SEMANTIC WEB

DISEASES ONTOLOGY  
Een overzicht van ziektes

Jochem van Dooren (JDN272)  
Jimi Cornelisse (JCE340)

1. **Domain modeling***1.1 Doel  
   1.2 Doelgroep*
2. **Data Reuse**  
   *2.1 Data Sources  
   2.2 Protégé  
   2.2.1 Object Properties*
3. **Applicatie**  
   *3.1 Algemeen  
   3.2 Queries  
   3.3 Grafische weergave  
   3.4 Evaluatie*

**Domain Modeling**

*1.1 Het idee*Onze applicatie zal zich focussen op ziektes, de symptomen van deze ziektes en

werkvelden/specialisaties van de ziektes. Het doel is om een database te verkrijgen waar gebruikers kunnen zoeken naar symptomen, soorten ziektes en sterftecijfersen daarmee uiteindelijk een overzicht zullen krijgen van welke ziektes en/of aandoeningendeze eigenschappen hebben. Een andere mogelijkheid is het krijgen van een overzicht van ziektes die een laag sterftecijfer hebben, maar ook nog eens een zeldzame ziekte zijn. Hierdoor kunnen er dus interessante correlaties opgevraagd worden door de gebruiker, maar kan het ook heel simpel blijven.

Met behulp van een grote dataset, zelf toegevoegde informatie, python, jquery en protégé kan er veel informatie over ziektes verwerkt worden in een ontology. Het domein van de ontologie zal vooral gaan over het representeren van ziektes en aandoeningen met daarop aansluitend de werkvelden en symptomen. Als voorbeeld kunnen Waterpokken genomen worden. ChickenPox valt onder verschillende ‘subjects’, bijvoorbeeld onder Pediatrics (het werkveld) enViralDiseases (overkoepeling). De applicatie zal vooral door mensen met weinig achtergrond kennis van de medische wereld gebruikt worden en om deze reden is er getracht om zo weinig mogelijk lastige medische benamingen voor de ziektes en symptomen te gebruiken, maar dit is in de praktijk niet altijd mogelijk. Echter zal Malaria gewoon malaria heten in de ontologie in plaats van de wetenschappelijke naam Plasmodium Falciparum.   
De applicatie moet meer dan enkel dingen oprakelen uit een database. De gebruiker moet meerdere restricties kunnen geven die door middel van inferencing zullen leiden tot nieuwe niet eerder bedachte resultaten. Er is dus een ontologie nodig om de juiste verbanden te kunnen maken binnen de dataset. Elke ziekte en aandoening heeft vele verschillende eigenschappen die allemaal op een manier met elkaar verbonden zijn en die elkaar kunnen aanvullen om tot nieuwe inzichten te komen. Juist om die reden is het gebruik van Semantic Web nodig bij de applicatie.  
  
*1.2 Doelgroep*De applicatie zal een zeer brede doelgroep kennen die niet goed samen te vatten is door een overkoepelende groep. De gebruikers kunnen dokters, patienten, familie van patienten, geneeskunde studenten, maar ook nieuwsgierige mensen zijn. Het kan als naslagwerk gebruikt worden voor studenten en voor patienten kan het knooppunten geven waarmee ze hun kennis verder kunnen uitbreiden en voor de nieuwsgierigen kan het interessant zijn om verbanden te vinden die niet zo erg voor de hand liggen, bijvoorbeeld het uitzoeken welke ziektes ook als biologische wapens gebruikt kunnen worden.

**Data Reuse**

*2.1 Data Sources*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | subject | subject | subject | subject | subject | subject | Subject |
| Ebola Virus Disease | AnimalVirology | ViralDiseases | Hemorrhagic Fever | Zoonoses | TropicalDisease | BiologicalWeapon | AnimalDiseases |

