"?

Antwoord: zie boven. Ik denk niet dat leden/voorzitters gelden als beroepen. Mocht je twijfelen, dan is het handig om ook de naam van de vereniging/commissie te annoteren. Als daar zaken tussen staan lijkt me dat die ook meegenomen moeten worden, gezien "lid" of "voorzitter" te algemeen is om op zichzelf een beroep te vormen.

3. Moet ik alleen de eerste keer dat een persoon of beroep (als het beroep door dezelfde persoon wordt uitgeoefend) genoemd wordt, deze annoteren of alle keren? En alleen wanneer voor- en achternaam genoemd worden of is een voor- of achternaam genoeg?

Antwoord: Ja! Alle keren dat een persoon+beroep voorkomt annoteren met zoveel mogelijk informatie, dus voor en achternaam.

4. Moet ik een beroep ook annoteren als de persoon die dit beroep uitvoert niet expliciet genoemd wordt?

Antwoord: Ja. Hierbij zal dan geen relatie voorkomen. Inderdaad geen relatie maar wel 'NotFamily' als attribute

5. Soms wordt een bepaald beroep niet duidelijk vermeld, maar wordt wel duidelijk dat een persoon bijvoorbeeld ook schrijver was, omdat er verteld wordt dat hij een

aantal werken heeft gepubliceerd. Moet ik iets daarvan dan annoteren als zijnde een beroep of deze omschrijving helemaal buiten beschouwing laten?

Antwoord: Dit moet denk ik per geval bekeken worden. In het geval van een schrijver zou ik "schrijver" als beroep annoteren.

6. Soms maak ik ook meerdere relaties, bijvoorbeeld wanneer er staat dat "Pietje is hoofdcommissaris geworden, hij trad daarbij in de voetsporen van Dirk." Dan markeer ik "hoofdcommissaris" als beroep van het BioSubject, maar ik maak zowel een relatie van Pietje naar hoofdcommissaris als van Dirk naar hoofdcommissaris. Is dit goed of moet ik dan alleen een relatie maken van het BioSubject naar het beroep?

Antwoord: Dat lijkt me perfect! 'In de voetsporen van' geeft natuurlijk wel wat onzekerheid (misschien was Dirk wel algemeen agent, of had hij een andere functie die als gelijksoortig werd gezien), maar dat lijkt me in dit geval geen reden om het niet te annoteren.

7. In de zinsnede 'Wel ontmoette hij hier de hoogleraren Theodorus van Kooten, een taalkunde en dichter', markeer ik zowel taalkundige, dichter en 'hoogleraren' als beroep van Theodorus van Kooten. 'Hoogleraren' is natuurlijk niet een beroep, 'hoogleraar' wel. Is dit een probleem of moet ik gewoon 'hoogleraren' annoteren en linken aan Theodorus?

Antwoord: Deze viel mij ook op. 'Hoogleraren' verwijst in dit geval naar Theodorus en naar nog een tweede persoon die ook genoemd wordt. Mij lijkt 't het beste om 'hoogleraren' als beroep te noteren en beide personen eraan te koppelen met als attribute 'NotFamily'. Dit zou wel betekenen dat in de analyse meervoud en enkelvoud gladgestreken moeten worden.

8. Het beroep koopman is vrij algemeen. Als er staat 'koopman in olie en traan', moet ik dat dan in zijn geheel annoteren of alleen 'koopman'?

Antwoord: Ik denk het wel. Deze zin in de handleiding lijkt er wel op te wijzen namelijk: "mark general occupations with further details "directeur van Diamantslijperij Asscher": mark "directeur van Diamantslijperij Asscher" in full"

9. Moet ik in de zinsnede 'Veel Amsterdamse Patriotten waren namelijk beurshandelaren en kooplieden', de beroepen markeren? En zo ja, moet ik er attributen aan koppelen of niet?

Antwoord: Ik zou zeggen van niet. Het lijkt om een groep te gaan en het is niet duidelijk of de hele groep bestaat uit beurshandelaren.

Aanvulling: Na overleg denken we dat de laatste regel van de guidelines hier van toepassing is. 'The person is not mentioned: "hij woont naast het gebouw van de Unie van bakkers": mark 'bakkers' as occupation but don't link 'bakkers' to a person'. Volg deze regel dus.

