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

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.
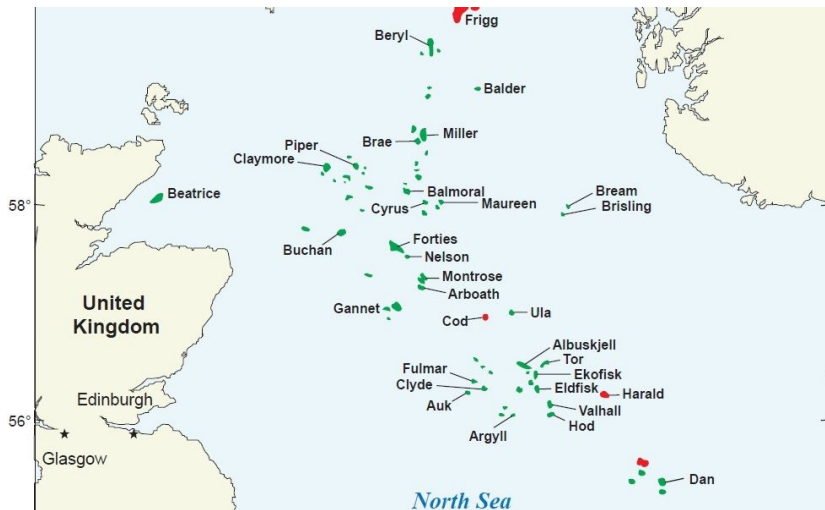
https://en.wikipedia.org/wiki/Stratigraphy
https://www.glossary.oilfield.slb.com/en/terms/s/stratigraphy

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

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.

What is the group of the Ekofisk Formation ?

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 ▶

Chalk Group

Shetland Group

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

Answers are different from
One website to another

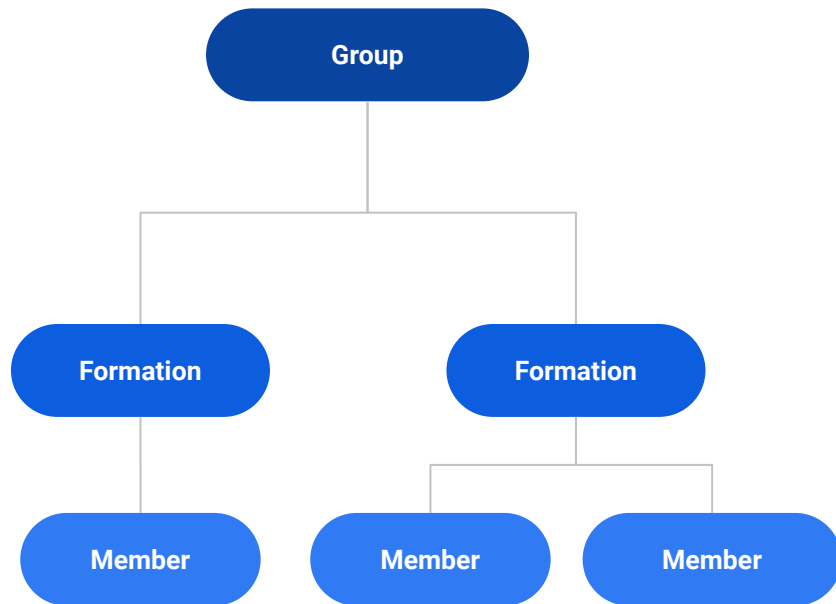| Lithostrat. unit |
| --- |
| BLODØKS FM |
| DELFIN FM (INFORMAL) |
| EKOFISK FM |
| HARDRÅDE FM |
| HIDRA FM |
| HOD FM |
| JORSALFARE FM |
| KVITNOS FM |
| KYRRE FM |
| NISE FM |
| SPRINGAR FM |
| SVARTE FM |
| TOR FM |
| TRYGGVASON FM |

What are the members of Ekofisk ?

