Anxiety Disorders & Risk Factors Ontology

Final Project Report Knowledge and Data 2024 - [redacted]

Maggie C[redacted] - [redacted]
Holly M[redacted] - [redacted]
Kayra Ö[redacted] - [redacted]
M[redacted] Fatih A[redacted] - [redacted]

Date: 28-10-2024

1 Introduction

1.1 Goal, motivation and users

Anxiety disorders are one of the most common mental health issues and have a significant impact on people's daily lives. Understanding the risk factors and genetic predispositions can inform public health strategies, medical research, and an individual's life choices. Therefore, analysing large datasets and merging them with relevant ontologies can help show connections between environmental and biological risk factors.

Stakeholders for this project include: medical researchers, healthcare providers, public health officials, and individuals or patients. This analysis provides researchers with insights into genetic markers and other risk factors linked to anxiety. By utilising the created ontology researchers can look into new treatments, diagnostic processes and preventive measures for anxiety. Similarly, health care providers can also use it to better treat and diagnose anxiety in their patients. Public health officials can use findings from data retrieved through SPARQL queries to develop nationwide mental health initiatives aimed at reducing anxiety in the population. Individuals or potential patients can view data graphs generated from this ontology to see if they are predisposed to anxiety and make relevant and informed lifestyle changes to manage their mental health to the best of their ability.

1.2 Competency Questions

1) How do patterns of reluctance to sleep, sleep resistance, and time taken to fall asleep influence an individual's development of anxiety disorders?

This question aims to answer the effects of various sleep-related problems on anxiety disorders. Sleep health plays a major role in anxiety disorders, as anxiety often worsens sleep quality [1]. Understanding how sleep disturbances influence anxiety disorders may provide stakeholders with insights into how to approach and deal with these conditions. Visualisations of sleep data will be done to showcase how sleep quality is an important indicator of an individual's anxiety disorders.

2) How do family psychiatric histories of substance abuse and psychiatric diagnoses correlate with the probability of anxiety disorders?

This question was chosen to explore hereditary factors and their link with anxiety disorders. By identifying these patterns, the project can provide valuable insights into familial

mental health risks, which can help individuals be aware of their risk of developing an anxiety disorder and take preventative steps as well as be aware of the signs when an anxiety disorder starts developing. Data visualisations on the family histories of substance abuse and psychiatric diagnoses will be conducted to underlie how these parameters are big factors on anxiety.

3) To what extent are genetic markers for anxiety disorders also associated with post-traumatic stress disorder and obsessive-compulsive disorder?

This question narrows down on the genetic aspects of mental disorders and examines the interaction between anxiety disorders and other mental illnesses, specifically post-traumatic stress disorder (PTSD) and obsessive-compulsive disorder (OCD). The American Psychological Association (APA) indicates that separation anxiety disorder is correlated with and comorbid with PTSD and OCD [2]. However, no other comorbidities have been indicated on other anxiety disorders with ADHD and OCD. Stakeholders can gain insights into how many genes associated with anxiety are also linked to the aforementioned disorders.

2 External sources identified

2.1 Existing Ontologies identified

Medical Subject Headings

Source: https://bioportal.bioontology.org/ontologies/MESH/?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMESH%2FD001523&lang=en

Description: The Medical Subject Headings (MeSH) ontology, developed by the National Library of Medicine, is an ontology for indexing and cataloguing biomedical and health-related information. It categorises a wide range of medical topics, including diseases, chemicals, and procedures, using a structured hierarchy that supports both broad and specific searches.

Motivation: This ontology was chosen due to its extensive “Mental Disorders” class, which organises mental disorders into categories based on type, such as Anxiety Disorders, Mood Disorders, and Personality Disorders. We will be specifically focusing on the Anxiety Disorders classes. This classification is particularly valuable for mapping and understanding various mental disorders, as it provides a well-established framework.

NCI Thesaurus (NCIt) OBO Edition

Source: *http://purl.obolibrary.org/obo/NCIT_C2878*

Description: The NCI Thesaurus (NCIt) is a comprehensive ontology that covers a wide range of topics within the cancer domain, including cancer-related diseases, findings and abnormalities. The NCIt OBO Edition is part of the Open Biological and Biomedical Ontology (OBO) Library, enhancing its usability with other ontologies in the biomedical domain.

Motivation: This ontology was chosen due to its coverage of “Psychiatric Disorders”, which complements the Medical Subject Headings (MeSH) ontology when it comes to types of anxiety disorders. While MeSH provides a foundation, it lacks some psychological disorder classifications. By using NCIt, this project aims to provide a more complete representation of various anxiety disorders.

