# INFO116 - Gruppeoppgave

### "Making sense of whisky"

Kandidatnummer: 105, 122, 251, 157, 148

# Introduksjon

For denne oppgaven har gruppen vår arbeidet på en ontologi for konseptet whisky, med utgangspunkt i nettstedet whisky.com. Dette er et nettsted som har en stor database med informasjon om whiskyer, der mye av informasjonen er generert av brukerne av nettstedet. Vi vil vise eksempler på hvordan vi har konstruert ontologien vår med bruk av Protegè, SPARQL spørringer for uthenting av data, og annotering av nettsider fra whisky.com med RDFa-lite, JSON-LD i kontekst av ontologien vår og schema.org.

Vi vil forklare hva vi har fått til, hvorfor ontologien ble konstruert slik den ble og hvilke spørringer man kan bruke på den. Hvorfor vi valgte å redigere den originale kildekoden til whisky.com og hvilken nytte semantisk annotering gir i denne sammenheng.

# Ontologi

Når vi skulle lage ontologien vår startet vi med å se litt på tidligere oppgaver som vi hadde gjennomført på lab-øving. Vi forsto da at vi ville ha en struktur der vi kunne beskrive individer som en gitt whisky ville ha en nær relasjon til. Dette medførte at vi fikk klasser som Whisky med subklasser som SingleMaltWhisky, BlendWhisky osv. I tillegg kom f.eks; Country, Distillery, Region, WhiskyBottler o.l.

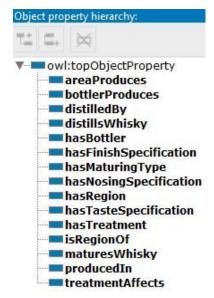


Vi strukturerte ontologien på denne måten fordi vi ønsket å være i stand til å kunne lage spørringer som, "Gi meg en whisky som er produsert i Islay, er av typen SingleMaltWhisky og er eldre enn 5 år".

Under konstruksjonen av ontologien tilpasset vi den løpende med egenskapene vi ville beskrive whiskyene med.

Eksempelvis alkoholprosent, flaskestørrelse, produsert av o.l.

Til venstre kan du se alle klassene/typene som er med i vår ontologi. Innenfor disse klassene lager vi enkeltindivider som vi har mulighet til å knytte til andre individer via objekt-egenskaper.



Som en ser til venstre på dette bildet kan vi knytte whiskyindivider til et destilleri ved å bruke *distilledBy*. Vi kan også lage et Region-individ, knytte dette med *isRegionOf* til et Country-individ, for så senere å bruke *producedIn* med en gitt region eller land på en whisky.

I tillegg la vi opp inverse egenskaper til enkelte der vi følte behovet. Dette vil si at *hasBottler* er det inverse av *bottlerProduces*. Den første knytter en whisky til en bottler, og den andre knytter en bottler til en whisky. Dette gjør vi for å ha flere muligheter når man vil lage spørringer, men også

fordi det gir et bredere bilde av sammenhengen mellom f.eks: bottler og whisky.

I tillegg hadde vi flere egenskaper som knytter

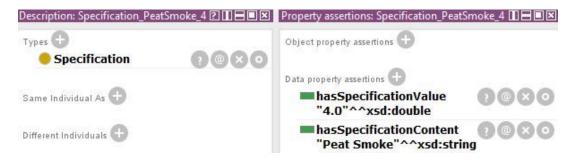
enkelt verdier til individer. Ved hjelp av disse kan vi knytte; alkoholprosent, alder, rating, id, navn o.l, til de gitte individene.

To egenskaper som kanskje skiller seg spesielt ut her er hasSpecificationContent og hasSpecificationValue. Disse beskriver det vi kaller Specification-individ i vår ontologi. Som vi ser under har dette individet to verdier, "4.0" og "Peat

owl:topDataProperty
hasABV
hasAge
hasBottlesize
hasBottlingDate
hasDistillationDate
hasID
hasName
hasRating
hasSpecificationContent
hasSpecificationValue

Data property hierarchy:

Smoke". Vi brukte dette til å si at en whisky har en egenskap ved seg som er "Peat Smoke" og at dette har en verdi av 4.