10. Frasen als 'eigenaar van een eigen bedrijf' en 'leiding van de gehele onderneming' markeer ik als beroep. Is dit fout? Met andere woorden: moet je het alleen markeren als er letterlijk een beroep in het enkelvoud staat, of ook als er heel duidelijk naar verwezen wordt?

Antwoord: we denken dat dit soort beschrijvingen gedetailleerd genoeg zijn om door te gaan als beroep (ze zeggen immers wel iets over sociaal-economische status)

11a. Bij de zinsnede: 'buitengewoon hoogleraar in de plantenziektekunde', weet ik niet wat ik moet annoteren. We hebben namelijk gezegd dat we bij hele algemene beroepen (zoals directeur) ook te noteren op welke institutie de beroepen betrekking hebben. Nu is hoogleraar niet zo algemeen als directeur, maar is het toch ook wel redelijk algemeen. Moet ik 'in de plantenziektekunde' dus annoteren of niet? Tweede deel van mijn vraag: 'buitengewoon' annoteren of niet? Wezenlijk verschillen buitengewoon hoogleraar en hoogleraar namelijk niet echt.

Antwoord: het lijkt ons het beste de hele zinsnede te annoteren. "buitengewoon hoogleraar" is een duidelijk beroep. Daarnaast lijkt de specificatie (in de plantenziektekunde) ons nuttig om mee te nemen. Dit voorbeeld is vergelijkbaar met een andere in de guidelines:

"mark general occupations with further details "directeur van Diamantslijperij Asscher": mark "directeur van Diamantslijperij Asscher" in full".

11b. "Hij ving in 1844 zijn studie voor ingenieur aan". Moet ingenieur hier als beroep worden geannoteerd? Zo ja, moet die dan gelinkt worden aan "hij" of aan degene naar wie "hij" verwijst die met naam en toenaam wordt genoemd?

Antwoord: We denken dat dit afhangt van de situatie. Is hij daadwerkelijk ingenieur geworden of blijkt uit de rest van de biografie dat hij een ander beroep is gaan beoefenen? In dat laatste geval lijkt het ons het beste om het niet te annoteren.

Correctie: 'ingenieur' moet hier sowieso als beroep worden geannoteerd. De vraag is eerder of het aan de persoon gekoppeld moet worden. Ik zou dit hier sowieso niet doen, omdat dat hier nog niet staat. Deze koppeling kan gemaakt worden op het moment dat ook expliciet genoemd wordt dat hij als ingenieur aan de slag gaat.

12a. Vraag n.a.v. deze vraag:

Soms maak ik ook meerdere relaties, bijvoorbeeld wanneer er staat dat “Pietje is hoofdcommissaris geworden, hij trad daarbij in de voetsporen van Dirk.” Dan markeer ik "hoofdcommissaris" als beroep van het BioSubject, maar ik maak zowel een relatie van Pietje naar hoofdcommissaris als van Dirk naar hoofdcommissaris. Is dit goed of moet ik dan alleen een relatie maken van het BioSubject naar het beroep?

Antwoord: Dat lijkt me perfect! ‘In de voetsporen van’ geeft natuurlijk wel wat onzekerheid (misschien was Dirk wel algemeen agent, of had hij een andere functie die als gelijksoortig werd gezien), maar dat lijkt me in dit geval geen reden om het niet te annoteren.

12b.

Als ik voor meer personen een relatie maak naar hetzelfde beroep in een dergelijke constructie, kan ik niet voor de afzonderlijke relaties een andere practicioner annoteren. Dus ik kan in dit geval alleen allebei als BioSubject of als NotFamily annoteren. Kortweg: ik kan niet twee verschillende attributes van één profession annoteren. Wat moet ik in zo'n geval doen?

Antwoord: Goed punt! Hier moeten we iets op bedenken. Voor nu leek het ons het beste om de eerste relatie te annoteren.

13. “Hij volgde hier onder meer college bij hoogleraar Frederik Willem Pestel, een specialist in natuurrecht.” Moet ik in deze zin als beroep alleen “hoogleraar” annoteren? Of ook “specialist in natuurrecht”?