LOWER MEMBER

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

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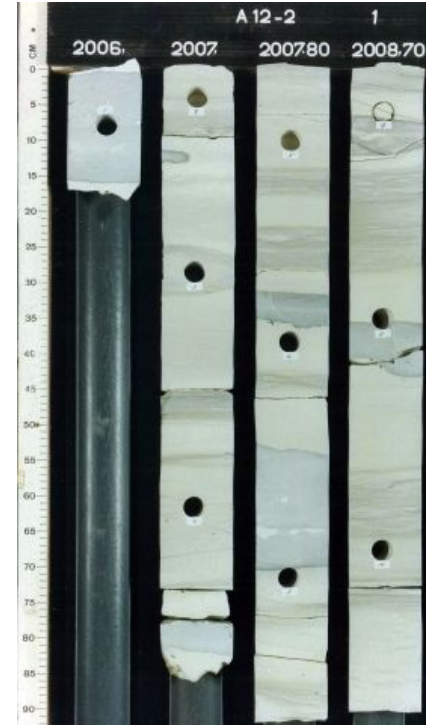

http://nhm2.uio.no/norges/litho/ekofisk.php
https://factpages.npd.no/en/strat/pageview/litho/formations/33

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

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

According to Norwegian Petroleum Directorate, the depth is 3258.

One formation has many wells

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 NPD		Norlex update (changes marked in green)			
Depth	Unit	Top	Base	Unit	Comment
97.00	NORDLAND GP	97.00		NORDLAND GP	
2995.00	ROGALAND GP	2995.00		ROGALAND GP	
2995.00	BALDER FM	2995.00		BALDER FM	
3006.00	SELE FM	3006.00		SELE FM	
3013.00	LISTA FM	3013.00		LISTA FM	
3095.00	VIDAR FM	3095.00		VIDAR FM	
3147.00	LISTA FM	3209.00		VÅLE FM	
3209.00	VÅLE FM	3258.00		SHETLAND GP	
3258.00	SHETLAND GP	3258.00		EKOFISK FM	
3258.00	EKOFISK FM	3354.00		TOR FM	

http://nhm2.uio.no/norges/litho/well_tops.php?well_name=1/3-1

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.

Question 9

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.

Properly compose the previous answers

Data Collection & Cleaning

- 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	22.05.2003	2994	3013

wlbName	lsuTopDepth	lsuBottomDepth	lsuName
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	FORMATION	SHETLAND GP	Ekofisk Formation	https://factpages.npd.no/factpages/default.aspx?nav1=strat&nav2=PageView|Litho|Formations&nav3=33	33	143
			Name			
			Named after the Ekofisk Field in Norwegian block 2/4 (Deegan & Scull 1977).			
			Well type section			
			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.			
			Well reference sections			
			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.			
			Thickness			
			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.			
			Lithology			
			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.			

Lithostratigraphy Unit	Name	Well type section	Well reference sections	Thickness	...*(title)
Ekofisk Formation	Named after…	Norweign well 2/4-5	Norweign well ⅓-1....	The formation is 127...	...*(content)

Unify terminology

- Depositional environment (130 rows)
- Depositional environments (15 rows)
- Depositional envitronment (1 rows)

- Reference sections (9 rows)
- Reference section (1 rows)

- Well reference section (80 rows)
- Well reference sections (34 rows)

- Basal Stratotype (77 rows)
- Basal stratotype 1) (1 rows)

- UPPER MEMBER OF THE EKOFISK FORMATION (1 rows)
- Upper member of the Tor Formation (1 rows)

- Characteristics of the upper boundary (38 rows)
- Characteristics of the tipper boundary (1 rows)

- 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.