2.2 External SPARQL endpoints identified

DBpedia
Source: https://dbpedia.org/sparql
Description: DBpedia is a community-driven project that extracts information from Wikipedia, allowing users to query information about various topics, including mental health.
Motivation: This dataset allows us to extract relationships and properties related to mental disorders.

2.3 External non-RDF Datasets identified

Anxiety Disorders Dataset

Source: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/N42LWG
Description: This dataset includes behavioural, psychophysiological, and demographic information from a sample of 193 preschool-aged children (3-5 years old) to assess potential risk factors for Generalised Anxiety Disorder (GAD) and Separation Anxiety Disorder (SAD). This data was originally gathered for the study “Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach.” It can be used to build machine learning models aimed at quantifying anxiety risk among young children. It has been acquired in Excel format.
Motivation: By using this dataset, the project can explore risk factors specific to anxiety, which allows us to also create interesting visualisations related to those risk factors, such as sex and socioeconomic status.

Anxiety Disorder Genetic Dataset

Source: https://www.alliancegenome.org/disease/DOID:2030
Description: This dataset looks at genes, alleles, and genetic models that are associated with diseases including anxiety, ptsd, ocd, panic disorders and acute stress disorder. It states whether

some variant of a gene is “implicated in”, meaning to be shown to cause or modify a disease for a human, and is “a marker for”, meaning when there is evidence of an association but insufficient evidence to establish causality. It has been acquired in Excel format.
Motivation: By using this dataset, the project can explore genetic risk factors that predispose individuals specifically to anxiety. This then allows us to strengthen implications gathered from the first dataset, and to create interesting visualisations that show possible links between risk factors from the previous dataset, such as sex and socioeconomic status, and genetic factors.

3 Design of Ontology

3.1 Methodology

This project’s scope consists of mental disorders, more specifically anxiety disorders. The external ontologies are going to be used to extract information about the disorders in the form of classes, and additional classes will be created in *Protégé* to bridge the gap between the ontologies. The datasets will provide information regarding risk factors of the disorders, and will form the properties. The SPARQL query will then be used to populate the instances, and OWL restrictions will be applied. *Pellet* has been selected as the reasoner of the ontology, and will be run consistently throughout the ontology creation process in *Protégé* to ensure no inconsistencies occur.

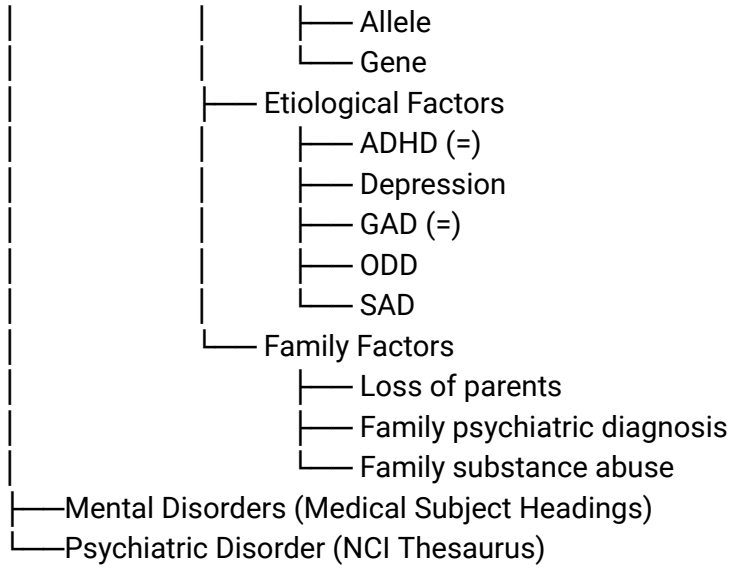
3.2 Conceptualisation

There are a total of 20 classes in addition to classes from external ontologies. The class ‘Mental Disorders’ is a specific subclass derived from the Medical Subject Headings ontology and ‘Psychiatric Disorder’ is the entirety of NCI Thesaurus ontology. In total, the ontology consists of 17 properties (12 object properties, 5 data properties), and 2 restrictions.

The ontology is defined with the prefix ‘anx’, which is associated with the base URI ‘http://AnxietyDisordersandRiskFactors.com#’.

3.2.1 Classes

owl:Thing
├── Group17-Classes
│ ├── Disorder (=)
│ ├── Species
│ ├── Patient
│ │ ├── Female
│ │ └── Male
│ └── Risk Factors
│ ├── Genetic Factors



**Classes with the (=) symbol have been mapped as equivalent to other classes from external ontologies.*

3.2.2 Properties

Object Properties:
Below are the object properties that are defined in the ontology.