Antwoord: “specialist (...)” is hierbij geen onderdeel van het beroep, maar meer een eigenschap die de persoon naast zijn beroep heeft. In onze ogen dus niet annoteren.

14. Moet ‘machthebber’ in ‘de Franse machthebber Napoelon Bonaparte’ worden geannoteerd als beroep?

Antwoord: alleen annoteren als het woord machthebber duidelijk staat voor een meer formele categorie. In dit geval zou je kunnen kiezen voor wel annoteren, gezien het duidelijk om Napoleon als staatshoofd gaat.

15. Moet een beroep ook geannoteerd worden bij de volgende constructie: “hij was in ... werkzaam” (een voorbeeld is “ambtsbetrekkingen”. Het lijkt hier dan namelijk vaak om een bredere beroepsgroep te gaan.

16. Wanneer ‘Maria’ (de moeder van Jezus) of andere personen worden genoemd die ten tijde van de biografie al lang niet meer leven én een religieuze connotatie hebben, annoteer ik het niet. Is dit de bedoeling? En bij figuren die al lang niet meer bestaan,



maar die wel 'historisch worden gehanteerd' (stel je de volgende fictieve zin voor: 'Maria - geboren 35 v.C. - komt uit Israël')? Snappen jullie wat ik bedoel, en is de vraag relevant?

Antwoord: personen moeten alleen geannoteerd worden als hun beroep expliciet genoemd wordt. Als er ergens staat dat Maria ook huisvrouw was of Jozef timmerman, dan wel (en dan maakt het ook niet uit dat ze wel of niet leven, in een andere tijd, etc.). In het voorbeeld dat je geeft, staat geen beroep en moet Maria dus ook niet geannoteerd worden.

17. Het fragment dat eindigt op '112\_05', is een telegramstijl biografie over dezelfde persoon die in een ander tekstfragment ook wordt behandeld (Willem Sluiter). Uit het 112\_05-fragment zelf kan ik niet opmaken wat zijn beroep is (zonder context en verheldering om welke persoon het gaat staat er plots: 'Geref. Pred.'), maar met voorkennis van het andere fragment kan ik weten dat het betekent dat Willem Sluiter Gereformeerd predikant is geweest. Moet ik 'Geref. Pred.' annoteren als het beroep van Willem Sluiter op basis van mijn voorkennis of moet ik elke tekst kritisch op zichzelf bekijken en dus doen alsof ik niet weet waar 'Geref. Pred.' voor staat?

Antwoord: inderdaad wel annoteren. Deze bron maakt veel gebruik van afkortingen. Als je niet weet wat een afkorting betekent, maar er voldoende reden is aan te nemen dat het om een beroep gaat, graag navragen.

18. Suggestie: zou het in de toekomst mogelijk zijn hotkeys in te stellen voor het aanmaken van attributen? Voor het aanmaken van één attribuut moet je namelijk vier keer klikken. Hotkeys (waar mogelijk) hiervoor zou het annoteren een stuk efficiënter maken.

19. Als er een beroep/functie genoemd staat en deze wel gerelateerd is aan een persoon, maar de naam van deze persoon niet genoemd wordt, wel een relatie tussen deze twee leggen. Of überhaupt niet markeren als persoon?

Antwoord: Losse beroepen moeten ook worden geannoteerd.

20a. Als er een beroep wordt genoemd maar er staat dat de persoon (BioSubject) deze functie afwijst, dan wel het beroep annoteren? Welk attribuut ken je het beroep dan toe, alsnog BioSubject?

Antwoord: Wel annoteren, maar niet koppelen aan biosubject. Zie vorige vraag.

20b. In de volgende zin ben ik niet zo zeker van wat ik als beroep moet annoteren: "Kort daarvoor was hij in de rang van reserve-kolonel opnieuw in militaire dienst getreden als

hoofd van de sectie kunstbeschermingsofficieren.” Ik twijfel over reserve-kolonel, militaire dienst en kunstbeschermingsofficieren.

