Logboek masterproef

Naam student: Andreas De Witte

Titel: Client-side evaluatie van GeoSPARQL opvragingen over heterogene gegevensbronnen

onderzoeksgroep

Naam: IDLab Tel: +329 33 14920 Promotors:

- dr. ing. Pieter Colpaert
- dr. ir. Ruben Taelman Begeleiders:
- Brecht Van de Vyvere
- Julian Andres Rojas Melendez

Week 1 (17/02/2020)

Last week, I started with reading about Linked Data, RDF, SPARQL, semantic web and I had a first look at geosparql and the triple fragment pattern. After gaining some basic knowledge, I tried improving my SPARQL knowledge. At the last day, I started reading the documentation of Comunica and tried the tutorials "Comunica tutorial: Querying Data" and "Comunica tutorial: Creating a REDUCED actor". The first tutorial worked pretty well, which was really interesting, because now I could actually try out some SPARQL queries on a RDF data source which I could control for my queries. I also think that I might have found a bug in Comunica, but I'd like to have a look at it with someone else, because I just might be wrong too. The second tutorial was rather frustrating than a succes, since I kept getting an error during the "npm install" before the actual start of the exercise. This week, I'll try to get a better viewing of the total of Comunica and it's different modules. Also, I'd love to read the paper "Computing and querying strict, approximate, and metrically refined topological relations in linked geographic data." because I will need this for the further development of my thesis.

Week 2 (24/02/2020)

Last week, I started with trying the last tutorial about "Reduced Actor". Afterwards, I continued with reading the docs of comunica itself, to get a better understanding of it's modular architecture. I spent the rest of my time reading papers and documentation about geo locations and geosparql. This week, I will try to make a first implementation of geosparql.

Week 3 (02/03/2020)

Last week, I started writing for my state of the art of my thesis. However, I do not really enjoy writing about what I found on the internet, so I delayed this task a little bit. Instead, I started on them implementation of a geosparql actor in comunica. I did bump into a few problems tho. At first, I wasn't really sure about which class to extend for my actor, which gave me several problems? When I finally did find out which was the right one, I didn't know I needed to change this in the config files aswell. However, I did keep getting the same result, which was not what I was expecting. I used a dataset from the geosparql documentation's examples, so I had so geographic figures. However, comunica kept telling me that my datatype had a certain form, and the value was an empty string. Later on, I found out that the problem was something inside the tag, which prevented comunica from interpreting the result. I also tried changing the dataset (with an online converter) to jsonld, but

comunica couldn't handle this input eiter. After I edited the dataset to something easier to read, I bumped into the next problem: how am I going to calculate geospatial relations. to solve this solution, I changed the data to geojson and used the Terraformer library. This was actually harder than it should be, but I blame my inexperience with typescript, so this doesn't really count as a problem. Afterwards, I had trouble using the contains function. In the end, the problem was that contains can be called from any "Terraformer.Prime" object, but the argument in the contains method needed to have the type "Geometry", which is a special type from the library and which means only a subset of all Terraformer objects. Now, the current state of my implementation is the following: I can do a check for the contains, and it gives me a result. However, I do expect 3 objects inside the polygon that I used, but I get only 1 result. I think the problem is with the other results being on the edges of the requested one, but I did not test this yet. During this week, I'm planning on writing a bigger part of my state of the art section in my thesis, and I would like to finish/fix the contains function and implement another function (to be decided which) aswell.

Week 4 (09/03/2020)

Last week, I started from a partial working implementation of the geospatial function "contains". However, the library I used (Terraformer) contained some bugs on edge cases (polygon inside polygon, with common edge didn't "contain", line inside polygon from point to another point didn't "contain"). I changed to the library "turfjs", which is a modular geospatial library. This library has way less bugs and supports way more functions. I made an implementation for most of the topological function. I had some problems this week, with the difference between functions and predicates. if I want to use predicates, I most likely will have to transform the query before executing it. Furthermore, for combining multiple function, the sparqlee library needs extending. This should be possible by implementing the custom functions functionality. Last but not least, I did some more writing for my state of the art, but this is not sufficient yet to receive feedback on it. Next week, I'll start with implementing the calculations from one projection to another, before calculating the topological functions. After that, I'll be diving into the sparqlee code, but I'm not sure how long this will take.

Week 5 (16/03/2020)

Last week, I wanted to finish the calculations from one projection to another, so I can calculate the geospatial relations correct. I also wanted to finish the custom functions in sparqlee, but because of me feeling a little bit ill in combination with underestimating the work, I wasn't able to do either. I started with sparqlee, and started reading and learning how the code works. However, because of it's complexity, I decided to start with the implementation of the projections. This gave me far more trouble than it should've given me. The majority of the problems were because the typings weren't working too well, but this appeared to be a problem with my yarn installation. For this week, I'm going finish the calculations of the projections and I'm going to try to finish the custom functions in sparqlee. The rest of the week, I'll be spending writing for the thesis itself.

Week 6 (23/03/2020)

Last week, I planned on making the calculations to go from one geographic projection to another and create the "custom functions" functionality in sparqlee. However, I ended up making the calculations, but then shift my focus to writing for the thesis itself, in order to make it to the deadline of 31/03. This week, I'll continue writing for my thesis as much as I can. I want to finish my state of the art and make one extra chapter about the implementation of GeoSPARQL. After that, I'll try making an abstract and I'll be thinking about the title of my thesis.