Final Data

	ID	title	content
0	Adventdalen Group	Name	The group is named after a major valley in central Spitsbergen.
1	Adventdalen Group	Well type section	The type area is Central Spitsbergen.
2	Adventdalen Group	Thickness	Thickness on Svalbard: ca. 750-1600 m, known thickness on the Barents Sea Shelf: ca. 1000-1750 m.
3	Adventdalen Group	Lithology	The Adventdalen Group comprises shales, siltstones and sandstones of Late Jurassic to Early Cretaceous age in Svalbard and throughout the Barents Sea Shelf.
4	Adventdalen Group	Distribution	The group is widely exposed along the margins of the Central Tertiary Basin on Spitsbergen, as well as in eastern Spitsbergen (Sabine Land) and on Kong Karls Land. It con
5	Adventdalen Group	Age	Late Jurassic - Early Cretaceous.
6	Adventdalen Group	Depositional environment	The group is dominated by dark marine mudstones, but includes also deltaic and shelf sandstones as well as thin, condensed carbonate beds. Important hydrocarbon sour
7	Adventdalen Group	Subdivision	Six formations are defined within the group on the Barents Sea Shelf: the Fuglen, Hekkingen, Klippfisk, Knurr, Kolje and Kolmule formations.
8	Adventdalen Group	Compiled from	Dallmann, W. K. (ed.) 1999: Lithostratigraphic lexicon of Svalbard. Review and recommendations for nomenclature use. Upper Palaeozoic to Quaternary Bedrock. Norwegia
0	Agat Formation	Name	Named after the gas-condensate Agat Discovery in Norwegian block 35/3.
1	Agat Formation	Well type section	Norwegian well 35/3-4 from 3589 m to 3345 m, coordinates N 61°51'54.54", E 03°52'26.99" (Fig 5.19), 95 m of cores, mainly from the lower half of the formation.
2	Agat Formation	Well reference section	Norwegian well 35/3-5 from 3620 m to 3219 m, coordinates N 61°47'46.71",E 03°54'44.01" (Fig 5.20). 65 m of cores from the upper part of the formation.
3	Agat Formation	Thickness	In the type well the gross thickness of the formation is 244 m, and in the reference well 401 m. The gross thickness varies in that range in the wells in block 35/3.
4	Agat Formation	Lithology	In the type well the formation consists of white to light grey, fine- to medium-grained, moderately to well-sorted sandstones alternating with grey claystones. The sandstone
5	Agat Formation	Lower member	The lower boundary is defined where sandstones become the dominant lithology and is placed at the base of the first marked coarsening-upwards sandstone unit or distinc
6	Agat Formation	Upper member	The upper boundary is placed at the top of the upper sandstone layer. This boundary is especially distinct on the gamma-ray log since the overlying sediments are dominate
7	Agat Formation	Distribution	The formation is encountered in the area around the Måløy Fault Blocks in Norwegian blocks 35/3-36/1 (Fig 5.21) and is expected to be present along the western boundary
8	Agat Formation	Age	Aptian-Albian (possibly Early Cenomanian).
9	Agat Formation	Depositional environment	Marine environment influenced by gravity flows of sediment.
10	Agat Formation	Source	Isaksen, D. and Tonstad, K. (eds.) 1989: A revised Cretaceous and Tertiary lithostratigraphic nomenclature for the Norwegian North Sea. NPD-Bulletin No. 5, 59 pp.
0	Akkar Member	Name	Norwegian for "squid".
1	Akkar Member	Well type section	Norwegian well 7121/5-1 coordinates 71° 35'54.88"N, 21° 24'21.78"E (Fig 4.46).
2	Akkar Member	Well reference section	Norwegian well 7120/12-1 coordinates 71°6'48.71"N, 20° 45'20.13"E (Fig 4.47).
3	Akkar Member	Thickness	The gross thickness of the member is 55 m in the type well, and 38 m in the reference well.
4	Akkar Member	Lithology	Grey to dark grey shales, interbedded sandstones, coal.
5	Akkar Member	Lower member	The base of the Akkar Member (and the Fruholmen Formation) is defined by a marked increase in gamma ray and neutron porosity logs, but often more moderate increases
6	Akkar Member	Age	Norian (based on palynology).
7	Akkar Member	Depositional environment	Open marine.
8	Akkar Member	Compiled from	Dalland, A., Worsley, D. and Ofstad, K. (eds.) 1988: A lithostratigraphic scheme for the Mesozoic and Cenozoic succession offshore mid- and northern Norway. NPD-Bulleti
0	Alge Member	Name	Norwegian for "alga".
1	Alge Member	Well type section	Norwegian well 7119/12-1 coordinates N 71°6'08.00'' E 19°47'40.29''.