Antwoord en: ik zou gaan voor “hoofd van de sectie kunstbeschermingsofficieren”. Binnen dit beroep is hij dan “reserve-kolonel”, maar de eerste lijkt me belangrijker.

21. Als er een beroep wordt genoemd, en er geen persoon bij staat om het beroep aan te linken, moet ik dan een attribuut toevoegen? En zo ja, dan is het attribuut ‘NotFamily’ neem ik aan?

Antwoord: Losse beroepen moeten worden geannoteerd en inderdaad met NotFamily.

22. Als het goed is moeten we altijd de naam die het dichtst bij het beroep staat annoteren. Echter, soms staat de naam van het BioSubject niet volledig in het verloop van de biografie, maar staat de volledige naam wel bovenaan (verder weg van het beroep). Moet ik dan de volledige naam annoteren of de kortere naam die dicht bij het beroep staat?

Antwoord: We koppelen lokaal. Annoteer dus de dichtstbijzijnde verschijning van de bijbehorende persoon (kan ook ‘hij’ of ‘zij’ zijn)

23. Vorige week was er bij mij wat fout gegaan met te veel personen annoteren. Dus om voor mij nu alles helder te hebben nog een keer een vraag over personen en beroepen annoteren. Klopt het dat elk persoon maar één keer geannoteerd moet worden, en dat alle beroepen die verder voorkomen in een biografie dus gelinkt moeten worden aan die ene (persoon)annotatie, ook als dat beroep vele zinnen later voorkomt?

Antwoord: Nee, we annoteren lokaal. Koppel dus het beroep aan de dichtstbijzijnde verschijning van de bijbehorende persoon (kan ook ‘hij’ of ‘zij’ zijn). We hebben hiervoor gekozen omdat dat beter strookt met de software die het straks gaat analyseren.

24. Uit de volgende zin wordt duidelijk dat de persoon garnalenvisser was, maar het staat er niet letterlijk als beroep. Moet ik het dan alsnog annoteren, zo ja, wat moet ik dan annoteren?

“Om zijn lage pensioen aan te vullen, legde Naerebout zich daarom toe op de garnalenvisserij. Twee van zijn kleinkinderen assisteerden hem daarbij.”

Antwoord: Klopt, “garnalenvisserij” dien je te annoteren. Zie ook vraag 15 en het antwoord van Richard in de kantlijn.

25. Dezelfde vraag voor de volgende zin: “De lettertypes verkreeg hij van Fransiscus, die Hebreeuws en Aramees doceerde.” Hieruit wordt duidelijk dat Fransiscus docent

Hebreeuws en Aramees was, maar er staat niet 'docent' als beroep. Moet dit alsnog geannoteerd worden?

Antwoord: Niet annoteren: we annoteren geen werkwoorden (produceren, bakken, diamantslijpen) ook al lijkt het op een beroep.

26. Ik heb het gevoel dat er iets geannoteerd moet worden in de volgende zin, maar ik weet niet zo goed wat, omdat het gaat over het vervullen van functies op drie verschillende plekken die niet perse los van elkaar geannoteerd kunnen worden: "Rotterdammer die functies vervulde op een accountantskantoor, een ministerie en bij de gemeente Rotterdam." Moet ik hier iets annoteren en zo ja wat?

Antwoord: Zie vraag 15. 'Functies vervullen...' valt onder dezelfde categorie als 'werkzaam in' en die annoteren we.

27. (1) In de zinsnede 'Zeker is dat voor een veldwachter als Hendrik Schoonakker (...)', annoteer ik veldwachter niet, omdat het de hele tijd al gaat over Hendrik en zijn veldwachter zijnde, en aangezien er hier wordt gesproken over Hendrik 'als veldwachter'; redelijk indirect dus. Is dit correct of moet ik dit wel annoteren? (2) Daarnaast wordt er in datzelfde fragment telkens gesproken over 'de veldwachter', waarbij het duidelijk is dat het over Hendrik gaat, en wordt er gesproken over 'de burgemeester', waarbij het ook telkens duidelijk is dat het om een persoon gaat die in de biografie eerder genoemd is. Moet ik 'de veldwachter' en 'de burgemeester' telkens annoteren en linken aan de personen die (al worden ze niet in een zin genoemd) hiermee bedoeld worden?