Hvilken egenskap dette definerer av en gitt whisky bestemmer en gjennom hasNosingSpecification, hasTasteSpecification og hasFinishSpecification, som knytter en gitt whisky til et Specification-individ.



# **SPARQL**

### Setting

Vi har sett for oss at vi har et scenario der vi er i prosessen med å utvikle en app for et nettsted som omhandler whisky. Dataene er per dags dato relativt ustrukturerte. Ved hjelp av vår løsning er målet å kunne kjøre nyttige spørringer som kan ta høyde for mange ulike aspekt ved de gitte whiskyene på nettstedet. Vi har konstruert over 10 spørringer for å vise fleksibiliteten man har med ontologien. Vi har lagt ved scenario eller "use-case" til disse. Alle spørringene ligger ved i en tekstfil (i tillegg er der en fil med resultat sett), her viser vi 5 (utdrag fra hver case) av dem:

#### Case 1

Kunden ønsker å ha et enkelt recommender-system for å finne lignende whiskyer som brukeren også kan være interessert i, etter et søk allerede har forekommet.

Finn whiskyer med lignende smak

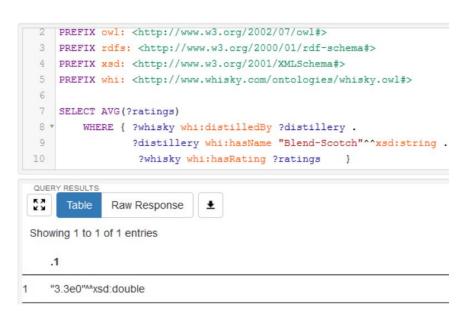
```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <a href="http://www.w3.org/2002/07/owl#>">PREFIX owl: <a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#></a>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a>
PREFIX whi: <a href="http://www.whisky.com/ontologies/whisky.owl">http://www.whisky.com/ontologies/whisky.owl</a>
SELECT DISTINCT ?name
WHERE { ?whisky whi:producedIn ?place .
?place whi:hasName "Islay" ^ ^xsd:string .
?whisky whi:hasTasteSpecification ?tasteSpecification .
?tasteSpecification whi:hasSpecificationContent ?originalContent .
?similarWhiskies whi:hasTasteSpecification ?otherSpecification .
?otherSpecification whi:hasSpecificationContent ?originalContent .
?similarWhiskies whi:hasName ?name .
FILTER (?similarWhiskies != ?whisky) }
"Bakers 7 Years Case Proof"^^<http://www.w3.org/2001/XMLSchema#string>
```

<u>Forklaring</u>: Her ser vi at kunden har et søk der han vil ha whiskyer som ble produsert i "Islay". Etter dette spør vi hvilke smaker som tilhører de gitte whiskyene som kunden fant. Så spør vi om hvilke av whiskyene i ontologien som har smaker til felles med de som kunden hadde søkt etter, og så viser vi navnet til disse. For resultatet er filter funksjonen brukt for å fjerne whiskyene som kunden allerede fant i første søk.

#### Case 2

I appen skal det også være mulig å finne gjennomsnittsratingen for en gitt produsent eller destilleri, på grunnlag av hva de har produsert.

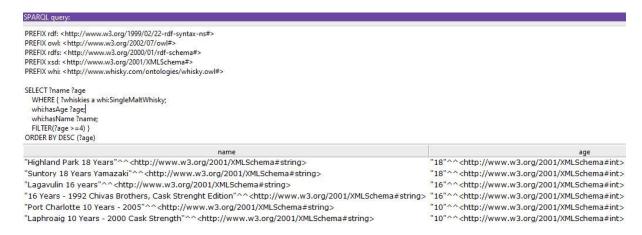
Gjennomsnittsrating på et destilleri



#### Case 3

Kunden ønsker å finne whiskyer basert på spørringer med mange faktorer. I vår ontologi har vi beskrevet whiskyene med følgende informasjon: produsent, destilleri, etterbehandling, modning (les: cask), produksjonssted, lukt, smak, ettersmak, rating (vurdering), alkoholprosent, alder, navn, flaskestørrelse, destilleringsdato og tappedato.