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.
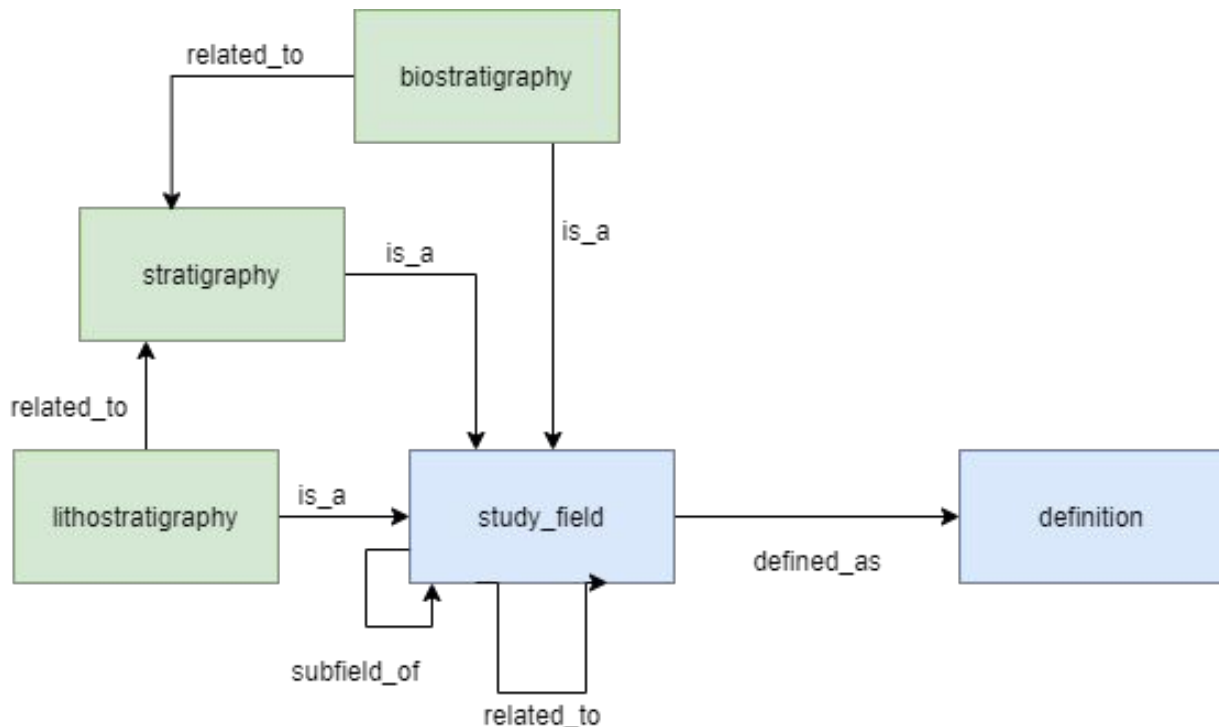
Questions related to formations

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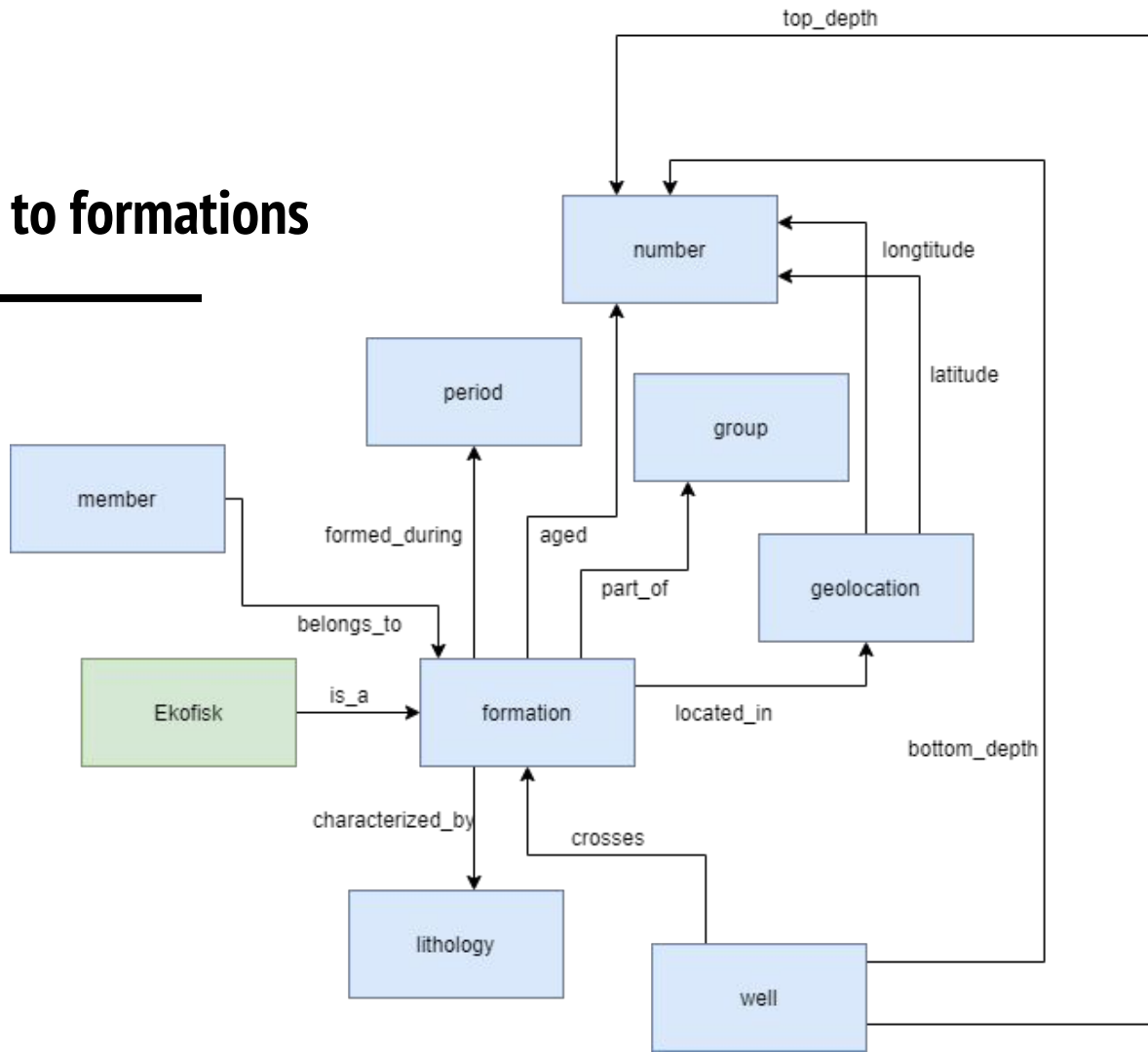- Can be later expanded over more study areas.
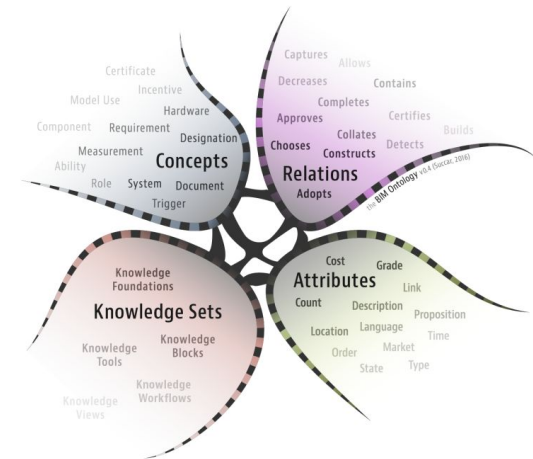- The entities and relations are named