Antwoord: "een veldwachter als Hendrik" wijst erop dat Hendrik een veldwachter was. Gewoon annoteren als beroep dus.

28. Wanneer moet ik in een beroep een bepaalde plaatsnaam of gebiedsnaam meenemen? Wanneer er bijvoorbeeld staat: Franse minister van Buitenlandse Zaken, annoteer ik het helemaal, omdat het (denk ik?) relevant lijkt voor de baan onder welk gebied je baan valt. Moet ik dit wel of niet doen? In mijn huidige annotatie-ronde stuitte ik zojuist op de frase 'gemeentesecretaris van Nijmegen'; alles annoteren of alleen gemeentesecretaris?

29. Wat te doen bij 'hoogleraar in het bestuursrecht, [vakgebied], [vakgebied], [vakgebied]'? Alles in één annotatie stoppen of alleen het eerste vakgebied annoteren? Ik heb voor het eerste gekozen.

Antwoord: Zie vraag 11, alles erin (vb. "Hoogleraar in de Oosterse Talen uitgezonderd Hebreeuws" helemaal, is allemaal deel van de officiële titel)

30. Vervolg vraag 10: Wat te doen in de volgende zin: ‘Door Van der Heijden is hij ook, in 1940, betrokken geraakt bij de redactie van het Maandblad De Naamloze Vennootschap, waaraan hij tot aan de dag van zijn overlijden leiding heeft gegeven.’ Hier volgt duidelijk uit dat ‘hij’ (verwijst naast BioSubject) leiding heeft gegeven aan de redactie van een blad, al is het verstoep in een hele lange zin. Annoteren of niet? Ik heb gekozen voor niet annoteren, omdat de zin te lang is en er anders te veel irrelevante elementen worden geannoteerd.

Antwoord: zie vraag 15, redactie zou je kunnen bestempelen als een beroepsgroep. Gezien het feit dat hij expliciet als deel (en zelfs leider) van die groep wordt aangemerkt kan je “redactie van het Maandblad De ... Vennootschap” annoteren. (hoewel hier ook de vraag speelt of de naam van het Maandblad meegenomen moet worden)

31. Moet ik in de volgende zin ‘schrijver’ als beroep annoteren: “hij begon als lokettist bij het postkantoor, om zich in zijn vrije tijd te ontwikkelen als schrijver.” Schrijver is hier als vrijetijdsbesteding genoemd, maar het is wel een beroep. Moet ik BioSubject als attribute toekennen of alleen annoteren als algemeen beroep? Of helemaal niet annoteren?

Antwoord: schrijver wel annoteren (als BioSubject ook). Het gaat om een beroep, zij het wat secundair. Als er gestaan had “in zijn vrije tijd schreef hij” niet.

32. “Meijer Bleekrode kwam uit een diamantbewerkerfamilie”. Moet ik diamantbewerkerfamilie hier annoteren?

Antwoord: gun deze man z’n agency ;) niet annoteren! Evt. zou je “diamantbewerkerfamilie” los als beroep kunnen annoteren en de attribute “FamilyMember” mee kunnen geven

33. Wat moet ik als beroep annoteren in de volgende zin: “... moest hij in het diamantvak werken”. Ik heb nu diamantvak geannoteerd.

Antwoord: zie antwoord van Richard bij vraag 15; beroepsgroepen annoteren (en diamantvak zou je kunnen zien als beroepsgroep)

34. Moet ik ‘notabelen’ annoteren als beroep? Ik twijfel hierover omdat het wel een status aangeeft die voortkomt uit bepaalde gezaghebbende functies.

Antwoord: niet annoteren, wijst inderdaad op status, niet op beroep

35. Wat te doen in zinnen als deze: “hij was in 1940 buitengewoon en in 1945 gewoon hoogleraar in de sterrenkunde”. Ik heb nu alleen “gewoon hoogleraar in de sterrenkunde” geannoteerd, want “buitengewoon” op zichzelf is niks. Maar zo gaat er wel een beroep verloren. Dit komt vaker voor, ook bijv. “hij was lid van de gemeenteraad en daarna van de Tweede Kamer” of “hierna was hij nog zeven jaar Tweede en Eerste Kamerlid”.