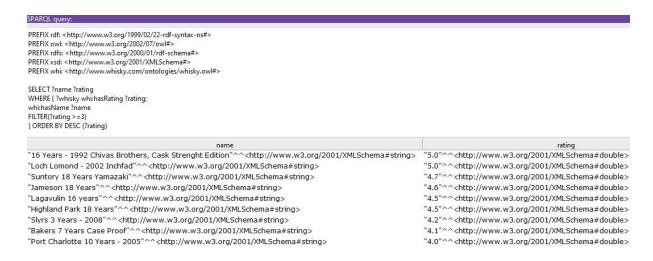
 Her er brukeren ute etter whiskytypen SingleMaltWhisky, med alder til og med 4 og over. Resultatet er sortert etter alder på whiskeyen i synkende rekkefølge.
 Kolonnenenavn og alder vises for brukeren.



#### Case 4

Siden vi har støtte for å beskrive whiskyer med "rating" har vi lyst til å sortere de ulike resultatene (slikt som man finner fra spørringer som tilhører case 3) på denne variabelen.

 Denne queryen henter alle whiskeyene, og gir oss rating som er høyere enn 3, og sorterer resultatet synkende. Vi ser gjerne for oss at dette er en spørring man kan nøstes med andre spørringer som en gitt bruker har benyttet.



#### Case 5

I appen skal der være funksjonalitet som kan tillate en gitt bruker å finne en whisky basert på faktorer gitt i case 3, men også kunne velge å få vist den som har høyest/minst alkoholprosent og høyest/minst lukt, smak eller ettersmak, eventuelt en kombinasjon av disse.

 Her har brukeren valgt å finne en whisky med høyest alkoholprosent innenfor alle Single Malt Whisky.

```
SPARQL query:

PREFIX owl: <a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#">http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#">http://www.whisky.com/ontologies/whisky.owl#">http://www.whisky.com/ontologies/whisky.owl#</a>

SELECT ?name ?abv

WHERE { ?whiskies a whi:SingleMaltWhisky;
whi:hasABV ?abv;
whi:hasABV ?abv;
whi:hasName ?name }

ORDER BY DESC (?abv)
LIMIT 1

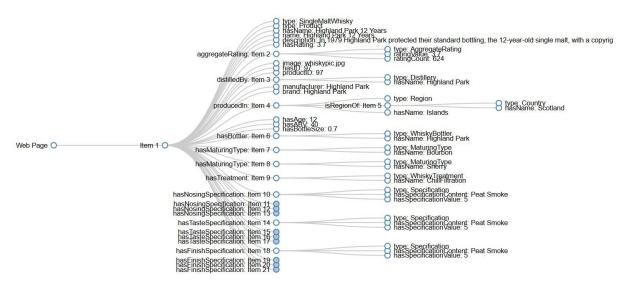
name

abv
```

<sup>&</sup>quot;16 Years - 1992 Chivas Brothers, Cask Strenght Edition" "58.4"^^<http://www.w3.org/2001/XMLSchema#double>

# Annotering av nettsider

### RDFa-lite og JSON-LD



Vi tok et utvalg av 5 whiskyer fra whisky.com og annoterte informasjonen som var tilgjengelig på disse. På bildet over ser vi en graf av RDFa dataene som ligger inne i en av de annoterte sidene (noen av dataene er kollapset for å gi et bedre overblikk). Vi brukte både vår egen ontologi og schema.org for å beskrive informasjonen tilgjengelig. For å validere annoteringsprosessen ble "<a href="http://rdfa.info/play/">http://rdfa.info/play/</a>" brukt for å lage grafen over, "<a href="https://search.google.com/structured-data/testing-tool">https://search.google.com/structured-data/testing-tool</a>" for å skjekke både JSON-LD og RDFa, samt "<a href="https://json-ld.org/playground/">https://json-ld.org/playground/</a>" for ekstra skjekk av JSON-LD.

```
<!-- The description of sch:Product -->
cp property="sch:description">In 1979 Highland Park protected their
    standard bottling, the 12-year-old single malt, with a copyright.
    Today it is still the most important
    malt in the range of the distillery.

<img src="rating.png" />
<!-- The aggreagate rating of the whisky -->
cp id="rating"><span property="hasRating">3.7</span> (624)

<!-- The aggreagate rating of the sch:Product -->
<span property="sch:aggregateRating" typeof="sch:AggregateRating">
    <span property="sch:ratingValue" content="3.7"></span>
    <span property="sch:ratingCount" content="624"></span>
</span>
</span>
```