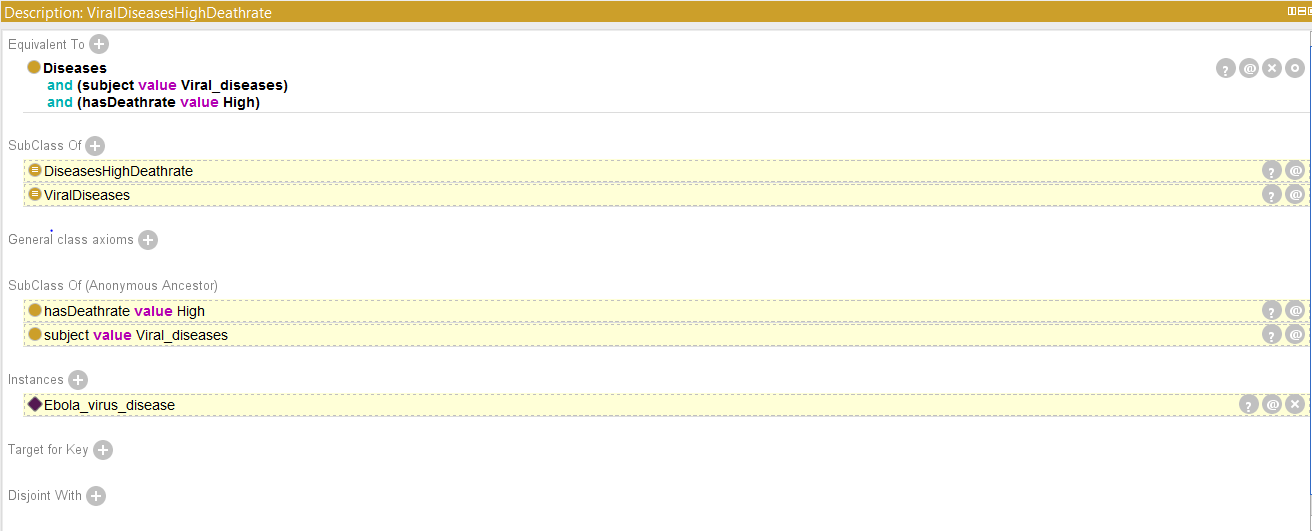
De applicatie maakt gebruik van een extensieve dataset, waarin er ruim drie druizend ziektes zijn toegevoegd. Aan deze ziektes zijn ook nog eens werkvelden en symptomen toegevoegd, waardoor de dataset rijk aan informatie wordt. Deze data komt van DBPedia, waaruit dus alle ziektes en hun subjects genomen zijn. We hebben gekozen voor deze dataset, omdat de dataset zeer uitgebreid is. Doordat er zoveel instances zijn, kunnen we de applicatie altijd nog verder uitbreiden zonder al teveel obstakels tegen te komen. Hieronder is een voorbeeld van de subjects die allemaal bij het Ebola Virus Disease horen.

Zoals te zien is, ontstaat er een scala aan verschillende aanknopingspunten die tot interessante inferencing kunnen leidenzo zou er bijvoorbeeld gekeken kunnen worden naar welke ziektes zowel een BiologicalWeapon zijn als een TropicalDisease en zodra deze restricties gegeven zijn dan kan de computer met behulp van inferencing aangeven dat Ebola Virus Disease aan deze restricties voldoet en dat er eventueel andere ziektes daar ook bij horen. De applicatie herbruikt de verschillende subjects dus en kan vervolgens hiermee links leggen tussen data.   
  
De tweede dataset is handmatig ingevuld en is niet compleet voor alle ziektes, omdat hier geen volledige dataset voor te vinden was. Het importeren van de letterlijke data was niet ideaal voor de toepassing van deze data dus is de keuze gemaakt om het handmatig te doen. Sterftecijfers zijn interessant in verband met ziektes en daarom is de keuze gemaakt om deze sterftecijfers toe te voegen aan een aantal ziektes[[1]](#footnote-2). De exacte sterftecijfers zijn niet per definitie heel erg nuttig, omdat er dan niet gemakkelijk gegroepeerd kan worden. Daarom zijn de sterftecijfers in drie categorieën opgesplitst; High, Medium en Low. De categorieën zijn als volgt opgesplitst:

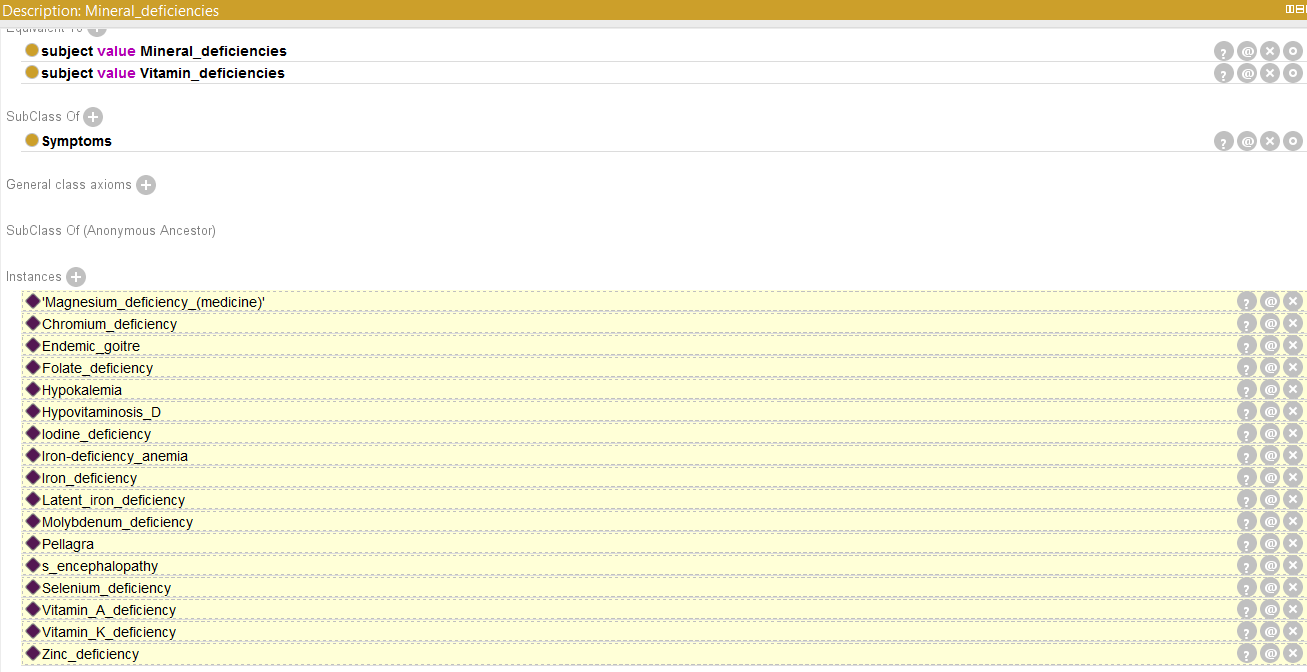
|  |  |
| --- | --- |
| HIGH | 50+ doden per 100.000 |
| MEDIUM | 5-50 doden per 100.000 |
| LOW | 0-5 doden per 100.000 |

De ziektes zijn dus in meerdere klasses opgesplitst en vervolgens hebben deze bij de ziektes van toepassing een extra restriction genaamd hasDeathrate gekregen, waar de waarde vervolgens High, Medium of Low is. Hiermee kunnen ook nieuwe verbanden geinferenced worden. Zo is het mogelijk om InfectiousDiseases met een High deathrate te kunnen inferencen.

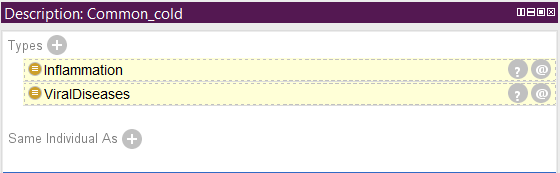
*2.2 Protégé*

Een paar voorbeelden van deze inferencing in Protégé staan hieronder afgebeeld:  
  


Hierboven is te zien dat onder ViralDiseases die een hoge deathrate kennen, Ebola Virus disease valt. Bovendien is het natuurlijk logisch dat ViralDiseaseswith High deathrate een subclass is van zowel ViralDiseases als Diseases met een hoge deathrate.

Er kan ook geinferenced worden op basis van de symptomen in de afbeelding hieronder zijn in protégé bij het symptoom MineralDeficiencies de ziektes en aandoeningen die dit symptoom hebben geinferenced:

De inferencing bij de individual Common Cold ziet er als volgt uit:



Er kan vervolgens een query geschreven worden waarmee Common Cold opgeroepen kan worden door een aandoening te hebben die zowel in de categorie van ViralDiseases valt als het symptoom van Inflammation heeft.  
  
 *2.2.1 Object properties*Subject**:** Subject is de object property waar de meeste data mee gelinkt is. Hiermee wordt dus de relatie aangegeven tussen twee objecten in de ontologie. Onder subject vallen de categorieën van de ziektes en de symptomen.  
  
hasDeathrate: hasDeathrate is een object property die aangeeft welke van de drie categorieën (High, Medium, Low) een ziekte heeft op basis van de deathrate van de ziekte

**Applicatie:***3.1 Algemeen:*De applicatie zal doormiddel van de gebruiker door middel van *checkboxes*de mogelijkheid geven om ziektes op te halen. Er kunnen meerdere checkboxes aangeklikt worden waardoor de resultaten specifieker zullen worden. In de screencast zal duidelijker worden hoe de applicatie precies werkt.   
  
