
Geological Knowledge Base Construction

Gayane Vardanyan
Yuhsuan Chen

The Idea

1

The Objective

Help internal stakeholders find relevant information faster and more efficiently.

2

The Methodology

Create a knowledge base that would help answer the questions.

3

Components

- The Knowledge Base
- Natural Language Processing

4

KPIs

- Reduction of search time
- High efficiency of answers

Five Main Stages of Work

- 1) Understand the domain related questions**
- 2) Data collecting & cleaning**
- 3) Create a knowledge base graph**
- 4) Parse questions into query to interact with knowledge graph**
- 5) Answer questions using the knowledge base**

Research Domain Questions

Question 1



What is stratigraphy ?

Stratigraphy is a geology study involved the study of the rock layer(strata). It includes three main subfields, lithostratigraphy, biostratigraphy and chronostratigraphy.

Lithostratigraphy: Studies the wells log, and physic characteristic of the rocks, including texture, mineral content and color.

Biostratigraphy and chronostratigraphy: Studies fossils to determine the absolute or relative age of the formation.

<https://en.wikipedia.org/wiki/Stratigraphy>

<https://www.glossary.oilfield.slb.com/en/terms/s/stratigraphy>

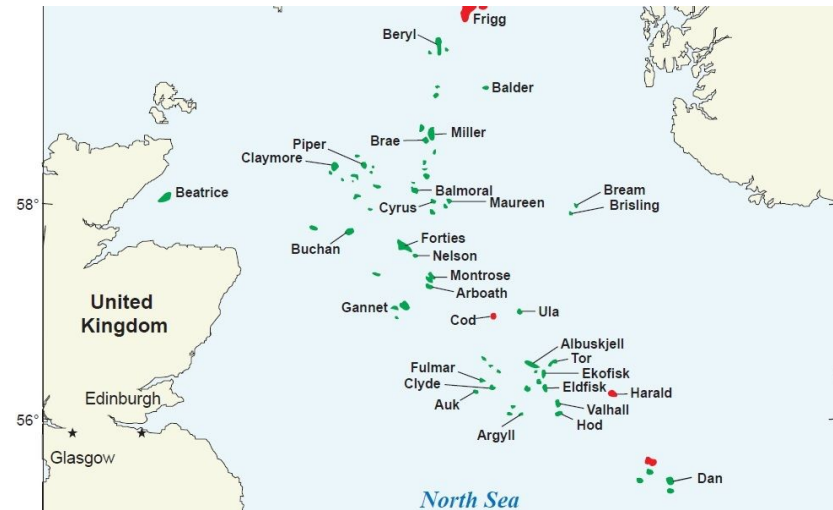
Question 2



Where is Ekofisk Formation ?

The Ekofisk Formation in the North Sea extends throughout the basinal areas of the Central Graben, Outer Moray Firth and South Viking Graben (Knox and Holloway 1992) and the Southern North Sea (Lott and Knox 1994).

Ekofisk Oil field, in block 2/4 of the Norwegian sector of the North Sea about 320 km southwest of Stavanger.



<http://nhm2.uio.no/norges/litho/ekofisk.php>

<https://certmapper.cr.usgs.gov/data/PubArchives/WEcont/regions/reg4/P4/tps/AU/au402513.pdf>

Question 3



What are the wells crossing Ekofisk Formation ?

- **well 1/3-1** from 3354 m to 3258 m, coordinates N 56°51'21.00", E 02°51'05.00".
No cores.
- **22/1-2A** from 2982.5 m to 2935 m, coordinates N 57°56'12.20", E 01°02'55.80".
No cores.
- **2/5-1** from 3132 m to 3041 m, coordinates N 56°38'19.95", E 03°21'07.94".
Cored through the upper 78 m.

Question 4



What is the group of the Ekofisk Formation ?

Chalk Group

Shetland Group

The group has now been expanded to include the formations of the former Chalk Group.

| Lithostrat. unit |
|--------------------------------------|
| BLODØKS FM |
| DELFIN FM (INFORMAL) |
| EKOFISK FM |
| HARDRÅDE FM |
| HIDRA FM |
| HOD FM |
| JORSALFARE FM |
| KVITNOS FM |
| KYRRE FM |
| NISE FM |
| SPRINGAR FM |
| SVARTE FM |
| TOR FM |
| TRYGGVASON FM |

CK - Chalk Group ▼

CKEK - Ekofisk Formation

CKGR - Ommelanden Formation

CKHM - Houthem Formation

CKMA - Maastricht Formation

CKGP - Gulpen Formation

CKVA - Vaals Formation

CKAK - Aachen Formation

CKOP - Oploo formation

CKTX - Texel Formation ►

<https://www.dinoloket.nl/en/stratigraphic-nomenclature/ekofisk-formation>

<https://factpages.npd.no/en/strat/pageview/litho/formations/33>

Question 5



What are the members of Ekofisk ?

LOWER MEMBER

The lowermost part (Ekofisk tight zone) consists of a low porosity to tight zone with a higher terrigenous clay content.

The larger part consists of the informal Ekofisk reworked zone with mainly reworked Maastrichtian chalks (Tor Formation) deposited as various mass flows and peridotite-facies chalks.

UPPER MEMBER

This zone is composed of mainly homogenous chalks with a low clay content, debris flows of reworked Danian chalks and minor turbidites.

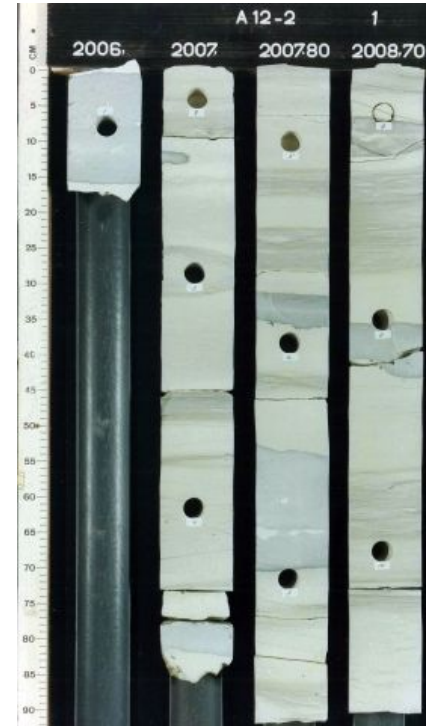
A lower tight to low porosity zone (Tommeliten tight zone) is present in parts of the Central Trough.

