A Declarative Semantics for SNOMED CT Expression Constraints

March 9, 2015

Contents

1 Axiomatic Data Types **2**
- 1.1 Atomic Data Types . . . 2
- 1.2 Composite Data Types . . . 3

2 The Substrate **3**
- 2.1 Substrate Components . . . 3
- 2.2 Substrate . . . 3
  - 2.2.1 Strict and Permissive Substrates . . . 6
  - 2.2.2 Strict Substrate . . . 6
  - 2.2.3 Permissive Substrate . . . 6

3 SCTIDS or Error Return **6**

4 Interpretation of Expression Constraints **7**
- 4.1 expressionConstraint . . . 7
  - 4.1.1 unrefinedExpressionConstraint . . . 7
  - 4.1.2 refinedExpressionConstraint . . . 8
  - 4.1.3 simpleExpressionConstraint . . . 8
  - 4.1.4 compoundExpressionConstraint . . . 9
  - 4.1.5 conjunctionExpressionConstraint . . . 9
  - 4.1.6 disjunctionExpressionConstraint . . . 10
  - 4.1.7 exclusionExpressionConstraint . . . 10
  - 4.1.8 subExpressionConstraint . . . 11
- 4.2 refinement . . . 11
  - 4.2.1 conjunctionGroup . . . 12
  - 4.2.2 disjunctionGroup . . . 12
  - 4.2.3 subRefinement . . . 13
- 4.3 attributeSet . . . 13
  - 4.3.1 conjunctionAttributeSet . . . 14
  - 4.3.2 disjunctionAttributeSet . . . 14
  - 4.3.3 subAttributeSet . . . 15

4.4 attributeGroup . . . 15
4.5 attribute . . . 15
4.5.1 expressionAttribute . . . 16
4.5.2 concreteAttribute . . . 16
4.6 AttributeSubject . . . 17
4.7 Attribute . . . 17
4.8 AttributeSet and AttributeGroup . . . 19
4.9 Compound attribute evaluation . . . 20
4.10 Group Cardinality . . . 20
4.11 Cardinality . . . 21

5 Substrate Interpretations 22
5.1 attributeName . . . 22
5.2 attributeExpressionConstraint . . . 23
5.3 concreteAttributeConstraint . . . 23
5.4 FocusConcept . . . 25
5.4.1 focusConcept . . . 25
5.4.2 memberOf . . . 25
5.5 ConceptReferences . . . 26
5.5.1 conceptId . . . 26
5.5.2 conceptReference . . . 26

6 Glue and Helper Functions 27
6.1 Types . . . 27
6.2 Result transformations . . . 27

7 Appendix 1 29

8 Appendix 2 30

1 Axiomatic Data Types

1.1 Atomic Data Types

This section identifies the atomic data types that are assumed for the rest of this specification, specifically:

- **SCTID** – a SNOMED CT identifier
- **TERM** – a fully specified name, preferred term or synonym for a SNOMED CT Concept
- **REAL** – a real number
- **STRING** – a string literal
- **GROUP** – a role group identifier
- $\mathbb{N}$ – a non-negative integer
- $\mathbb{Z}$ – an integer

$[SCTID, TERM, REAL, STRING, GROUP]$

We will also need to recognize some well known identifiers: the *is_a* attribute, the *zero_group* and *attribute_concept*, the parent of all attributes

$$
\begin{array}{|l}
is\_a : SCTID \\
zero\_group : GROUP \\
attribute\_concept : SCTID \\
refset\_concept : SCTID
\end{array}
$$

1.2 Composite Data Types

While we can't fully specify the behavior of the concrete data types portion of the specification at this point, it is still useful to spell out the anticipated behavior on an abstract level.

- **CONCRETEVALUE** – a string, integer or real literal
- **TARGET** – the target of a relationship that is either an *SCTID* or a *CONCRETEVALUE*

$$CONCRETEVALUE ::= string\langle\langle STRING\rangle\rangle \mid integer\langle\langle \mathbb{Z}\rangle\rangle \mid real\langle\langle REAL\rangle\rangle$$
$$TARGET ::= object\langle\langle SCTID\rangle\rangle \mid concrete\langle\langle CONCRETEVALUE\rangle\rangle$$

2 The Substrate

A substrate represents the context of an interpretation.

2.1 Substrate Components

Quad Relationships in the substrate are represented a 4 element tuples or "quads" which consist of a source, attribute, target and role group identifier. The *is_a* attribute may only appear in the zero group, and the target of an *is_a* attribute must be a *SCTID* (not a *CONCRETEVALUE*)

Quad

$$
\begin{array}{|l}
s : SCTID \\
a : SCTID \\
t : TARGET \\
g : GROUP \\
\hline
a = is\_a \Rightarrow (g = zero\_group \land object^{\sim} t \in SCTID)
\end{array}
$$

2.2 Substrate

A substrate consists of:

- **concepts** The set of *SCTID*s (concepts) that are considered valid in the context of the substrate. *References to any SCTID that is not a member of concepts MUST be treated as an error.*

- **relationships** A set of relationship quads (source, attribute, target, group)
- **parentsOf** A function from an SCTID to its asserted and inferred parents
- **equivalent_concepts** A function from an SCTID to the set of SCTID's that have been determined to be equivalent to it.
- **refsets** The reference sets within the context of the substrate whose members are members are concept identifiers (i.e. are in *concepts*). While not formally spelled out in this specification, it is assumed that the typical reference set function would be returning a subset of the *refsetId*/*referencedComponentId* tuples represented in one or more RF2 Refset Distribution tables.