Object Property	Domain	Range	Characteristics
Has_Risk_Factor	Patient	Risk Factor	Irreflexive
Is_Implicated_In	Genetic Factors	Disorder	Asymmetric, Irreflexive
Is_Not_Implicated_In	Genetic Factors	Disorder	Asymmetric, Irreflexive
Is_Implicated_Via_Orthology	Genetic Factors	Disorder	Irreflexive
Is_Model_Of	Genetic Factors	Disorder	Irreflexive
Susceptibility_To	Species	Disorder	Asymmetric, Irreflexive
Is_Marker_For	Genetic Factors	Disorder	Irreflexive
Is_Marker_Via_Orthology	Genetic Factors	Disorder	Irreflexive
Is_Expressed_In	Genetic Factors	Disorder	Irreflexive
Is_Observed_In	Disorder	Species	Irreflexive

Severity_Of	Species	Disorder	Irreflexive
Sexual_Dimorphism_In	Species	Disorder	Irreflexive

Table 3.2.2.1: Object properties, domains, ranges, and characteristics of the ontology

Data Properties:
Below are the data properties that are defined in the ontology.

Data Property	Domain	Range	Characteristics
hasAge	Patient	xsd:integer	Functional
isInPoverty	Patient	xsd:boolean	Functional
hasHoursToFallAsleep	Patient	xsd:integer	Functional
HasSleepDisturbances	Patient	xsd:integer	Functional
HasSleepResistance	Patient	xsd:integer	Functional

Table 3.2.2.2: Data properties, domains, ranges, and characteristics of the ontology

3.2.3 Restrictions

- Genetic Factors
  - Restrictions:
    - EQUIVALENT TO is_implicated_in some ‘Disorder’
    - SUBCLASS OF is_marker_for some ‘Disorder’