Question 6



What is the lithology of Ekofisk ?

White, chalky limestones contain rare white and grey nodular and bedded chert layers and thin, grey to green clay laminae. Some glauconite can occur in the basal interval.



https://en.wikipedia.org/wiki/Bouldnor_Formation

<http://nhm2.uio.no/norges/litho/ekofisk.php>

<https://factpages.npd.no/en/strat/pageview/litho/formations/33>

<https://www.dinoloket.nl/en/stratigraphic-nomenclature/ekofisk-formation>

Question 7



Describe at Best Ekofisk Formation?

Ekofisk formation is named after the Ekofisk Oil Field. The formation is widespread in the southern and central North Sea. Its thickness is up to 140 m and mainly form by white and chalky limestones. It's from the Danian age, the oldest age of the Paleocene period and it belongs to Shetland group.

Question 8



What is the top of the Ekofisk Formation for the well 1/3-1 ?

Formation tops in well 1/3-1

NPD fact sheet [1/3-1](#)

Click on unit name for all wells containing unit top.

According to Norwegian Petroleum Directorate, the depth is 3258.

| According to NPD | |
|------------------|-----------------------------|
| Depth | Unit |
| 97.00 | NORDLAND GP |
| 2995.00 | ROGALAND GP |
| 2995.00 | BALDER FM |
| 3006.00 | SELE FM |
| 3013.00 | LISTA FM |
| 3095.00 | VIDAR FM |
| 3147.00 | LISTA FM |
| 3209.00 | VÅLE FM |
| 3258.00 | SHETLAND GP |
| 3258.00 | EKOFISK FM |

| Norlex update (changes marked in green) | | | |
|---|------|-----------------------------|---------|
| Top | Base | Unit | Comment |
| 97.00 | | NORDLAND GP | |
| 2995.00 | | ROGALAND GP | |
| 2995.00 | | BALDER FM | |
| 3006.00 | | SELE FM | |
| 3013.00 | | LISTA FM | |
| 3095.00 | | VIDAR FM | |
| 3209.00 | | VÅLE FM | |
| 3258.00 | | SHETLAND GP | |
| 3258.00 | | EKOFISK FM | |
| 3354.00 | | TOR FM | |

Question 9



What is the period and age of Ekofisk ?

Danian

the Danian is the oldest age of the Paleocene period. The Danian age started from the Cretaceous–Paleogene extinction event 66 Ma. to 61.6 Ma, being followed by the Selandian age.

Data Collection & Cleaning

Data Scrape

- For questions: **What is _____?**
- There are 2 sources: Wikipedia and Schlumberger.

Stratigraphy

From Wikipedia, the free encyclopedia

This article is about the branch of geology. For stratigraphy in [archaeology](#), see [Stratigraphy \(archaeology\)](#).

Stratigraphy is a branch of [geology](#) concerned with the study of [rock](#) layers ([strata](#)) and layering (stratification). It is primarily used in the study of [sedimentary](#) and layered [volcanic rocks](#). Stratigraphy has two related subfields: [lithostratigraphy](#) (lithologic stratigraphy) and [biostratigraphy](#) (biologic stratigraphy).

stratigraphy

1. n. [Geology, Reservoir Characterization]

The study of the history, composition, relative ages and distribution of [strata](#), and the [interpretation](#) of strata to elucidate Earth history. The comparison, or [correlation](#), of separated strata can include study of their [lithology](#), [fossil](#) content, and relative or absolute age, or [lithostratigraphy](#), [biostratigraphy](#), and [chronostratigraphy](#).

Data Scrape

- For questions: **What is the top/bottom of _____ formation for the well _____ ?**
- Norwegian Licence for Open Government Data: <http://data.norge.no/nlod/en/1.0>

Wellbores penetrating

| Wellbore name | Wellbore completion date | Top depth [m] | Bottom depth [m] |
|--------------------------|--------------------------|---------------|------------------|
| 1/2-1 | 04.06.1989 | 3407 | 3514 |
| 1/2-2 | 02.02.2006 | 3418 | 3434 |
| 1/3-1 | 11.11.1968 | 3258 | 3354 |
| 1/3-2 | 27.07.1969 | 3207 | 3270 |
| 1/3-3 | 24.03.1983 | 3201 | 3288 |
| 1/3-4 | 08.05.1983 | 2754 | 2797 |
| 1/3-5 | 11.02.1985 | 3288 | 3384 |
| 1/3-6 | 22.06.1991 | 3103 | 3201 |
| 1/3-8 | 27.05.1997 | 3377 | 3478 |
| 1/3-9 S | 31.07.1998 | 3319 | 3409 |
| 1/3-11 | 30.08.2008 | 3502 | 3595 |
| 1/3-12 S | 22.07.2010 | 3473 | 3581 |
| 1/5-2 | 15.04.1974 | 3069 | 3152 |
| 1/5-4 S | 22.05.2002 | 2994 | 3013 |

| wlbName | IsuTopDepth | IsuBottomDepth | IsuName |
|----------|-------------|----------------|------------|
| 1/2-1 | 3407 | 3514 | EKOFISK FM |
| 1/2-2 | 3418 | 3434 | EKOFISK FM |
| 1/3-1 | 3258 | 3354 | EKOFISK FM |
| 1/3-11 | 3502 | 3595 | EKOFISK FM |
| 1/3-12 S | 3473 | 3581 | EKOFISK FM |
| 1/3-2 | 3207 | 3270 | EKOFISK FM |
| 1/3-3 | 3201 | 3288 | EKOFISK FM |
| 1/3-4 | 2754 | 2797 | EKOFISK FM |
| 1/3-5 | 3288 | 3384 | EKOFISK FM |
| 1/3-6 | 3103 | 3201 | EKOFISK FM |
| 1/3-8 | 3377 | 3478 | EKOFISK FM |
| 1/3-9 S | 3319 | 3409 | EKOFISK FM |
| 1/5-2 | 3069 | 3152 | EKOFISK FM |
| 1/5-4 S | 2994 | 3013 | EKOFISK FM |
| 1/6-1 | 3144 | 3247 | EKOFISK FM |