Til venstre ser vi at "sch:" er brukt som en prefix for å henvise til schema beskrivelser, og ellers har vi hasRating, som er fra vår egen ontologi. Flere steder er lik informasjon beskrevet med både schema og ontologien.

```
cript type="application/ld+json">
 "@context": "http://www.whisky.com/ontologies/whisky.owl#",
                                                                      Her ser vi bruk av JSON-LD for å
 "@type": "SingleMaltWhisky",
                                                                      implementere data både fra schema
 "hasAge": "12",
"hasABV": "40",
                                                                      og ontologien vår. Dette er en av
                                                                      whiskyene vi valgte å annotere som
 "hasName": "Highland Park 12 Years",
                                                                      heter "Highland Park 12 Years".
 "hasMaturingType": [{"@type": "MaturingType", "hasName": "Bourbon"},
 {"@type": "MaturingType", "hasName": "Sherry" }],
 "hasName": "ChillFiltration"}],
                                                <script type="application/ld+json">
  'producedIn": {
   "@type": "Region",
                                                    "@context": "http://schema.org/",
                                                    "@type": "Product",
   "isRegionOf": {
                                                    "url": "HighLand_P12Years.html",
     "@type": "Country",
                                                    "productID": "39",
     "hasName": "Scotland"
                                                    "brand": "Highland Park",
                                                    "manufacturer": "Highland Park",
                                                    "name": "Highland Park 12 Years",
 "distilledBy": {
    "@type": "Distillery",
                                                    "image": "whiskypic.png",
                                                    "aggregateRating": {
   "hasName": "Highland Park"
                                                      "@type": "AggregateRating",
                                                      "ratingValue": "3.7",
                                                      "ratingCount": "624",
   "@type": "WhiskyBottler",
                                                      "bestRating": "5"
   "hasName": "Highland Park"
                                                    "description": "In 1979 Highland Park protected their standard bottling,
```

All den annoterte informasjon over gjør at vi kan teste dette med "rich snippet" verktøyet fra google; for å vise at informasjonen kan bli tolket og evt. videre prosessert. JSON-LD under.

Product	ERRORS 0 WARNINGS ^
@type	Product
url	HighLand_P12Years.html
productID	39
name	Highland Park 12 Years
image	whiskypic.png
	In 1979 Highland Park
	protected their standard
	bottling, the 12 - year - old
description	single malt, with a
	copyright. Today it is still the
	most important malt in the
	range of the distillery.
brand	
@type	Thing
name	Highland Park
manufacturer	
@type	Organization
name	Highland Park
aggregateRating	
@type	AggregateRating
ratingValue	3.7
ratingCount	624
bestRating	5

https://search.google.com /structured-data/testing- tool/SingleMaltWhisky	0 ERRORS WARNINGS
	https://search.google.com
@type	/structured-data/testing-
	tool/SingleMaltWhisky
url	HighLand_P12Years.html
hasAge	12
hasABV	40
hasBottlesize	0.7
hasName	Highland Park 12 Years
hasMaturingType	
	https://search.google.com
@type	/structured-data/testing-
	tool/MaturingType
hasName	Bourbon
hasMaturingType	
	https://search.google.com
@type	/structured-data/testing-
	tool/MaturingType
hasName	Sherry
hasTreatment	
	https://search.google.com
@type	/structured-data/testing-
	tool/WhiskyTreatment
hasName	ChillFiltration
producedIn	
@type	https://search.google.com
	/structured-data/testing-tool/Region
hasName	Islands

(JSON-LD for vår ontologi)