3.3 Formalization / Implementation

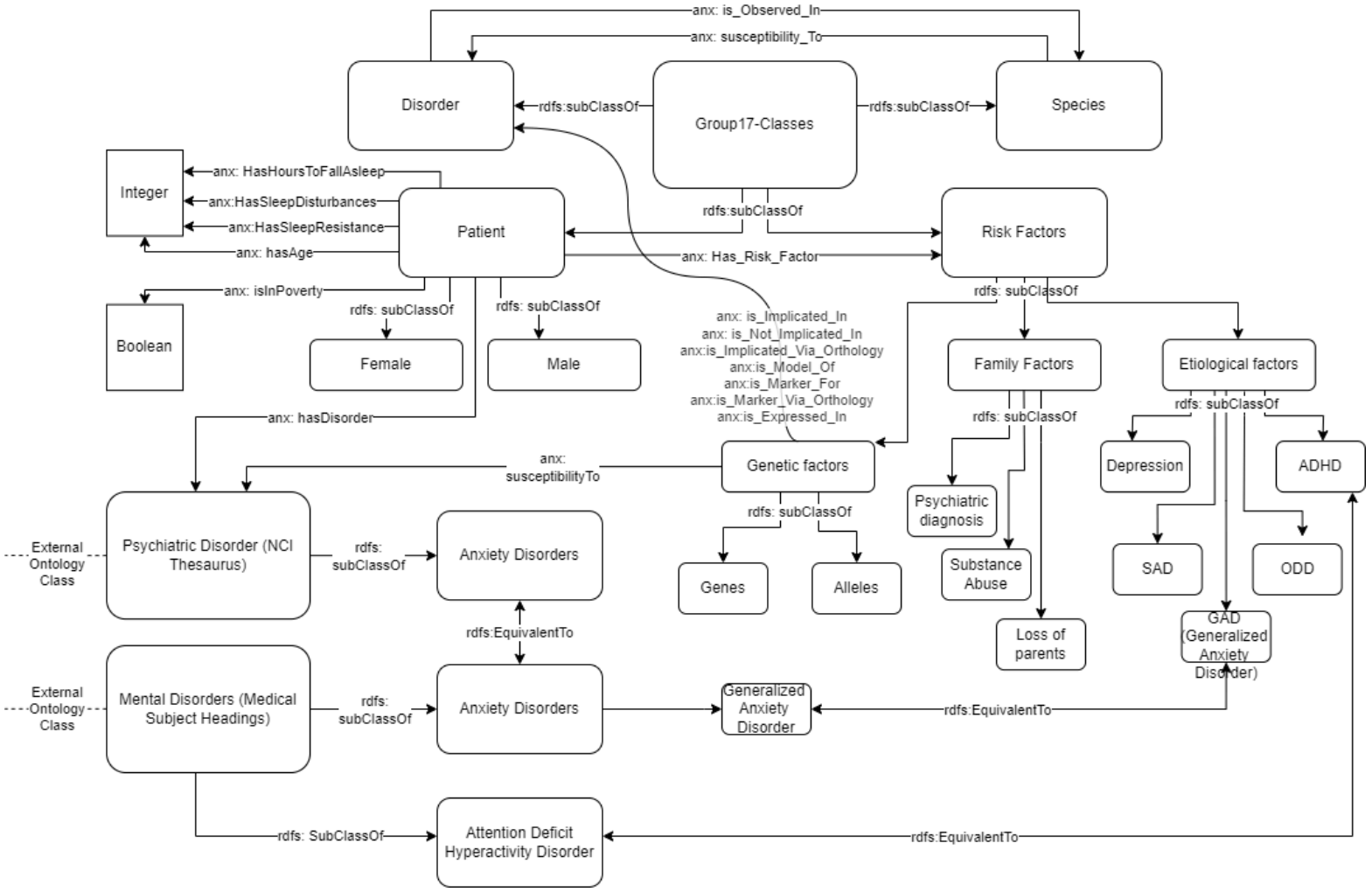


Figure 3.3.1: Conceptualization diagram of the ontology

4 Data Integration and conversion

The conversion was done in OntoRefine, and the two external non-rdf datasets identified were utilised for the conversion. The first external dataset provided the instances (patients), their risk factors (as classes), and properties linking the risk factors to the instances. The dataset was in the form of two excel files, and during the conversion process the two files were imported together into OntoRefine and treated as one dataset. Some columns in the dataset had values that weren't formatted correctly with respect to the ontology, so some values' formats have been altered via the "Transform…" button in the "Edit Cells" section of the column that had the issue (e.g. "M" was changed to "Male" to comply with the data integration process easily). The alteration was done via the General Refine Expression Language (GREL) (code for previous example in GREL: if(value == "M", "Male", value)).

Next, in the "Edit RDF Mapping" section, many mappings have been created to populate the instances of the ontology. First, all subjects have been initialised to a *Sex* class in compliance with the *Male* and *Female* subclasses of the *Patient* class. To extract all subject columns, the '@'

symbol was placed in between the namespace and the value, and then declared to be whichever gender data they have in the *Sex* column (i.e. :@Subject a :@Sex).

Furthermore, all subjects are mapped to properties from the original ontology, which were created by us, in the following format: :@Subject anx:hasAge @Age^^xsd:integer. It should be noted that the pre-existing property has a different namespace— this is because for any property that was coded in and was not from an external ontology, the “anx:” namespace was used. For external ontologies, the “:” namespace shorthand was used. The next steps consisted of recursively deciding whichever information is necessary from the dataset, creating properties to link the information together, and extracting the instances.

The second dataset brought forth *Species*, *Genes*, *Alleles*, *Disorders* (classes); *Disease Qualifiers*, *Genetic Associations*, and *Allele Associations* (object properties) from the extractions. New object properties for the “*anx:*” were also initialised (expressedIn, isObservedIn) and populated with the extractions. The second dataset was in the form of two TSV files, and instead of doing the importation of the two files together this time, the first TSV file was mapped in the RDF Mapper of OntoRefine. Next, the mapping was downloaded in JSON format, and for the other TSV file the JSON mapping was uploaded and the structure was changed (as one TSV file concerned genes while the other concerned alleles). Once the mappings were completed, they were merged. The resulting turtle files were imported into the ontology, and the reasoner was run. After the process of eliminating the reasoner conflicts, the conversion phase was completed.

5 The resulting knowledge graph

The integrated knowledge graph unifies mental disorders, risk factors, and patient data from external sources, creating a structured representation of anxiety disorders and their associated influences. External mental disorder classes are mapped from the external ontologies under "Mental Disorders (Medical Subject Headings)" and "Psychiatric Disorders (NCI Thesaurus)", and can be accessed alongside project-defined classes grouped as "group17-Classes." This grouping serves as a single repository for accessing data, instances, and inferences related to the project.