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.

Protégé

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
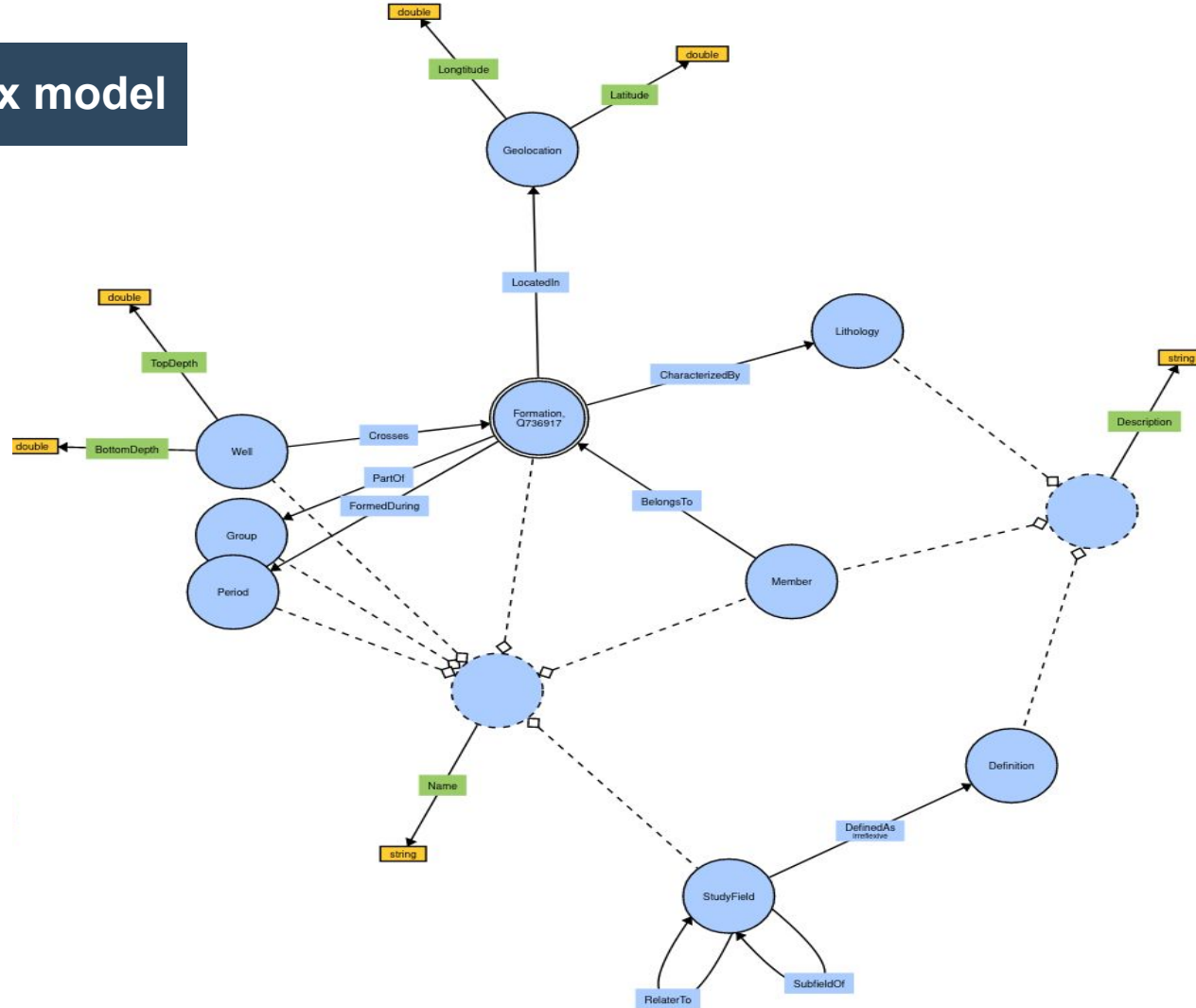
Using Java & Jena Library