Bildet som viser JSON-LD med vår egen ontologi er kun et utdrag av informasjonen som ligger der, fordi listen er svært lang med lukt, smak o.l. Bildet som viser JSON-LD for schema.org er så godt som indentisk til hvordan verktøyet tolket RDFa for den annoterte siden, dette er fordi vi har prøvd å bruke samme "beskrivelser" både i JSON-LD og RDFa. Denne er derfor ikke avbildet eksplisitt.

nttp://www.whisky.com/ontologies /whisky.owl#SingleMaltWhisky	0 0 ERRORS WARNINGS
@type	http://www.whisky.com/ontologies
	/whisky.owl#SingleMaltWhisky
http://www.whisky.com/ontologies /whisky.owl#hasName	Highland Park 12 Years
http://www.whisky.com/ontologies /whisky.owl#hasRating	3.7
http://www.whisky.com/ontologies /whisky.owl#hasAge	12
http://www.whisky.com/ontologies /whisky.owl#hasABV	40
http://www.whisky.com/ontologies /whisky.owl#hasBottleSize	0.7
http://www.whisky.com/ontologies /whisky.owl#distilledBy	
@type	http://www.whisky.com/ontologies /whisky.owl#Distillery
http://www.whisky.com /ontologies/whisky.owl#hasName	Highland Park
http://www.whisky.com/ontologies /whisky.owl#producedIn	
@type	http://www.whisky.com/ontologies /whisky.owl#Region
http://www.whisky.com /ontologies/whisky.owl#hasName	Islands
http://www.whisky.com /ontologies /whisky.owl#isRegionOf	
@type	http://www.whisky.com/ontologies /whisky.owl#Country
http://www.whisky.com /ontologies /whisky.owl#hasName	Scotland
http://www.whisky.com/ontologies /whisky.owl#hasBottler	
@type	http://www.whisky.com/ontologies /whisky.owl#WhiskyBottler
http://www.whisky.com /ontologies/whisky.owl#hasName	Highland Park

På bildet til venstre ser vi hvordan verktøyet til google har tolket RDFa-informasjonen som tilhører ontologien vår, og som vi har annotert en nettside med.

Dette er også kun et utdrag av informasjonen som blir vist, siden listen er svært lang.

(RDFa-Lite data med bruk av ontologien)

### Kommentarer angående kildekoden

Vi begynte først med å laste ned en webside fra whisky.com og forsøkte å tagge siden, men vi innså at det kom til å bli svært utfordrende å gjøre det med god kvalitet. Websiden har over ni tusen linjer kode, og annoterings-prosessen i seg selv var utfordrende nok i starten før en ble vant til det. <u>Vi kontaktet derfor foreleser for å høre om det gikk ann å lage en mockup side som tok utdrag fra den originale kildekoden som vi trengte. Dette var akseptert så lenge vi tok hensyn til den originale strukturen som tilhørte sidene.</u>

Dette gav oss mulighet til å presentere de samme dataene som var nødvendig for vår bruk av annotering med RDFa og JSON-LD, men på et par hundre linjer i stedet for rundt ni tusen.

Først ble det laget en enkel utgave av siden. Vi matet deretter inn informasjon om en whisky og begynte arbeidet med å tilføye RDFa og JSON-LD til denne. Etter vi var tilfreds med denne ene whiskyen, brukte vi dette som en mal for resten vi ville annotere. Eksempler på dette ble vist tidligere i teksten.

### Fordeler med den tilføyde semantikken

Med semantikken som er blitt lagt til tilbyr nettsidene nå rik og strukturert informasjon. Dette gjør for eksempel at søkemotorer kan enklere finne relevant informasjon på websidene, og presentere detaljerte data til brukerne. For eksempel kan produktinformasjon som alkoholprosent, smak osv. hjelpe brukeren raskere med å finne informasjonen de er ute etter, uten å nødvendigvis klikke seg inn på websiden.

Vi la både til RDFa-Lite og JSON-LD til nettsidene, dette gjør dataene mer søkbare, generelt sett mulige å kategorisere, og kan knyttes til både schema.org og ontologien vår (om vi hadde "hostet" denne ontologien på nettet)

Kanskje man googler whisky som har en spesiell type smak. Ved hjelp av den semantiske informasjonen kan websiden nå tilby disse dataene.

