TDT4215 WEB INTELLIGENCE SPRING 2015

GROUP PROJECT REPORT

Query-based system for information about drugs

Introduction

This report will provide information about the group project in the course “Web-intelligens”. Our goal is to create an ontology with information about drugs, which can be queried. We have also going to create an user-friendly interface for the querying and its results.

Motivation

In this section we will describe our ontology and product briefly, as well as our motivation behind these choices.

Our ontology and product

The ontology this group has created addresses medical drugs and their effects. This includes explicit information about symptoms treated by these drugs, active ingredients, adverse effects and dosage. The ontology is based on casuistries provided, containing different persons with multi-chronically illness taking several drugs accordingly.

With the use of this ontology, it is possible to get information about each drug as well as their interactions through a SPARQL-endpoint. With interactions we mean a situation occurring by taking a substance which alters the effect of another drug in use[1]. This could also include various foods.

Our final product is a website where users can search for one or several drugs. Doing a search for only one drug will provide full information about the drug, while searching for several will result in overlapping adverse effects,substance and counteracting effects. There will only be possible to get information about the drugs we have in our ontology, seeing as this will only be used as an example(see next section).

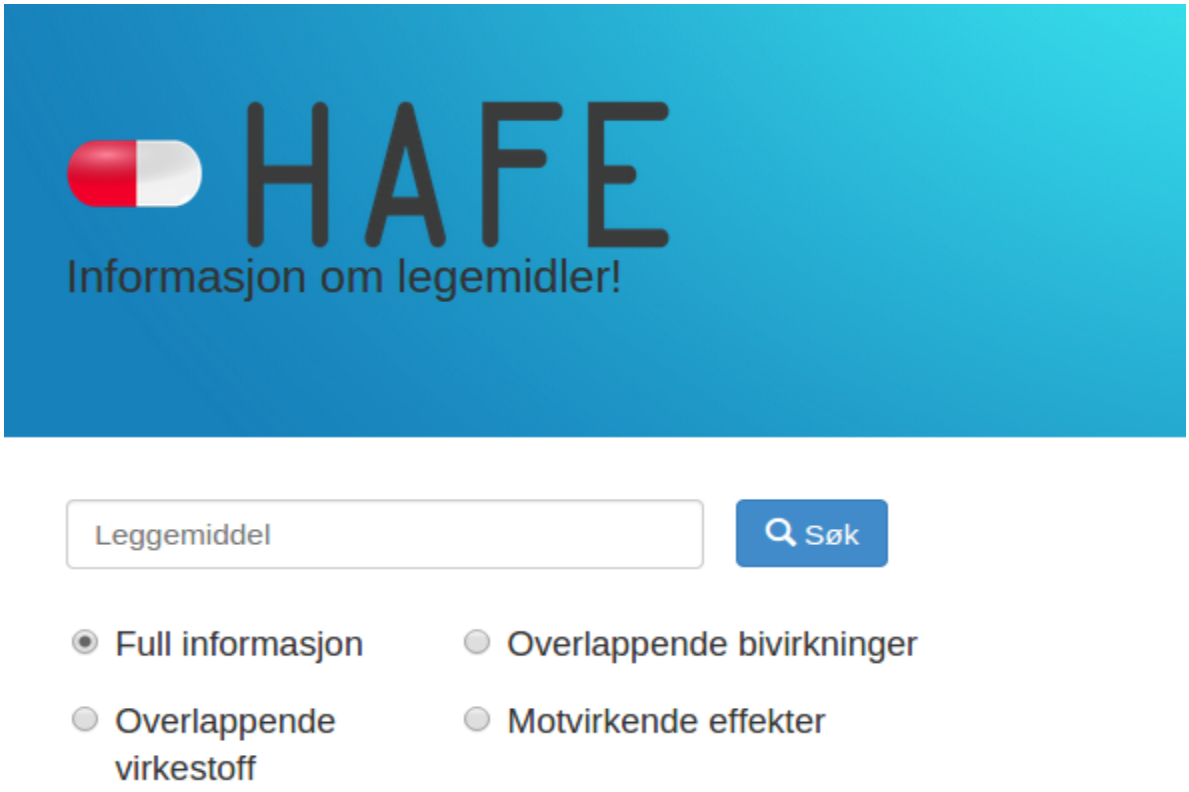

Figure 1: The interface of the website

In hindsight our design of the ontology is probably not the way it should be done with few classes and a lot of individuals. When we looked at other ontologies, we saw that we should/could have done it differently. We decided that even though the design wasn't optimal, we would go through with it due to the lack of time left. This is a lesson learned for later work.