Data Mapping

Download csv file from: <https://factpages.npd.no/en/strat/tableview/overview>
Contains more than 150 lithography units (include group ,formation and member)

| | | | | | | |
|------------|---------------|-------------|--|---|----|-----|
| EKOFISK FM | FORMATI ON | SHETLAND GP | <p>Ekofisk Formation</p> <p>Name</p> <p>Named after the Ekofisk Field in Norwegian block 2/4 (Deegan & Scull 1977).</p> <p>Well type section</p> <p>Norwegian well 2/4-5 from 3164 m to 3037 m, coordinates N 56°34'29.77", E 03°12'13.03" (Fig 5.30). No cores.</p> <p>Well reference sections</p> <p>Norwegian well 1/3-1 from 3354 m to 3258 m, coordinates N 56°51'21.00", E 02°51'05.00" (Fig 5.24). No cores. UK well 22/1-2A from 2982.5 m to 2935 m, coordinates N 57°56'12.20", E 01°02'55.80" (Fig 5.25). No cores. Norwegian well 2/5-1 from 3132 m to 3041 m, coordinates N 56°38'19.95", E 03°21'07.94" (Fig 5.31). Cored through the upper 78 m.</p> <p>Thickness</p> <p>The formation is 127 m thick in the type-well, 96 m in 1/3-1, 47.5 m in 22/1-2 A and 91 m in 2/5-1. In the Norwegian sector, seismic interpretation indicates that a thickness of more than 150 m is found in the northwestern part of the Central Trough.</p> <p>Lithology</p> <p>In the type well, the formation comprises white, tan or beige, hard, dense, sometimes finely crystalline limestones, although softer chalky textures are also present. The formation usually consists of white to light grey, beige to brownish, mudstones or wackestones with occasional packstones/grainstones and pisolitic horizons, often alternating with argillaceous chalks, chalky limestones or limestones. Thin beds of grey, calcareous, often pyritic shales or clays are most common in the lower part while brownish-grey cherts occur rarely to abundantly throughout the formation.</p> | https://factpages.npd.no/factpages/default.aspx?nav1=strat&nav2=PageView Litho Formations&nav3=33 | 33 | 143 |
|------------|---------------|-------------|--|---|----|-----|

| Lithostratigraphy Unit | Name | Well type section | Well reference sections | Thickness | ...*(title) |
|------------------------|----------------|---------------------|-------------------------|-------------------------|---------------|
| Ekofisk Formation | Named after... | Norweign well 2/4-5 | Norweign well 1/3-1.... | The formation is 127... | ...*(content) |

velocity decreases towards the marly beds of the [Vale Formation](#), where the marly beds represent the changes towards chert (cf. [Mørkved et al. 2000, p. 50](#)).

Unify terminology

- Even though the data is from same website, the terminology still various from different lithology units (plurals, case, typo, order)
- Based on similarity of the words. Different Clustering function.

This feature helps you find groups of different cell values that might be alternative representations of the same thing. For example, the two strings "New York" and "new york" are very likely to refer to the same concept and just have capitalization differences, and "Gödel" and "Godel" probably refer to the same person. [Find out more...](#)

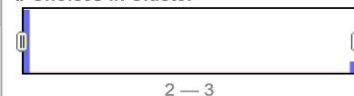
Method

Keying Function

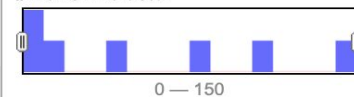
7 clusters found

| Cluster Size | Row Count | Values in Cluster | Merge? | New Cell Value |
|--------------|-----------|--|--------------------------|------------------------------|
| 3 | 146 | <ul style="list-style-type: none"> Depositional environment (130 rows) Depositional environments (15 rows) Depositional enviroinment (1 rows) | <input type="checkbox"/> | Depositional environment |
| 2 | 10 | <ul style="list-style-type: none"> Reference sections (9 rows) Reference section (1 rows) | <input type="checkbox"/> | Reference sections |
| 2 | 114 | <ul style="list-style-type: none"> Well reference section (80 rows) Well reference sections (34 rows) | <input type="checkbox"/> | Well reference section |
| 2 | 78 | <ul style="list-style-type: none"> Basal Stratotype (77 rows) Basal stratotype 1) (1 rows) | <input type="checkbox"/> | Basal Stratotype |
| 2 | 2 | <ul style="list-style-type: none"> UPPER MEMBER OF THE EKOFISK FORMATION (1 rows) Upper member of the Tor Formation (1 rows) | <input type="checkbox"/> | Upper Member |
| 2 | 39 | <ul style="list-style-type: none"> Characteristics of the upper boundary (38 rows) Characteristics of the tipper boundary (1 rows) | <input type="checkbox"/> | Characteristics of the upper |
| 2 | 2 | <ul style="list-style-type: none"> LOWER MEMBER OF THE EKOFISK | <input type="checkbox"/> | Lower Member |