# Gruppearbeid og konklusjon

For å ha kontroll på prosjektstatus både under samlingene våre, og når vi jobbet hjemmefra brukte vi mye "<a href="https://todoist.com/">https://todoist.com/</a>". For å koordinere samlinger brukte vi rombooking via UiB slik at vi kunne samles for diskutere hvordan vi skulle gå løs på selve oppgaven, og avtale hvilke målsetninger vi burde sette innenfor gitte tidsbegrensninger. Dette fungerte svært bra. Etterhvert ble vi mer komfortable med Protegè og skjønte hvordan vi skulle annotere nettsidene. Vi sørget tidlig for å sette opp et "Github-repository" for å ha solid versjonskontroll på alle data vi arbeidet med.

Samtidig som arbeidet med det rent tekniske pågikk ble det laget utkast av informasjon som vi ønsket å få med i rapporten. Vi førte også logg av oppmøte under samlingene vi hadde under hele prosessen. Vi legger ved logg lenger nede som gir en viss indikasjon på deltagelse, det ble dog arbeidet relativt mye mellom disse møtene.

Hele prosessen har gitt oss et godt innblikk i hvordan man kan skape ontologier i Protegè, og kjøre SPARQL spørringer mot disse. I tillegg har vi lært å legge til semantisk informasjon mot websider med både RDFa-Lite og JSON-LD.

### Logg

### 24.10.17 - Gruppemøte

Deltagere: kanditat-157, kanditat-105, kandidat-122, kandidat-251

Vi satte opp en prosjektplan i Todoist, opprettet repository i github, og påbegynt rapport i Google Docs. Vi begynte å se på hva informasjon som ligger på enhetene på whisky.com

### 27.10.17 - Gruppemøte

Deltagere: kanditat-157, kandidat-251, kandidat-122

Satt opp alle klasser ut i fra informasjonen websiden har tilgjengelig.

### 31.10.17 - Gruppemøte

Deltagere: kanditat-157, kanditat-105, kandidat-251

Arbeid med ontologi i protege. Forsøkt å legge opp en test-flaske for å se hva vi mangler og hva som fungerer eller ikke fungerer.

### 01.11.17 - Gruppemøte

Deltagere: kandidat-251, kanditat-157

Arbeid med å legge opp properties i ontologien. Store utfordringer med å finne ut hva vi skal bruke fra whisky.com fordi det er svært varierende mellom typene om hva som ligger på de av informasjon. Enkelte viser knapt ABV, mens andre har distillation date, bottling date, bottle size, bottler, series, maturing osv. Enkelte har ikke en gang distillery eller brand.

### 06.11.17 - Lab-gruppe

Deltagere: kandidat-251, kanditat-157, kandidat-122, kanditat-105

Påbegynt arbeid med å sette opp sparql-queries. Tenkt opp et use-case for bruken av queryene og ontologien vi har laget.

### 08.11.17 - Gruppemøte

Deltagere: kanditat-157, kandidat-251, kanditat-105

Arbeid med å legge opp tasting, nosing og finish på en sample whisky for å demonstrere mulighetene ved querying.

### 09.11.17 - Gruppemøte

Deltagere:kanditat-157, kandidat-251, kanditat-105, kandidat-122

Arbeid med å lage use cases\scenario som vil være relevante og nyttige utgangspunkt for queryene som skal utformes.

### 10.11.17 - Gruppemøte

Deltagere: kanditat-105, kanditat-157, kandidat-251, kandidat-122

Påbegynt utforming av SPARQL-queries. Rydding i rapporten. Avtalte at vi jobber videre hjemme hver for oss med å få ferdig de ulike queryene.

### 13.11.17 - Lab-gruppe

Deltakere: kanditat-105, kandidat-122, kanditat-157, kandidat-251, kanditat-148 Påbegynt å utforme whisky sider med RDFA. Planlagt å lage en mal side som alle kan jobbe ut fra. Testing og øving på RDFa-annotering

### 16.11.17 - Gruppemøte

Deltagere: kandidat-251, kanditat-157, kanditat-105, kanditat-148

Kollektiv gjennomgang av vår første semantiske webside, før vi bruker denne som mal og legger opp flere typer whisky.