Motivation

Three out of four in our group are interested in health informatics, and had in prior conversations with Øystein Nytrø at the institute about a possible master thesis. He has been engaged with health informatics for decades and have several collaborations within this sector. Øystein has several groups writing master theses for him every semester. After some discussions he came up

with an idea to use this project as a preliminary study in the field of health informatics, as well as helping out a group currently writing their master thesis for Øystein.

The group are looking at how people(typically with an multi-chronically illness) can find information about the medical drugs they take in an understandable fashion. Their result will most likely be prototypes and functional requirements for such a system. Our website and ontology will be aiding the group writing master thesis as examples of how certain parts of their prototypes can be implemented.

Structure and annotation of content

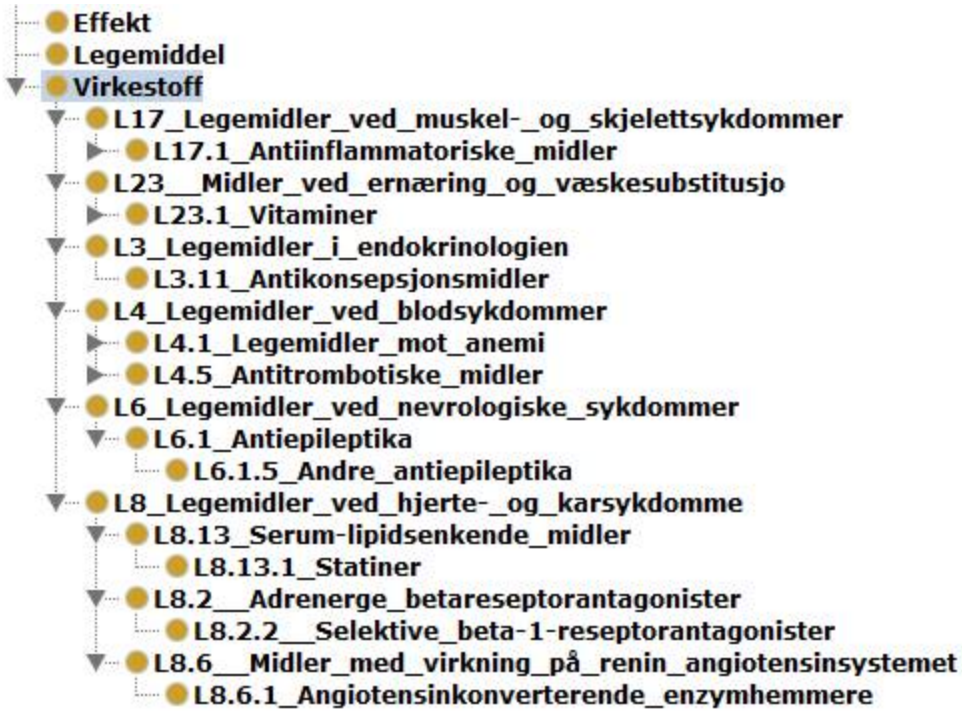

Figure 2: Structure of the ontology in Protégé.