The knowledge graph includes three major classes:

1. Patients
2. Risk factors, categorised into etiological factors, family factors and genetic factors
3. Mental disorders

Each patient is linked to various risk factors, including both familial and etiological influences. These connections are modelled using data properties and object properties that represent specific risk probabilities for various anxiety disorders such as acute stress disorder, generalised anxiety disorder, separation anxiety disorder and more. Each risk factor is linked to specific patients, reflecting how these factors contribute to their overall risk for various anxiety disorders. All data properties are defined as owl:Functional, ensuring that each Patient is associated with a single, unique value for each risk factor. Furthermore, all object properties are designated as owl:Irreflexive, indicating that no entity can have a relationship with itself through these properties.

Additionally, genes and alleles from the Anxiety Disorder Genetic Dataset are incorporated as subclasses of genetic factors. While these genetic markers are not directly

linked to individual patients, they provide insights into broader trends and associations between genetic predispositions and anxiety disorders. The graph also distinguishes which genes and alleles are expressed in different species and whether they are implicated in certain anxiety disorders. Both alleles and genes are represented as distinct instances in the ontology, enhancing the scope of the knowledge graph for research into genetic factors across species.

Overall, the knowledge graph includes a total of 1,325 patients, of which 918 have unknown sex, while 141 are identified as female and 266 as male. The average age of the patients in the graph is 3.4 years. In terms of genetic information, there are 18 instances of alleles and 1,341 instances of genes, illustrating the rich genetic data integrated into the graph.

6 Meaningful Inferences

Through the use of a reasoner (Pellet), several meaningful inferences were drawn from the integrated knowledge graph. These inferences allow stakeholders to uncover hidden relationships between anxiety disorders, genetic markers, and risk factors, providing valuable insights for research and treatment.

The Anxiety Disorders Dataset originally consisted of a “training” dataset and a “testing” dataset, structured for machine learning purposes. However, for this project, we integrated both datasets to utilise the full scope of the available data. A key challenge arose from the fact that the datasets contained different types of information for each individual. Notably, sex information was absent from the training dataset, while the testing dataset included this data.

During the data integration phase, patients from the testing dataset were correctly classified into the subclasses Male and Female, which are subclasses of Patient. This automatically categorised them as Patients in the graph. However, patients in the training dataset lacked sex information and were not initially classified as Patients. After running the Pellet reasoner, the system inferred that 918 individuals from the training dataset should also be categorised as Patients, based on the fact that all individuals linked to a has_risk_factor property must belong to the class Patient. This automatic assignment of class memberships ensured consistency across both datasets, aligning all individuals under the correct class structure. These inferences are crucial for future analyses that involve identifying patient profiles and their associated risks for specific anxiety disorders.

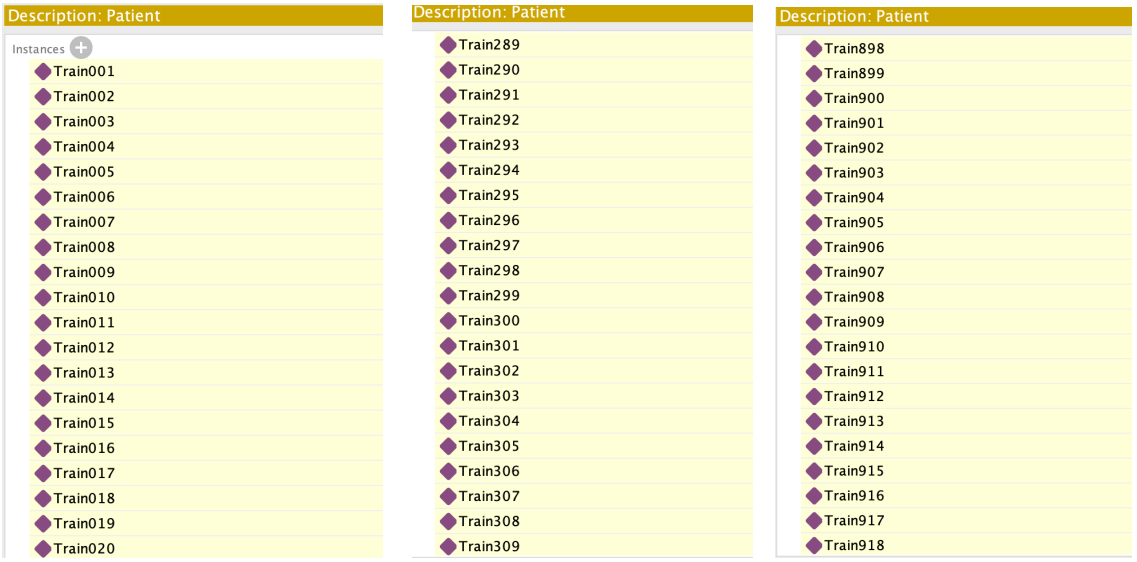


Figure 6.1: 918 Patient class inferences from the Training Dataset

Additional class memberships were inferred by the reasoner, particularly for risk factors. Since the range of the has_risk_factor property is Risk_Factor, instances such as family_psychiatric_diagnoses, loss_of_parents, and others were automatically classified as Risk_Factors. Notably, other mental disorders—including ADHD, Depression, GAD, ODD, and SAD—were also inferred to be risk factors. This inference is particularly valuable for exploring comorbidity, as it provides insights into the likelihood of developing an anxiety disorder when other mental health conditions are present.