- **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**

Define TBox

Define namespace → **Create the Class** → **Create the Property** → **Define domain & range**

Create the Class	Create the Property
Formation	Object property:
Group	e.g.
Member	Part Of
Definition	Located In
Lithology	
Geolocation	Data property:
Period	e.g.
StudyField	name
Well	Top Depth

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 TBox

Extract the data

Inserting instance

Write to rdf

Read 2 csv files, and properly clean the data to create URI for each entity

1. Refer to TBox
2. Set data type
3. Add it to model

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, XSDDatatype.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, XSDDatatype.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/to
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>
```

Finish creation of knowledge graph

Validation using Sparql Query

SPARQL

- A ***semantic querying language*** for databases

- ***PREFIX*** specified to use make use of anything defined in the ontology

- ***SELECT-WHERE*** clauses

- Variables specified by ***?***

- ***JOIN, SORT, AGGREGATE*** clauses for more RDBMS-like experience

- Popular ontologies that are often used:
  - ***PREFIX foaf: <http://xmlns.com/foaf/0.1/>***
  - ***PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>***
  - ***PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>***
  - ***PREFIX owl: <http://www.w3.org/2002/07/owl#>***

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> .

}

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> .

}

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

https://factpages.npd.no/en/wellbore/pageview/exploration/all/1382