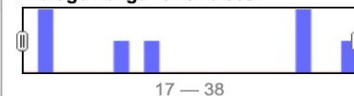
Choices in Cluster



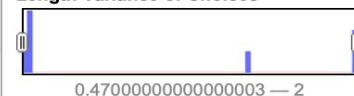
Rows in Cluster



Average Length of Choices



Length Variance of Choices



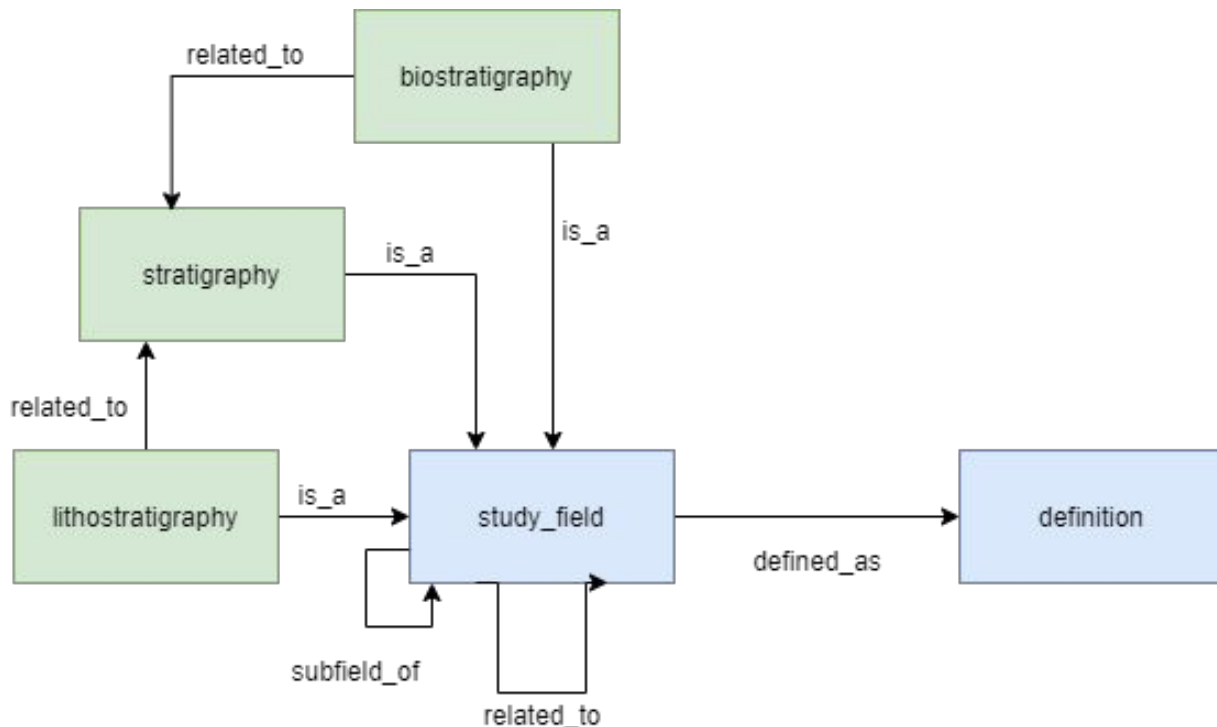
Final Data

| | ID | title | content |
|--|----|-------------------|--------------------------|
| | 0 | Adventdalen Group | Name |
| | 1 | Adventdalen Group | Well type section |
| | 2 | Adventdalen Group | Thickness |
| | 3 | Adventdalen Group | Lithology |
| | 4 | Adventdalen Group | Distribution |
| | 5 | Adventdalen Group | Age |
| | 6 | Adventdalen Group | Depositional environment |
| | 7 | Adventdalen Group | Subdivision |
| | 8 | Adventdalen Group | Compiled from |
| | 0 | Agat Formation | Name |
| | 1 | Agat Formation | Well type section |
| | 2 | Agat Formation | Well reference section |
| | 3 | Agat Formation | Thickness |
| | 4 | Agat Formation | Lithology |
| | 5 | Agat Formation | Lower member |
| | 6 | Agat Formation | Upper member |
| | 7 | Agat Formation | Distribution |
| | 8 | Agat Formation | Age |
| | 9 | Agat Formation | Depositional environment |
| | 10 | Agat Formation | Source |
| | 0 | Akkar Member | Name |
| | 1 | Akkar Member | Well type section |
| | 2 | Akkar Member | Well reference section |
| | 3 | Akkar Member | Thickness |
| | 4 | Akkar Member | Lithology |
| | 5 | Akkar Member | Lower member |
| | 6 | Akkar Member | Age |
| | 7 | Akkar Member | Depositional environment |
| | 8 | Akkar Member | Compiled from |
| | 0 | Alge Member | Name |
| | 1 | Alge Member | Well type section |

Knowledge base Modeling & Creation

Questions related to stratigraphy

- Blue rectangles indicate concepts.
- Green rectangles indicate instances.
- We have a triple format representation.
- Can be later expanded over more study areas.