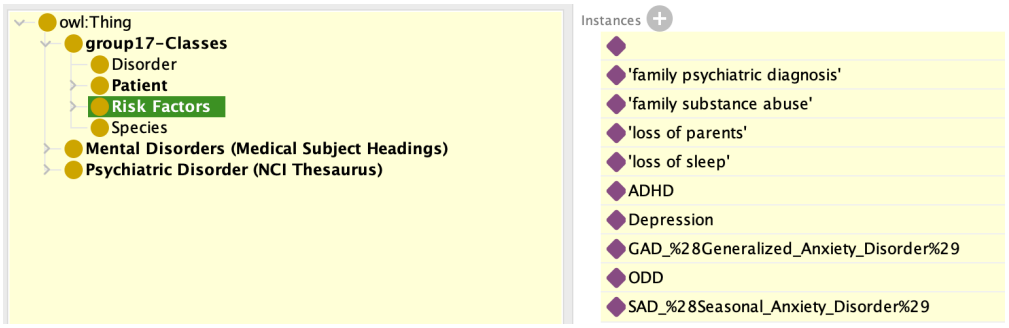


Figure 6.2: Risk Factors class inferences

7 Relevant SPARQL queries

Queries below provide a structured analysis of patient demographics, risk factors, and health conditions related to sleep and anxiety. By utilising the data converted from the datasets, they aim to find correlations between genetic markers, sleep patterns, and anxiety factors, offering insights that can support broader research into anxiety disorder risk and related conditions.

7.1 Query 1: Count of Patients

This query retrieves the count of male and female patients from the ‘testing’ dataset, categorised and ordered by the age group from low to high.

Query:

```
PREFIX anx: <http://AnxietyDisordersandRiskFactors.com#>

SELECT ?age
       (COUNT(?malePatient) AS ?maleCount)
       (COUNT(?femalePatient) AS ?femaleCount)
       (COUNT(?malePatient) + COUNT(?femalePatient) AS ?total)
WHERE {
  {
    ?malePatient anx:hasAge ?age .
    ?malePatient a anx:Male .
  }
  UNION
  {
    ?femalePatient anx:hasAge ?age .
    ?femalePatient a anx:Female .
  }
}
GROUP BY ?age
ORDER BY ?age
```

Output:

	age	maleCount	femaleCount	total
1	"2"^^xsd:integer	"53"^^xsd:integer	"41"^^xsd:integer	"94"^^xsd:integer
2	"3"^^xsd:integer	"43"^^xsd:integer	"30"^^xsd:integer	"73"^^xsd:integer
3	"4"^^xsd:integer	"36"^^xsd:integer	"38"^^xsd:integer	"74"^^xsd:integer
4	"5"^^xsd:integer	"32"^^xsd:integer	"29"^^xsd:integer	"61"^^xsd:integer
5	"6"^^xsd:integer	"2"^^xsd:integer	"3"^^xsd:integer	"5"^^xsd:integer

Table 7.1.1: Query 1 retrieved data

7.2 Query 2:

This query counts the number of patients for each risk factor from the ‘training’ and the ‘testing’ datasets, ordered by the number of patients from highest to lowest.

Query:

```
PREFIX anx: <http://AnxietyDisordersandRiskFactors.com#>

SELECT (STRAFTER(STR(?riskFactor), "#") AS ?riskFactorName) (COUNT(?patient) AS ?patientCount)
WHERE {
  ?patient anx:hasRiskFactor ?riskFactor .
}
GROUP BY ?riskFactor
ORDER BY DESC(?patientCount)
```

Output:

	riskFactorName	patientCount
1	"SAD_%28Seasonal_Anxiety_Disorder%29"	"276"^^xsd:integer
2	"GAD_%28Generalized_Anxiety_Disorder%29"	"272"^^xsd:integer
3	"loss_of_parents"	"155"^^xsd:integer
4	"family_psychiatric_diagnosis"	"102"^^xsd:integer
5	"family_substance_abuse"	"72"^^xsd:integer
6	"ODD"	"34"^^xsd:integer
7	"ADHD"	"22"^^xsd:integer
8	"Depression"	"17"^^xsd:integer

Table 7.2.1: Query 2 retrieved data

7.3 Query 3:

This query calculates the rounded average age of patients with varying numbers of sleep disturbances from the 'testing' dataset who are not in poverty, categorised and ordered by number of disturbances from low to high.