```python
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)
```

```python
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
```
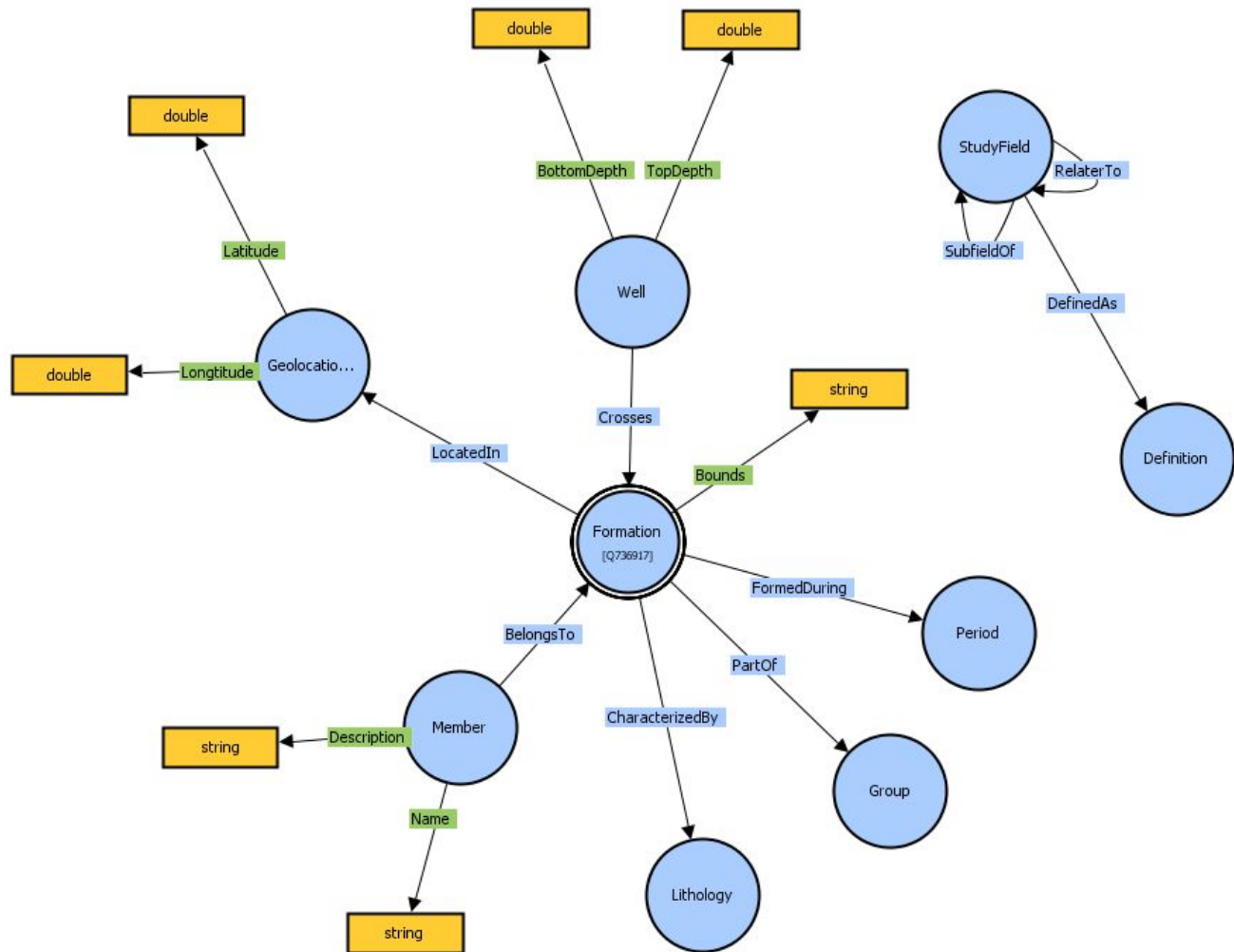
Loop parallelization to fasten up the scraping:

```python
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 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."

Summary of a Formation

Information we are interested in:

- The period and age of the formation
- The lithology description and characteristics of the formation
- The geolocation of the formation
- The group of the formation

Querying with Sparql

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

SELECT ?res
WHERE
 {
?formation stratig:Name   "Ekofisk Formation" ;
         stratig:FormedDuring ?period .
?period   stratig:Name   ?period_name .
?formation stratig:Bounds   ?boundary ;
         stratig:PartOf   ?group .
?group   stratig:Name   ?group_name .
?formation stratig:CharacterizedBy ?lithology .
?lithology stratig:Name   ?lithology_name
BIND(concat("The formation was originated in ", concat(?period_name, concat(" period and belongs to ", concat(?group_name, concat(". ", concat(?boundary, concat(" and lithology is", ?lithology_name)))))))) AS ?res)
          }

Summarization

```
if (input_question.contains("summary") || input_question.contains("best")) {
                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 ?res " +
                        "WHERE {" + formation_uri + " stratig:FormedDuring ?period ." +
                        "?period stratig:Name ?period_name ." +
                        formation_uri + "stratig:Bounds ?boundary ." +
                        formation_uri + "stratig:PartOf ?group ." +
                        "?group stratig:Name ?group_name ." +
                        formation_uri + "stratig:CharacterizedBy ?lithology ." +
                        "?lithology stratig:Name ?lithology_name . " +
                        "BIND( concat(\"The formation was originated in \", " +
                        "concat(?period_name, concat(\" period and belongs to \", " +
                        "concat(?group_name, concat(\". \", concat(?boundary, " +
                        "concat(\" and lithology is\", ?lithology_name))))))) AS ?res )}";
```

	res
1	"The formation was originated in Danian. period and belongs to shetland group. 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. and lithology is"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. ""