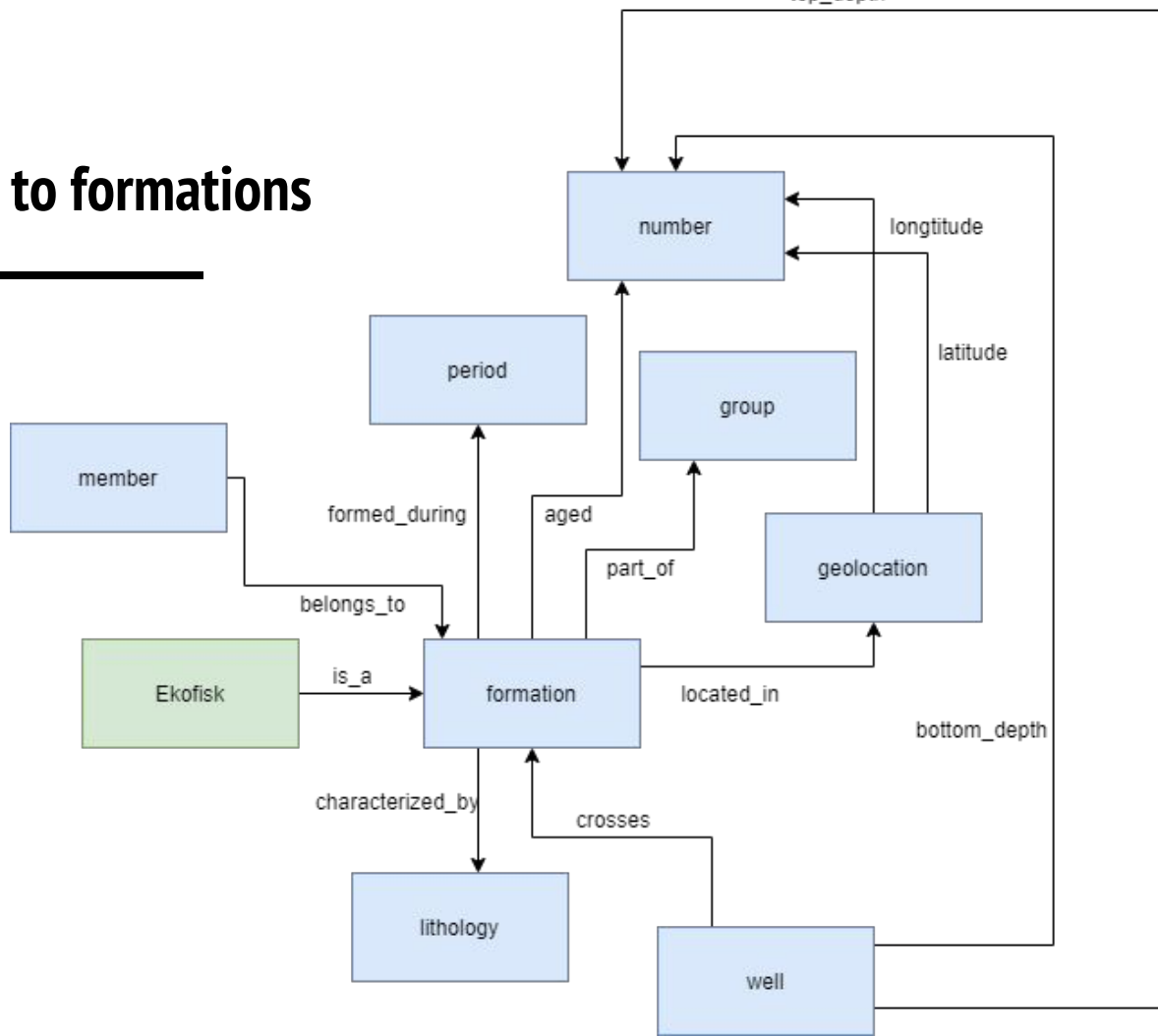


Stratigraphy



Questions related to formations

- Blue rectangles indicate concepts.
- Green rectangles indicate instances.
- We have a triple format representation.
- Can be later expanded over more study areas.
- The entities and relations are named

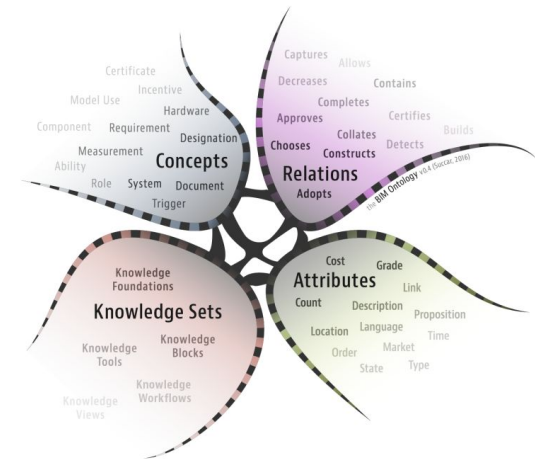




Ontology Editors

There's a large number of ontology editors:

- **Protégé** - Popular and pluggable ontology editor
- **NeON Toolkit** - A number of plugins are available. More suitable for huge projects
- **SWOOP** - Small and simple ontology editor
- **Neologism** - Online vocabulary editor and publishing platform.

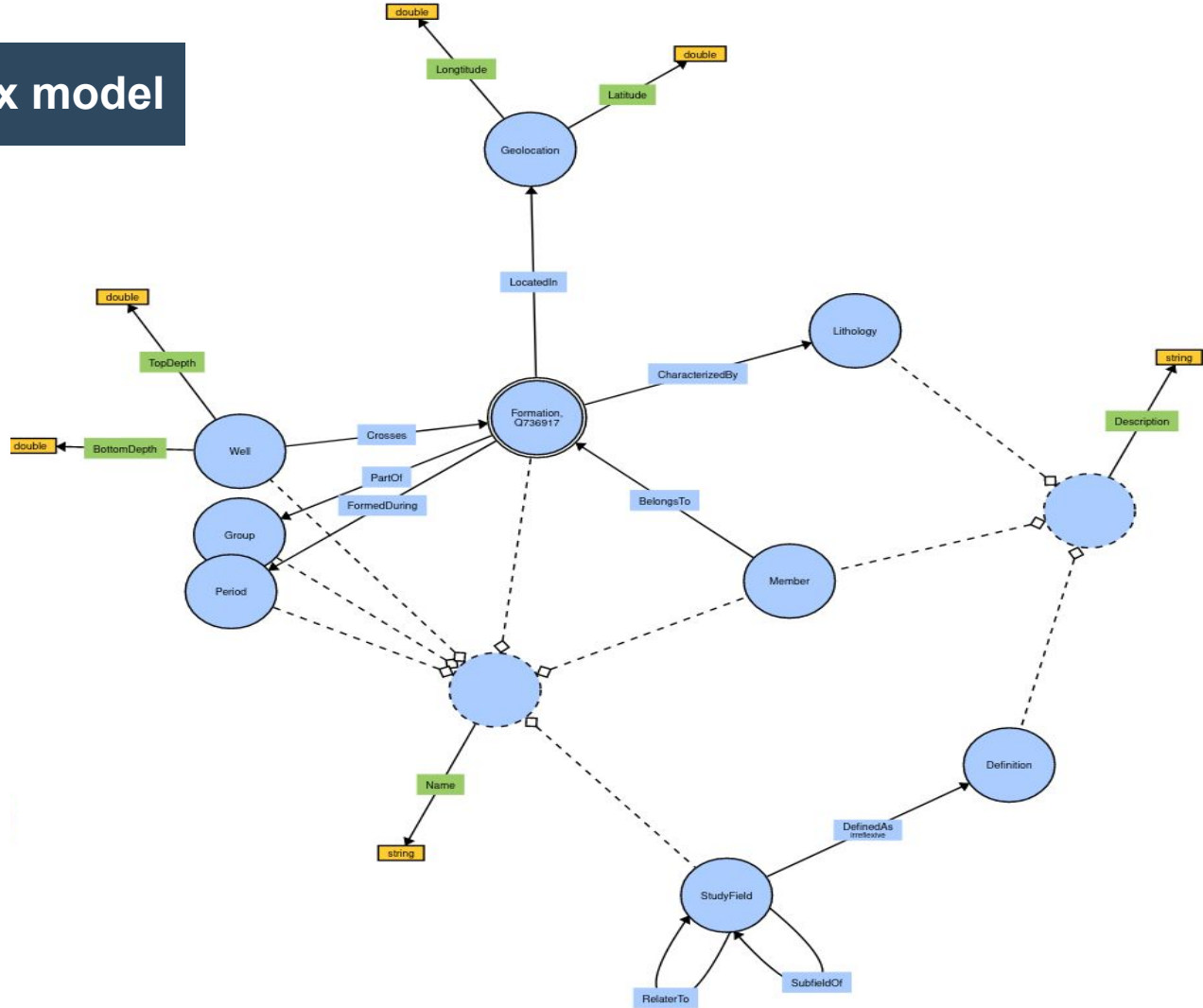




Many advantages that voted for it:

- Runs on a broad range of hardware platforms
- Has an extremely active user community
- Has a GUI and API
- Contains a graphical editor for Logical OWL Expressions
- Has a direct access to reasoners
- Provides multi-user support
- Supports multiple storage formats