Classes and properties

Effekt

Legemiddel

- harVirkemiddel
- harDose

Virkestoff

- harBivirkning
- harIndikasjon

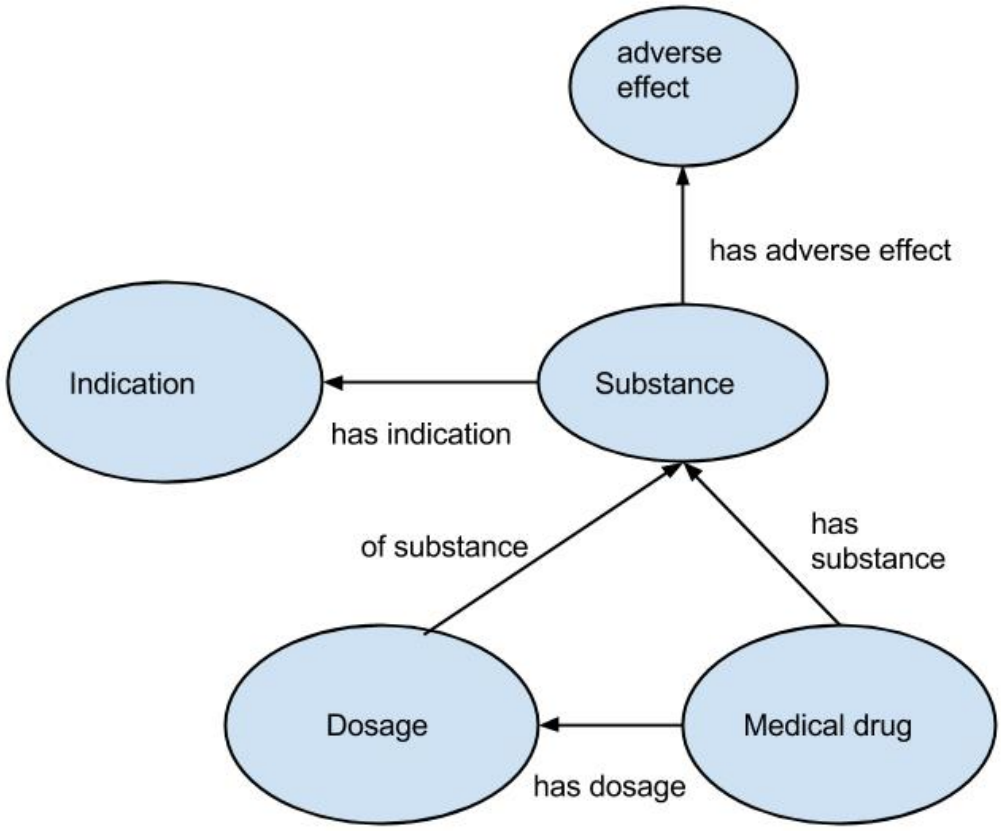

Figure 3: Graph representing the ontology and its relations.

The main resource for a medical drug is its substance. A substance has one or more adverse effects and indications or might even be a child of another substance. In our structure the

“harVirkemiddel” (hasSubstance) is the connection between a medical drug and a substance. The medical drug then inherit the substance adverse effect and indication. A medical drug also has a dosage of substance, this is a simple instance connected with the medical drug with the “harDose” relation.

As you can see on the example above, the “Virkestoff” class (Substance) has several subclasses. Each subclass under this class represents active ingredients, and they are named after which chapter in Legemiddelhåndboka[2](see “Creating local data”) they are taken from. We chose this approach since it gave us great overview. When extracting the information (next chapter), which was tedious work, having all the information labeled from the chapters as classes helped when linking the information to our ontology.

The “Effekt”(Effect) class was a thought we had later about trying to fetch both indications and adverse effects as a class, and then relate them to a substance later. This was just a thought and not really implemented, as of today the effects are relations and members of the class “Virkestoff” (Substance).

Querying the ontology

Here is a query we used to get the adverse effects of the two drugs Keppra and Albyl-e. This query involves using the label annotations for the medical drugs and filter them to lowercase. Next we get the adverse effect of the substance of the drug and it gets checked against the same result form the same query for the other drug.

We could show the queries for getting counteracting effects and overlapping substances as well, but we do not want overdo this report with alot of queries.

Overlapping adverse effect query:

```
SELECT * WHERE {
{select ?bi_label where {
                ?s rdfs:label ?med_label .
                filter (lcase(str(?med_label)) = lcase(str('Keppra'))) .
?s lmh:harVirkemiddel ?virkemiddel .
?virkemiddel lmh:harBivirkning ?bi .
?bi rdfs:label ?bi_label }}
```

```
{select ?bi_label where{
?s rdfs:label ?med_label .
filter (lcase(str(?med_label)) = lcase(str('Albyl-e'))) .
?s lmh:harVirkemiddel ?virkemiddel .
?virkemiddel lmh:harBivirkning ?bi .
?bi rdfs:label ?bi_label }}
        }
```

Implementation and usage of datasets

Pille

In order to make our ontology available on the web we use a Cliopatria-server called Pille[6] which is hosted at the institute. We were given full access and could freely do “whatever we want”. Cliopatria is a semantic web-server built using the swi-prolog language. Pille can also be used as a SPARQL-endpoint to query the different graphs it hosts, which we are doing in this case. There are other graphs hosted on Pille which is in our interest as well. The server hosts a graph representing FEST[4], which we have described in the project description. Simply put, FEST[3] is another data source like Legemiddelhåndboka(see “Creating local data”), but contains more detailed information about medical drugs. This way we will structure data using several sources, also by using another ontology. Our web application is also hosted on Pille, but available on another port. In order to upload our ontology to Pille we had to get familiar with the swi-prolog language which took a while.

Creating local data

Legemiddelhåndboka contains very specific details about anything regarding medicine, and is often used as a reference for doctors, nurses and other health care professional. In our case, this reference contains detailed information about active ingredients and which drugs these contain, with adverse effects, indications and dosage. For our example, some of this information may be too specific or time-consuming to handle in our ontology. After further research, we chose to cut out some of the adverse effects often listed in the chapters for active ingredients. These would

often be adverse effects in connection with e.g pregnancy, diabetics and so forth. This would make our ontology more complex, but since we had to create the instances ourselves and not having an automated process handling it for us, this would take a lot of time only for analyzing the content.