Query:

```
PREFIX anx: <http://AnxietyDisordersandRiskFactors.com#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?numOfDisturbances (ROUND(AVG(?age) * 100) / 100 AS ?avgAge)
WHERE {
  ?patient anx:has_sleep_disturbances ?numOfDisturbances;
           anx:hasAge ?age;
           anx:isInPoverty "false"^^xsd:boolean.
}
GROUP BY ?numOfDisturbances
ORDER BY ?numOfDisturbances
```

Output:

	numOfDisturbances	avgAge
1	"1"^^xsd:integer	"3.68"^^xsd:decimal
2	"2"^^xsd:integer	"3.26"^^xsd:decimal
3	"3"^^xsd:integer	"3.23"^^xsd:decimal
4	"4"^^xsd:integer	"3.49"^^xsd:decimal
5	"5"^^xsd:integer	"3.39"^^xsd:decimal
6	"6"^^xsd:integer	"2.86"^^xsd:decimal
7	"7"^^xsd:integer	"2.78"^^xsd:decimal
8	"8"^^xsd:integer	"4"^^xsd:decimal
9	"9"^^xsd:integer	"3.5"^^xsd:decimal
10	"10"^^xsd:integer	"5"^^xsd:decimal

Table 7.3.1: Query 3 retrieved data

7.3 Query 4:

This query computes the average hours to fall asleep and sleep resistance levels for patients from the 'training' dataset, categorised by Social and Generalised Anxiety Disorder.

Query:

```
PREFIX anx: <http://AnxietyDisordersandRiskFactors.com#>

SELECT (STRAFTER(STR(?riskFactor), "#") AS ?riskFactorName)
       (ROUND(AVG(?hoursToFallAsleep) * 100) / 100 AS ?avgHoursToFallAsleep)
       (ROUND(AVG(?sleepResistance) * 100) / 100 AS ?avgSleepResistance)
WHERE {
    ?patient anx:has_hours_to_fall_asleep ?hoursToFallAsleep ;
             anx:has_sleep_resistance ?sleepResistance ;
             anx:hasRiskFactor ?riskFactor .
}
GROUP BY ?riskFactor
```

Output:

	riskFactorName	avgHoursToFallAsl...	avgSleepResistance
1	"GAD_%28Generalized_Anxiety_Disorder%29"	"0.25"^^xsd:decimal	"29.21"^^xsd:decimal
2	"SAD_%28Seasonal_Anxiety_Disorder%29"	"0.34"^^xsd:decimal	"33.03"^^xsd:decimal

Table 7.4.1: Query 4 retrieved data

7.3 Query 5:

This query retrieves all genetic factors (genes or alleles) that are expressed in Homo sapiens and are identified as markers for anxiety disorders.

Query:

```
PREFIX anx: <http://AnxietyDisordersandRiskFactors.com#>

SELECT (STRAFTER(STR(?geneticFactor), "#") AS ?geneticFactorName)
WHERE {
  {
    ?geneticFactor anx:expressedIn anx:Homo_sapiens ;
                   anx:is_marker_via_orthology anx:anxiety_disorder .
  }
  UNION
  {
    ?geneticFactor anx:expressedIn anx:Homo_sapiens ;
                   anx:is_implicated_in anx:anxiety_disorder .
  }
}
```

Output:

	geneticFactorName
1	"ARC"
2	"BDNF"
3	"CREB1"
4	"DNMT3B"
5	"ELK1"
6	"GAD1"
7	"HDAC2"
8	"HTR1A"
9	"L1CAM"
10	"MC4R"
11	"NTRK2"
12	"POMC"
13	"CTNND2"
14	"DRD2"
15	"FNTB"
16	"GAD1"
17	"GDNF"
18	"HTR1A"
19	"MANEA"
20	"NPY"
21	"SLC6A4"

Table 7.5.1: Query 5 retrieved data

8 Data Science pipeline

This Jupyter notebook implements a data science pipeline to analyse and visualise the relationship between biological and environmental risk factors and the development of anxiety disorders. The pipeline utilises Pandas DataFrames for data manipulation and analysis. The notebook consists of the following steps:

- The ontology data from the `GROUP17-FinalOntology.ttl` file is parsed and converted into a Pandas DataFrame, allowing for easy querying and manipulation of the data.
- Various queries are run to extract insights related to sleep patterns, family histories, and genetic markers. The results are stored in Pandas DataFrames, which facilitate further analysis and statistical calculations.
- The notebook uses visualisation libraries such as Matplotlib and Seaborn to create plots that show the findings. These visualisations provide a clear understanding of the relationships between sleep disturbances, family psychiatric history, and genetic markers associated with anxiety disorders.