TBox model



Using Java & Jena Library

- **Construct TBox & populate ABox**
 - **Jena allow ABox to have semantic web features and follow the constraint of the TBox model**
 - **Process csv files and turn them into rdf files**
- **Parse User's question**
- **Connect to graph database and build Query -> Don't need to use GraphDB**

Define TBox



Example

```
final String ns = "http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#";  
OntClass Formation = ontModel.createClass(ns + "Formation");  
OntClass Group = ontModel.createClass(ns + "Group");  
ObjectProperty PartOf = ontModel.createObjectProperty(ns + "PartOf");  
PartOf.addDomain(Formation);  
PartOf.addRange(Group);
```

Define ABox Example



Define ABox

Iterate every row in the csv file

```
String formation_name = line[3];
String group_content = line[2];
// Inserting the formation
Individual formation = formationClass.createIndividual(ns + "Formation/" + Utils.cleanURI(formation_name));
Literal formation_Name_string = ontModel.createTypedLiteral(formation_name, XSSDatatype.XSDstring);
ontModel.add(formation, Name_prop, formation_Name_string);
// Inserting the group
Individual group = groupClass.createIndividual(ns + "Group/" + Utils.cleanURI(formation_name));
Literal group_content_string = ontModel.createTypedLiteral(group_content, XSSDatatype.XSDstring);
ontModel.add(group, Name_prop, group_content_string);
ontModel.add(formation, PartOf_prop, group);
.....
OutputStream out = new FileOutputStream("output-test.rdf");
RDFDataMgr.write(out, ontModel, Lang.RDFXML);
```

Output rdf

```
205 <j.0:Formation rdf:about="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation
206 <j.0:PartOf>
207 <j.0:Group rdf:about="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Group/tor
208 <j.0:Name>sotbakken group</j.0:Name>
209 </j.0:Group>
210 </j.0:PartOf>
211 <j.0:FormedDuring>
212 <j.0:Period rdf:about="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Period/t
213 <j.0:Name>Late Paleocene to Oligocene. </j.0:Name>
214 </j.0:Period>
215 </j.0:FormedDuring>
216 <j.0:CharacterizedBy>
```

```
281989 <j.0:Well rdf:about="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Well/vestland_group
281990 <j.0:BottomDepth rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
281991 >3667</j.0:BottomDepth>
281992 <j.0:TopDepth rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
281993 >3627</j.0:TopDepth>
281994 <j.0:Crosses rdf:resource="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/v
281995 <j.0:Name>7/12-10</j.0:Name>
281996 </j.0:Well>
281997 <j.0:Well rdf:about="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Well/lista_formatio
281998 <j.0:BottomDepth rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
281999 >2244</j.0:BottomDepth>
282000 <j.0:TopDepth rdf:datatype="http://www.w3.org/2001/XMLSchema#double"
282001 >2107</j.0:TopDepth>
282002 <j.0:Crosses rdf:resource="http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/l
282003 <j.0:Name>16/7-3</j.0:Name>
```

Validation using Sparql Query

What is stratigraphy?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?def where

{

?s stratig:DefinedAs ?definition .

?definition stratig:Description ?def .

}

What wells cross Ekofisk formation?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?well where

{

?well stratig:Crosses ?formation .

?formation stratig:Name

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/ekofisk_formation> .

}

What is the group of ekofisk formation?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?name where

{

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/ekofisk_formation> stratig:PartOf ?group .

?group stratig:Name ?name .

}

What is the top of the Ekofisk Formation for the well 1/3-1 ?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?top where

{

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Well/ekofisk_formation/1/3-1> stratig:TopDepth ?top .

}

What is the lithology of Ekofisk ?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?name where

{

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/ekofisk_formation> stratig:CharacterizedBy ?lithology .

?lithology stratig:Name ?name .

}

What is the period and age of Ekofisk ?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?name where

{

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/ekofisk_formation> stratig:FormedDuring ?period .

?period stratig:Name ?name

}

What are the members of Ekofisk formation ?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?member where

{

?member stratig:BelongsTo

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/ekofisk_formation> .

}

// no members for Ekofisk formation?

What is the top of the Ekofisk Formation for the well 1/3-1 ?

PREFIX stratig:

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

select ?top where

{

<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Well/ekofisk_formation/1/3-1> stratig:TopDepth ?top .

}

Weekly Progress

1. Geo location extraction
2. Parsing user question into query

To answer where is ekofisk formation, what needs to be done

| | |
|----------|---|
| 1 | Collect well geolocation data |
| 2 | Process the data |
| 3 | Convert latitude and longitude into a bounding box |
| 4 | Update the TBox and the ABox |
| 5 | Test |

Data Collection

```
def fetch_npi_id(np_id):
    with open(f"/home/Projects/gayane/bdrp/well_info/html_dump/{np_id}.html"
              ) as f:
        soup = BeautifulSoup(f.read(), 'html.parser')
        table = soup.find(id="8iT0S0T0")
        rows = table.find_all("tr")
        detail = {}
        for row in rows:
            cols = row.find_all("td")
            col_name = cols[0].text
            value = cols[1].text
            detail[col_name] = value
        results.append(detail)
```