Given the short period of time and the complexity of the domain, we found it sufficient to create the dataset manually by ourselves based on the casuistries. Ideally we would have extracted content from Legemiddelhåndboka automatically, but we quickly figured out that this would be to complex. The master students agreed on this with us. We might look into this later when we are starting our master thesis. When creating the dataset, we went through the casuistries, looked up every drug the different personas were using, and noted down adverse effects, active ingredients, indication(the illness it should cure) and dosage. Later we linked the drug with the information found in our ontology. This took a while and was tedious work, but was needed as the other option would take too much time.
Analyzing, structuring and creating the data based on Legemiddelhåndboka was definitely the most time-consuming part of this project. Furthermore, even though this is perhaps not in the scope of the course, understanding medical terms and the problem domain of the task at hand was challenging. To begin with we had no insight in this field, so we had to do some research and held several meetings with the master group and Øystein.

Using existing sources

The optimal solution for us would be to use the information in FEST on Pille.idi.ntnu.no, but we could not get it to work. We tried to use it remotely from Pille and to download it in RDF XML format, but we got errors during the download. We decided to use ICD 10[5] which is a international classification of diseases. We did not have time to use it for anything useful, but we could maybe see a use for it where we can link different active ingredients with the classification of diseases. Our solution does have an annotation property “harIndikasjon”(hasIndication) which connects our class of different active ingredients to the related disease classes. This is maybe not the best solution but there’s a lot to be done when we take this further to our master thesis.

Exportation of the query-results

After a couple of meetings with the students that might use this application we agreed on JSON as a format to represent the results from the queries. Did was already available on the SWI-prolog part of the Pille server, but it didn't really behave as we hoped it would. We therefore created an php script which pretty much extract the content from the website and prints it out as a JSON format.

Future work

Seeing as this is our intro to a master thesis, the work here is not finished even when the project in the course is due. Some of the functional requirements provided by the group already writing their thesis for Øystein is still not implemented. These functional requirements were all regarding the ontology and not the product of this particular project(the website). The next paragraphs discusses the requirements for the ontology yet to be implemented, as well as the use of automated source fetching and use of other sources.

One requirement for the master group was the possibility to query for either effects, indications and adverse effects for a specific patient group. Patient groups could be based on gender, age, pregnancy etc. This information can also be found in Legemiddelhåndboka. We decided early that we did not have time for this implementation. We talked with the master students and they agreed that this was not a high priority with the limited time we had. We have put a lot of thought into how this could be presented in the ontology, but nothing has been finalized yet. One possible solution is to create a class for "patient group" which would hold instances like "pregnant" and similar groups. Furthermore we think that the connection for drug and effects/indications/adverse effects can be solved by the class "Effekt" having object property of "patient group". This would not take long modeling-vise, but since we do not have an automated process for processing the content of Legemiddelhåndboka we will have to do it manually. As already stated this takes time, which were the main reason for not expanding the ontology at this point.

Initially we were set on fetching information from Legemiddelhåndboka by scraping the content of the website. We were given an example of how it could be done by our supervisor(Øystein), but realized early that this would be too time-consuming in order to fit our needs for the ontology. This will be our main priority after this project is finalized. The main challenge here is creating a structure of the information. There is no persistent structure in the content available on Legemiddelhåndboka, so there is hard for us to estimate the workload for this task.

Another aspect of this is the usage of the external sources (as mentioned earlier). We do believe that we did not get the time to fully use them to their full potential as we could have made some really interesting connections between drugs and actual diseases. We do hope that we can continue the work and make the ontology a bit more interesting with the proper use of the external sources we have available.

Conclusion

Throughout the project period we have understood the mindset behind an ontology and how to create them. Furthermore, we spent a lot of time understanding the medical terms and the field in general in order to proceed.

The biggest roadblock for us was understanding the medical field and the scope of the project. We benefited greatly from the meetings and exchanging information with our supervisor and the other master students. We also had some technical challenges such as getting an overview of our ontology in Protégé and understanding how to use SWI-Prolog on Pille. This ended up in being the biggest time consumers throughout the project period.

The project was a great opportunity for us to get to know the healthcare field as an small preliminary study for our master thesis. It is a lot of fun to get to know this complex field and getting to know it through making an ontology. It helped us to put a context to the content and gave us an idea of how to categorize and relate medical drugs to each other. It also was very much motivating to create something that could help others (both the other master students and

human beings that might have the need of a good system to understand the medical drugs they are using).

References

[1]http://en.wikipedia.org/wiki/Drug_interaction

[2]www.legemiddelhandboka.no

[3]http://www.legemiddelverket.no/Bruk_og_raad/FEST/hva_er_fest/Sider/default.aspx

[4]http://pille.idi.ntnu.no/browse/list_graph?graph=fest

[5]http://pille.idi.ntnu.no/browse/list_graph?graph=icd10no

[6]http://pille.idi.ntnu.no