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

- Function:
  - question_parsing()
    - If the question contains stratigraphy
    - If the question contains name
  - query_selection()
    - Check keywords
    - Build Query
- Parse output into answers in paragraph.
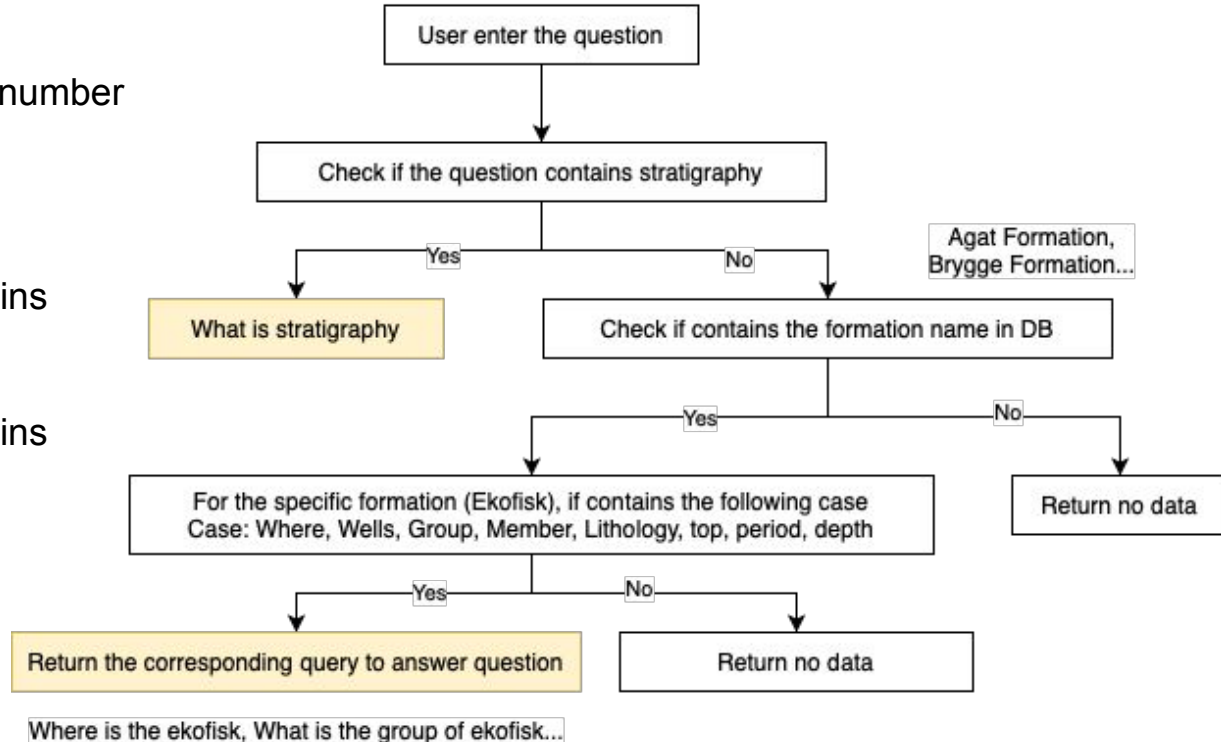
User enter the question

↓

Check if the question contains stratigraphy

Yes → What is stratigraphy

No → Check if contains the formation name in DB

Agat Formation, Brygge Formation...

Yes → For the specific formation (Ekofisk), if contains the following case
Case: Where, Wells, Group, Member, Lithology, top, period, depth

No → Return no data

Yes → Return the corresponding query to answer question

No → Return no data

Where is the ekofisk, What is the group of ekofisk...

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.

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"    |
```

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" |
--------------------
```

What is the lithology of _____ formation ?

```java
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. \"" |
-------------------------------------------------------------------------------
```

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. " |
------------------------------------
```

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" |
-------------------
```

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." |
-----------------------------------------------------------------------------------------------
```

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> |
--------------------------------------------------------
```

The Main Challenges

Challenges

1	Defining the scope of the project and following through with the objectives

2	Identifying and choosing the right data sources

3	Modelling geolocation into bounding box and technical conversion

4	Proper data clean

5	Learning new tools (Jena, Servlet)

Future improvement

1 **Integrate different stratigraphy sources page**

2 **Add flexibility for user's question/ auto correct**