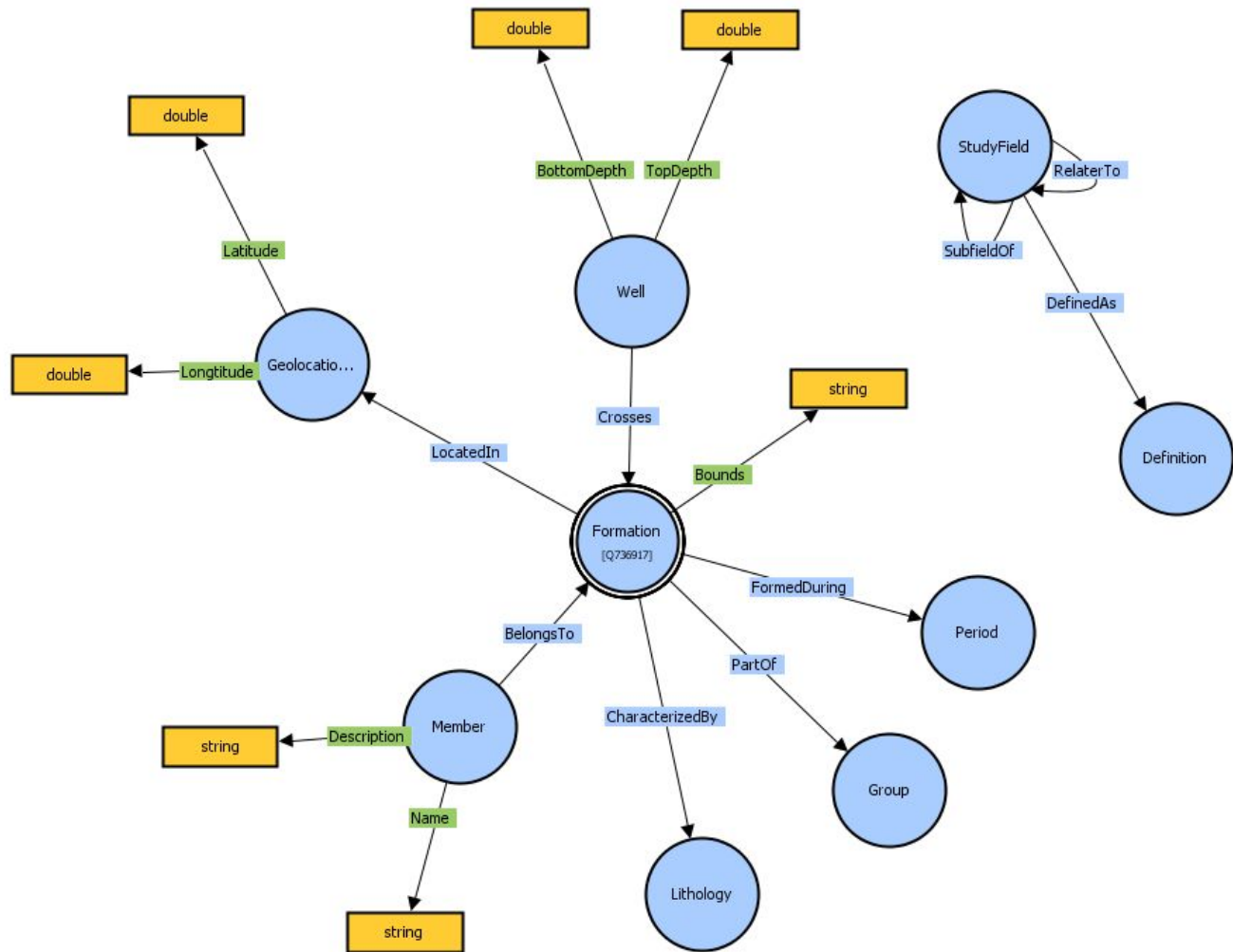
```
def get_nids():
    np_ids = []
    master_page = requests.get("https://factpages.npd.no/en/wellbore/
                                pageview/exploration/all")
    master_list_page = BeautifulSoup(master_page.content, 'html.parser')
    tables = master_list_page.find(id="tvCarriers").find_all("table")
    for table in tqdm(tables):
        atag = table.find("a")
        np_ids.append(atag["href"].split("/")[1])
    return np_ids
```

Data Collection

Loop parallelization to fasten up the scraping:

```
def download_to_local(ids):  
    with parallel_backend('threading'):  
        Parallel(n_jobs=50)(delayed(download_page)(npi_id) for npi_id in tqdm  
                             (ids))  
    files_present = [path.split(".")[0] for path in  
                    os.listdir("/home/Projects/gayane/bdrp/well_info/html_dump"  
                               )]  
    for np_id in ids:  
        if np_id not in files_present:  
            download_page(np_id)
```

Updating the TBox



Updating the ABox

- Modify the data injection process to:
 - Process the raw data obtained
 - Turn it into a bounding box.
- The data contained the original latitude and longitude values as strings:
 - 56°59' 32" N;
 - 2°29' 47.66" E.
- Geographically:
 - Latitude increases as you go north
 - Longitude increases as you go East
- The idea is:
 - Treat latitude as a north-south axis
 - Treat longitude as an east-west axis.
 - Compare hour, minute and second parts of the latitude/longitude incrementally
 - Return the bounding box formed by the north-, south-, east-, west- most points

Querying with Sparql

PREFIX stratig: <http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>

SELECT ?boundary

WHERE

```
{  
    ?formation stratig:Name "EkofiskFormation";  
    stratig:Bounds ?boundary  
}
```

| | boundary | ⌵ |
|---|--|---|
| 1 | "Formation is bounded in NS by60° 47' 38.94" N to 56° 7' 32.15" N.Formation is bounded in EW by6° 10' 4.7" E to 1° 32' 49.9" E." | |

Using Java & Jena Library

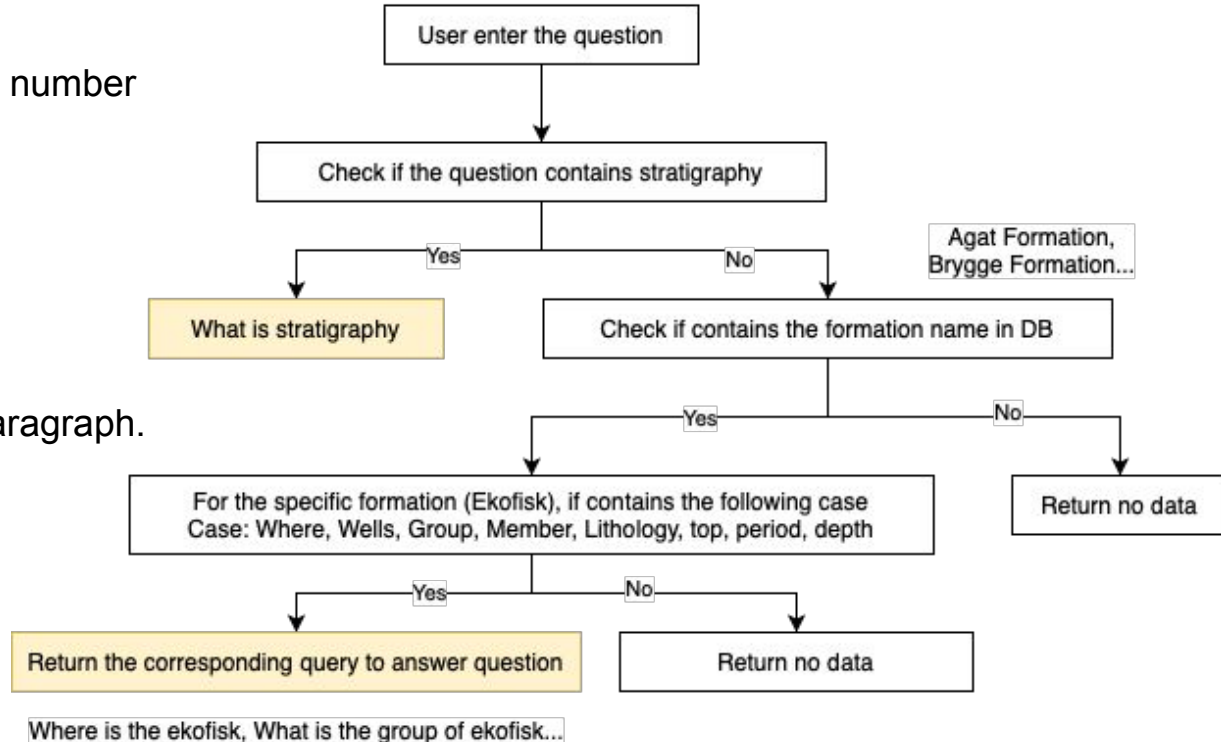
- **Construct TBox & populate ABox**
 - **Jena allow ABox to have semantic web features and follow the constraint of the TBox model**
 - **Process csv files and turn them into rdf files**
- **Parse User's question**
- **Connect to graph database and build Query -> Don't need to use GraphDB**