The notebook directly addresses the competency questions through targeted analysis:

Sleep Disturbances and Anxiety Disorders: The analysis investigates how patterns of sleep-disturbances, sleep-resistance, and time taken to fall asleep correlate with anxiety disorders. By visualising sleep-related data, the notebook shows the importance of sleep quality as an indicator of anxiety, illustrating the potential impact of sleep issues on mental health.

Family Histories and Anxiety Risks: The project explores hereditary factors by examining the correlation between family psychiatric histories including diagnoses and substance abuse with anxiety disorders. Data visualisations present these patterns, highlighting the risk familial mental health has on people.

Genetic Markers and Comorbidities: The analysis narrows down the genetic markers associated with anxiety disorders and their connections to PTSD and OCD. By visualising the overlap of genetic factors, the notebook provides insights into the genetic risks of anxiety and its potential links to other disorders.

Through these analyses, the pipeline answers the competency questions and provides insights for stakeholders in mental health.

9 Contributions and Justification

M[redacted] Fatih A[redacted]:

- Contribution to finding one of the external ontologies
- Revising the competency questions
- Contribution to creating, optimising, and handling errors in the ontology
- Creating the conceptualising graph of the ontology
- Creating SPARQL queries and writing the SPARQL Queries section

Maggie C[redacted]:

- Contributed to finding one of the external datasets and writing the descriptions and motivations for the external sources
- Wrote the description of integrated knowledge graph (Section 5)
- Made descriptions of Protégé (Pellet) inferences (Section 6)

Holly M[redacted]:

- Contributed to finding an external dataset and writing a description of an external source
- Wrote the goals, motivation, and users section
- Contributed to writing the competency questions
- Wrote the data visualisation pipeline section
- Created the data visualisations in the jupyter notebook and the README file

Kayra Ö[redacted]:

- Contribution to creating, populating, and handling errors in the ontology.
- Contribution to finding one of the external ontologies and its integration
- Conversion of external non-rdf datasets (3 datasets from 2 sources)
- Writing and implementation of Sections 3-4 (excluding 3.3 Formalization): Methodology, Conceptualisation, Object/Data Properties, Restrictions, Data Integration and Conversion

To optimise the writing process of this report, generative AI was used for finding spelling and grammar errors.

References

[1] S. L. Chellappa and D. Aeschbach, "Sleep and anxiety: From mechanisms to interventions," Sleep Medicine Reviews, vol. 61, no. 61, p. 101583, Feb. 2022, doi: https://doi.org/10.1016/j.smrv.2021.101583.

[2] American Psychiatric Association, Diagnostic and statistical manual of mental disorders, 5th ed. Arlington, VA: American Psychiatric Publishing, 2013.

[3] Martínez-Romero, M., Jonquet, C., O'Connor, M. J., Graybeal, J., Pazos, A., Musen, M. A. (2017). NCBO Ontology Recommender 2.0: An Enhanced Approach For Biomedical Ontology Recommendation. Journal of Biomedical Semantics, 8(21), 1-22. DOI:
https://bioportal.bioontology.org/ontologies/MESH/?p=summary

[4] Carpenter, K. (n.d.). Quantifying Risk for Anxiety Disorders in Preschool Children: A Machine Learning Approach. Harvard Dataverse.
https://dataverse.harvard.edu/citation?persistentId=doi:10.7910/DVN/N42LWG

[5] Alliance of Genome Resources Consortium. Genetics. 2024 May 7;227(1):iyae049. doi:
https://www.alliancegenome.org/disease/DOID:2030 PMID: 38552170.

[6] Edison Ong, Zuoshuang Xiang, Bin Zhao, Yue Liu, Yu Lin, Jie Zheng, Chris Mungall, Mélanie Courtot, Alan Ruttenberg, Yongqun He, Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration, Nucleic Acids Research, Volume 45, Issue D1, January 2017, Pages D347–D352, http://purl.obolibrary.org/obo/NCIT_C2878