Antwoord: Zie vraag 11 & 29. Titels van hoogleraren in z'n geheel annoteren. “gewoon” lijkt daar bij te horen (omdat naar ons weten “buitengewoon hoogleraar” ook een titel is).

36. Valt stiefvader onder OtherFamilyMember of NotFamily? Ik heb het nu als OtherFamilyMember geannoteerd.

Antwoord: Dat klopt, het is meer wel familie dan niet familie ;)

37. Moeten alle titels worden geannoteerd? Koning moet wel, edelman/jonkvrouw ook, maar titels als baron/barones etc.?

Antwoord: Aangezien we koning, edelman en jonkvrouw annoteren, lijkt het ons goed om dit ook mee te nemen.

38. Tijdens het lokaal annoteren, ook verwijzingen annoteren als ‘die’ of ‘deze’?

Voorbeeld: “Pietje was bevriend met een aartsbisschop. Deze wist veel invloed uit te oefenen op...” Aartsbisschop = beroep. Deze = person? Ander voorbeeld: “Pietje was een zeer betrokken docent, die...” docent = beroep. die = person? Want ‘die’ staat dichterbij dan ‘Pietje’.

Antwoord: Hierop weten wij het antwoord niet. Het lijkt ons te rommelig, maar misschien kan het wel.?

39. Waarschijnlijk gaat dit probleem niet vaak voorkomen dus is het niet relevant, maar ik heb nu dus een biografie voor me waarin twee tweelingbroertjes worden besproken, die beiden als het onderwerp (biosubject) worden behandeld gedurende de gehele biografie. Ik doe dus alsof ze allebei biosubject zijn en markeer beide namen niet. Is dit oké?

Antwoord: Die vraag zal inderdaad vrij uniek zijn. Ik denk dat het goed is zoals je het doet.

40. Wanneer een beroep zowel verwijst naar een biosubject als een non-biosubject, welk attribuut moet ik dan toevoegen? Blijft het onder de nieuwe regels hetzelfde en moet ik

dan het attribuut toevoegen dat hoort bij de eerste persoon die in de zin genoemd wordt?

Antwoord: Als ik me niet vergis, is het mogelijk om twee attributen/relaties toe te voegen. Zo niet, dan lijkt het me logisch om de dichtstbijzijnde persoon te annoteren.

41. Onder de nieuwe regels moeten we (als ik het goed begrijp) alle beroepen die in een biografie staan, annoteren. Wanneer er echter staat "X nam ontslag als procureur-generaal" (waarbij X staat voor biosubject), moet ik dan procureur-generaal annoteren met als attribuut biosubject? De biografie duidt immers aan dat hij de functie niet meer gaat vervullen.

Antwoord: Ja, gewoon annoteren.

42. Moet ik in het volgende geval annoteren?: "Hoewel Rutgers geen politiek leider of marxistische theoreticus was, (...)". 'Politiek leider' heb ik nu geannoteerd met als attribuut 'NotFamily'.

Antwoord: Lijkt me goed.

43. Wanneer er staat (v.b.) "dr. Jan Jansen was arts.", moet ik 'dr.' dan meenemen bij zijn naam?

Antwoord: Hier lijkt me deze regel van toepassing:

\* splits titel en naam

== vb. "Zijn vader was generaal [beroep] Thijssen [persoon]" – Attribute = Father

Het beroep staat er in dit geval achter, dus ik denk dat je de titel in zijn geheel weg kan laten.

44. Vraag 38 van over verwijzingen is een vraag die ik ook telkens heb als ik aan het annoteren ben. "De vader van Katz was een in zijn tijd vermaarde, links-georiënteerde advocaat en procureur." Elders eerder wordt deze "vader van Katz" bij naam genoemd. Maar aangezien we lokaal annoteren, heb ik "de vader van Katz" geannoteerd als 'Person'. Klopt dit?

Antwoord: Ik denk dat dat prima is.

45. Moet ik Prins annoteren als beroep als het iemands achternaam is?

Antwoord: Dat hoeft niet, want het is een achternaam.