Methods to parse text into sparql query.

Functions:

Check the formation name and well's number

- Main(): Execute Query
- Function: query_selection()
 - Check keywords
 - Build Query
- Parse output into answers in paragraph.



Question 1



What is stratigraphy ?

No need query to answer, as it's independent from the rest of classes

Stratigraphy is a geology study involved the study of the rock layer(strata). It includes three main subfields, lithostratigraphy, biostratigraphy and chronostratigraphy.

Question 2



What wells cross _____ formation ?

```
else if(input_question.contains("well")){  
  
    String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/" + subject + ">";  
    QueryString ="PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +  
        "SELECT ?wellnum "+  
        "WHERE { ?well stratig:Crosses "+formation_uri+" . "+  
        "?well stratig:Name ?wellnum . }";  
  
}
```

| ----- | |
|-------------|--|
| wellnum | |
| ===== | |
| "2/8-2" | |
| "16/1-23 S" | |
| "16/2-11 A" | |
| "15/9-17" | |
| "2/4-23 S" | |

Question 3



What is the group of _____ formation ?

```
else if(input_question.contains("group")){
```

```
String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/" + subject + ">";
```

```
QueryString ="PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +
```

```
"SELECT ?name "+
```

```
"WHERE { "+"formation_uri+" stratig:PartOf ?group .\n" +
```

```
"?group stratig:Name ?name . }";
```

```
}
```

```
-----  
| name |  
=====
```

| |
|------------------|
| "shetland group" |
|------------------|

```
-----
```

Question 4



What is the lithology of _____ formation ?

```
else if(input_question.contains("lithology")){
```

```
String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Lithology/" + subject + ">";
```

```
QueryString ="PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +
```

```
"SELECT ?name "+
```

```
"WHERE { "+formation_uri+" stratig:Name ?name . }";
```

```
}
```

```
what is the lithology of Garn formation
```

```
-----  
| name |
```

```
=====
```

| |
|---|
| "The Garn Formation consists of medium to coarse-grained, moderately to well-sorted sandstones. Mica-rich zones are present. The sandstone is occasionally carbonate-cemented." |
|---|

```
-----
```

Question 5



What is the age/period of _____ formation ?

```
if(input_question.contains("age")||input_question.contains("period")){

String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/" +subject+">";
QueryString ="PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +
"SELECT ?name "+
"WHERE { "+formation_uri+" stratig:FormedDuring ?period . " +
"?period stratig:Name ?name . }";

}
```

What is the age of Lyr Formation?

```
-----
| name |
=====
| "Valanginian to Early Aptian. " |
-----
```

Question 6



What is the member of _____ formation ?

```
else if(input_question.contains("member")) {  
    //example: fruholmen (has member)  
    String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/" + subject + ">";  
    QueryString = "PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +  
        "SELECT ?name " +  
        "WHERE { ?member stratig:BelongsTo "+formation_uri+" ."+  
        "?member stratig:Name ?name . }";  
}
```

what are the members of fruholmen formation?

| name |
|-----------------|
| "Akkar Member" |
| "Krabbe Member" |

Question 7



Where is _____ formation ?

```
else if(input_question.contains("where")||input_question.contains("location")) {  
  
    String formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Formation/" + subject + ">";  
    QueryString = "PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +  
        "SELECT ?boundary " +  
        "WHERE { "+formation_uri+" stratig:Bounds ?boundary . }";  
}
```

Where is ekofisk formation?

| boundary
=====

| "Formation is bounded in NS by 60° 47' 38.94'' N to 56° 7' 32.15'' N. Formation is bounded in EW by 6° 10' 4.7'' E to 1° 32' 49.9'' E." |

Question 8



What is the top of _____ formation for _____ well ?

```
if(!well_num.equals("")){  
    String  
    formation_uri="<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#Well/" + subject + "/" + well_num + ">";  
    System.out.println(formation_uri);  
    QueryString = "PREFIX stratig:<http://www.semanticweb.org/user/ontologies/2020/11/Stratigraphy_in_North_Sea#>" +  
        "SELECT ?topDepth "+  
        "WHERE { "+formation_uri+" stratig:TopDepth ?topDepth . }";  
}
```

what is the top of ekofisk formation for 1/3-1 well

```
-----  
| topDepth |  
=====
```

| |
|---|
| "3258"^^<http://www.w3.org/2001/XMLSchema#double> |
|---|

```
-----
```

Demo

What's next?

Example: What are the member of fruholmen formation?

After execute query, this is what we get:

| | |
|-----------------|--|
| ----- | |
| name | |
| ===== | |
| "Akkar Member" | |
| "Krabbe Member" | |
| ----- | |

But we want something more like human: The member of fruholmen formation includes Akkar Member and Krabbe Member.

Also add the best description for the formation.