*3.2 Queries*

De prefixes die gebruikt zijn voor alle queries zijn de volgende:

*PREFIX dbr: <http://dbpedia.org/resource/>*

*PREFIX dct: <http://purl.org/dc/terms/>*

*PREFIX dbp: <http://dbpedia.org/property/>*

*PREFIX dbc: <http://dbpedia.org/resource/Category:>*

*PREFIX db: <*[*http://dbpedia.org/*](http://dbpedia.org/)*>*

*PREFIX disease: <http://www.semanticweb.org/jimi/ontologies/2015/9/Ontology1443989410323.owl#>*

De construct query die gebruikt is om de data van dbpedia te kunnen gebruiken in onze eigen ontologie:

*CONSTRUCT*

*{*

*?disease a dbo:Disease .*

*?disease dct:subject ?subject.*

*}*

*WHERE {*

*?disease a dbo:Disease .*

*?disease dct:subject ?subject.*

*}*

Vervolgens nog een aantal queries die de applicatie gebruikt:

In deze query worden alle ziektes gezocht die een biological weapon disease zijn, maar ook een viral disease en een hoge death rate hebben. Deze query is hardcoded in de applicatie.

SELECT ?disease WHERE

{

?disease a disease:BiologicalWeaponDiseases.

?disease a disease:ViralDiseases.

?disease a disease:DiseasesHighDeathrate.

}

De volgende query is ook hardcoded en zoekt naar ziektes die een neurological disorder zijn, een rare disease en als symptom pain hebben.

SELECT ?disease WHERE

{

?disease a disease:NeurologicalDisorders.

?disease a disease:RareDiseases.

?disease a disease:Pain.

}

Als laatste nog een stuk code uit onze jquery file waarin gekeken wordt welke checkboxes aangevinkt. Op basis van de aangevinkte checkboxes wordt vervolgens een query geconstrueerd. In de onderstaande code wordt infectious diseases tegenover deathrates gezet.

if (checkbox9 == true && checkbox2 == true){

$('#QUERYBOX').html(SELECT ?disease WHERE {?disease a disease:DiseasesHighDeathrate. ?disease a disease:InfectiousDiseases}');

}

if (checkbox8 == true && checkbox2 == true){

$('#QUERYBOX').html(SELECT ?disease WHERE {?disease a disease:DiseasesMediumDeathrate. ?disease a disease:InfectiousDiseases}');

}

if (checkbox7 == true && checkbox2 == true){

$('#QUERYBOX').html(SELECT ?disease WHERE {?disease a disease:DiseasesLowDeathrate. ?disease a disease:InfectiousDiseases}');

}

*3.3 Grafische weergave*

Voor onze grafische weergave hebben we een screencast gemaakt die alle aspecten van de applicatie meteen belichten. Zie de youtube link hieronder.

https://www.youtube.com/watch?v=pGPzJ-ibK\_k

*3.4 Evaluatie*

Het doel van de Semantic Web applicatie was om ziektes op basis van hun eigenschappen te kunnen groeperen en opvragen. Hierin is de applicatie zonder meer geslaagd. De gebruikte ontologie is rijk aan informatie en er zijn veel bruggetjes die de verschillende data aan elkaar moeten linken. Hierdoor kan er een compleet en overzichtelijk beeld ontstaan van alle data in de ontologie.   
  
We zijn over het algemeen tevreden over het eindresultaat van ons project, maar we zijn wel tegen een aantal problemen opgelopen**.** We hebben veel moeite gehad met het vinden van een dataset waarin een uitgebreid overzicht van sterftecijfers van bepaalde ziektes in zit aangegeven, mede daardoor is er niet veel te inferencen op het gebied van deze sterftecijfers in combinatie met de ziektes en/of symptomen. In een ideale situatie was het mogelijk geweest om elke ziekte een sterftecijfer mee te geven en in het verlengde dus een grotere dataset te hebben om uit te inferencen, maar dit is helaas niet gelukt.  
  
De applicatie zelf, hoewel functioneel, is esthetisch niet hoogstaand. We hebben er binnen het tijdskader voor gekozen om een minder mooie, functionele applicatie te hebben in plaats van een mooie, niet functionele applicatie. Het idee dat de gebruiker kan *mix-and-matchen* tussen de verschillende eigenschappen is wel gelukt en dat is in onze ogen het belangrijkste.

1. <http://www.who.int/whr/2004/annex/topic/en/annex_2_en.pdf> [↑](#footnote-ref-2)