Week 7 (30/03/2020)

last week, I have been writing for my thesis. I finished my state of the art (there are still some todos in tho) and I came up with a title. I also wrote a first version of the abstract, but I'm not sure yet if the result will be how I described it in the abstract. I will probably rewrite some parts of the abstract as I get further with the thesis. This week, I'll start with rereading my state of the are to remove typo's and afterwards I'll send it to all my supervisors. For the rest of the week, I'm going to try to extend sparqlee with custom functions.

Week 8 (06/04/2020)

Last week, I've been trying to implement the custom functions in sparqlee. However, it's much more complicated than I originally thought and I already spend a lot of time on trying to make this work. Since this isn't that relevant for my thesis (only a smart part is needed for me), int the end I decided not to waste more time on this and instead just add some GeoSPARQL functions too sparqlee. This has not been finished yet, so this is what I'll be doing for the first part of this week. The rest of the week, I'll start preparing for the presentation. Also, I feel good about presenting the progress in a 10 minute presentation. However, I was thinking maybe we could do this earlier than the 24th. We need to prepare a similar presentation for school (on which we will be graded), so this would be the perfect opportunity to get feedback on our presentation. The deadline for the video for school is the 19th.

Week 9 (14/04/2020)

Last week, I made the GeoSPARQL functions possible on sparqlee. This was much more complicated than I thought. I originally tried to add custom functions, but after too long time of trying, I decided to implement te functions in sparqlee itself for now. After this was finished, I started preparing the use case for the demo I'm planning to do. This week, I'll continue working on the demo and on the presentation for the upcomming deadline.

Week 10 (20/04/2020)

Last week, I have been putting my focus on the presentation. Right from the beginning, I had a clear view on how I wanted to present what I made so far, especially for the demo part of it. In code, the demo worked almost perfectly from the start. However, when I wanted to build it in a gui (the jQuery-widget), I bumped into a lot of problems. First of all, I had a name mismatch between my actor and his own name in package.json. Next, I had a version mismatch between the jQuery-widget and the init sparql actor. These 2 were the hardest problems, because I had no experience with this and I actually had no clue this could give problems. Thanks to Ruben Taelman's help, I managed to fix these problems. The next problem was a situation where the widget removed my symbolic link with the init actor on install, so it kept changing to the online version instead of mine. Another nasty problem was me needing to link both sparqlee and my own actor to the jQuery widget aswell, while I felt like only the new init sparql actor should be enough. The last problem was something code specific, a library I used (proj4) was giving me a very encrypted error, but in the end I found out that it was behaving differently in a node environment than in a browser environment. I feel like my productivity is lower when I'm at home, due to distractions. However, I know I can keep working way longer when I'm at home (I have no clue why this is the case), so I actually work the same amount of time, it just takes longer. This week, I will start writing on the thesis again. First I will put Ruben Taelman's feedback into the thesis and afterwards, I'm going to make a chapter about the implementation.

Week 11 (27/04/2020)

last week, I wanted to start writing again, about the implementations. However, I misjudged the size of some tasks I had for a group project of another course, so I decided to shift the thesis writing by a week. I ended up only preparing for the presentations. This week, I'll be writing for the thesis, as I said I would do last week.

Week 12 (04/05/2020)

Last week, I reworked some parts of my thesis, based on feedback. I also wrote my next chapter about the implementation itself. Something I didn't plan, was the implementation of non-topological relations, I did as well. I actually had a really productive week last week. This week, I'm planning on writing my introduction and setting up a triple pattern fragment interface, to test one of my hypotheses. This is not too much work, but I would like to finish a paper for another course this week.

Week 13 (11/05/2020)

last week, I put my focus on a paper for another course. I underestimated the amount of work it took, so I couldn't finish everything that I wanted to do for the thesis. In the end, I only did setup a TPF interface locally. This also took longer than I thought it would, because there was a bug. I fixed the bug and opened a pull request to the master though. This week, I'm planning on testing my hypothesis. I will postpone the writing of my introduction a bit more.

Week 14 (18/05/2020)

Last week, I truly felt the downside of working from home. I felt like I had the least productive week in a long time. This might be because of the prolonged time of working (which feels like months of exam preparation for me). In the end, all I did was setting up the tpf interface and test if this works (which it did) and setting up the sparql endpoint. The latter I didn't know how to access the endpoint itself, but I just got this working today. I did some brainstorming about both my introduction chapter and my chapter about the different interface, but I didn't really do any writing. This week, I'll be putting all my code online (to the forked repositories on my own github), so I can link the code in my writing. I'll also be writing the chapter about the interfaces. I'm not sure how long this will take, but if I can finish all of this this week, I'll start with the introduction as well.

Week 15 (25/05/2020)

Last week, I finished the evaluation and wrote it down. This week, I'll be writing my introduction, conclusion, my extended abstract. The plan is that I could push the code I created to the forked repo's next week, so I can edit the needed parts for the demo presentation before I push it.

Week 15 update (29/05/2020)

I've had a small problem in my planning, so I had to shift some priorities. In short, I have an exam that's 4 days earlier than I thought it was. However, I'm optimistic and I believe the presentations will be good

Week 16 (02/06/2020)

last week, I finishes my introduction, but I noticed my first (and only) exam was way earlier than I initially thought it was (being today). Because of that, I shifted my planning and focussed on the exam first. This week, I will write the conclusion and my extended abstract. I will also be rereading and changing the whole master thesis. as last, I wont to put all the code I made (with some changes for the demo) online and reference to those links in the correct chapters in the thesis.

Week 17 (08/06/2020)

Last week, I finished my master's thesis and I started rereading it to correct mistakes. This week, I'll be correcting some more mistakes and finish my todo-list. I'll finish my demo and presentation as well.