The following functions can be computed from the basic set above

- **childrenOf** The inverse of the parentsOf function
- **descendants** The transitive closure of the childrenOf function
- **ancestors** The transitive closure of the parentsOf function
- **attributeIds** The descendantsOf the attribute_concept, including equivalents
- **refsetsIds** The descendants of the refset_concept, including equivalents

The formal definition of substrate follows, where c and r are given and the remainder are derived. The expressions below assert that:

1. All sources, attributes and SCTID targets of *relationships* are included in the substrate *concepts* list.
2. There is a *parentsOf* entry for every concept in the substrate *concepts* list.
3. Every sctid in the range of the *parentsOf* function is in the substrate *concepts* list.
4. Every *is_a* relationship entry is represented in the *parentsOf* function. (Note that the reverse isn't necessarily true).
5. There is an *equivalent_concepts* assertion for every substrate concept.
6. The *equivalent_concepts* function is reflexive (i.e. every concept is equivalent to itself)
7. If two concepts (c2 and c2) are equivalent, then they:
   - Have the same parents
   - Appear the subject, attribute and object of the same set of relationships
   - Appear in the domain of the same set of refsets
   - Both appear in the range of any refset that one appears in
8. Every refset is a substrate *concepts*
9. Every member of a refset is a substrate *concept*
10. *childrenOf* is the inverse of *parentsOf*, where any concept that isn't a parent has no children.
11. *descendants* is the transitive closure of the *childrenOf* function
12. *ancestors* is the transitive closure of the *parentsOf* function
13. No concept can be its own ancestor (or, by inference, descendant)
14. Every *attributeId* is a substrate *concept*
15. Every *refsetId* is a substrate *concept*

Substrate

$$
\begin{array}{l}
concepts : \mathbb{P}\, SCTID \\
relationships : \mathbb{P}\, Quad \\
parentsOf : SCTID \pfun \mathbb{P}\, SCTID \\
equivalent\_concepts : SCTID \pfun \mathbb{P}\, SCTID \\
refsets : SCTID \pfun \mathbb{P}\, SCTID \\
\\
childrenOf : SCTID \pfun \mathbb{P}\, SCTID \\
descendants : SCTID \pfun \mathbb{P}\, SCTID \\
ancestors : SCTID \pfun \mathbb{P}\, SCTID \\
attributeIds : \mathbb{P}\, SCTID \\
refsetIds : \mathbb{P}\, SCTID \\
\hline
\forall\, rel : relationships \bullet rel.s \in concepts \land rel.a \in concepts \land \\
\qquad (object^{\sim} rel.t \in SCTID \Rightarrow object^{\sim} rel.t \in concepts) \\
\\
\mathrm{dom}\, parentsOf = concepts \\
\bigcup(\mathrm{ran}\, parentsOf) \subseteq concepts \\
\\
\forall\, r : relationships \bullet r.a = is\_a \Rightarrow (object^{\sim} r.t) \in parentsOf\ r.a \\
\\
\mathrm{dom}\, equivalent\_concepts = concepts \\
\forall\, c : concepts \bullet c \in equivalent\_concepts\ c \\
\forall\, c1, c2 : concepts \mid c2 \in (equivalent\_concepts\ c1) \bullet \\
\qquad parentsOf\ c1 = parentsOf\ c2 \land \\
\qquad \{r : relationships \mid r.s = c1\} = \{r : relationships \mid r.s = c2\} \land \\
\qquad \{r : relationships \mid r.a = c1\} = \{r : relationships \mid r.a = c2\} \land \\
\qquad \{r : relationships \mid object^{\sim} r.t = c1\} = \{r : relationships \mid object^{\sim} r.t = c2\} \land \\
\qquad c1 \in \mathrm{dom}\, refsets \Leftrightarrow c2 \in \mathrm{dom}\, refsets \land \\
\qquad c1 \in \mathrm{dom}\, refsets \Rightarrow refsets\ c1 = refsets\ c2 \land \\
\qquad (\forall\, rsd : \mathrm{ran}\, refsets \bullet c1 \in rsd \Leftrightarrow c2 \in rsd) \\
\\
\mathrm{dom}\, refsets \subseteq concepts \\
\bigcup(\mathrm{ran}\, refsets) \subseteq concepts \\
\\
\mathrm{dom}\, childrenOf = concepts \\
\forall\, s, t : concepts \bullet t \in parentsOf\ s \Leftrightarrow s \in childrenOf\ t \\
\forall\, c : concepts \mid c \notin \bigcup(\mathrm{ran}\, childrenOf) \bullet childrenOf\ c = \emptyset \\
\\
\forall\, s : concepts \bullet \\
\qquad descendants\ s = childrenOf\ s \cup \bigcup\{t : childrenOf\ s \bullet descendants\ t\} \\
\forall\, t : concepts \bullet \\
\qquad ancestors\ t = parentsOf\ t \cup \bigcup\{s : parentsOf\ t \bullet ancestors\ s\} \\
\forall\, t : concepts \bullet t \notin ancestors\ t \\
\\
attributeIds \subseteq concepts \\
refsetIds \subseteq concepts
\end{array}
$$

2.2.1 Strict and Permissive Substrates

Implementations may choose to implement "strict" substrates, where additional rules apply or "permissive" substrates where rules are relaxed.

2.2.2 Strict Substrate

A **strict_substrate** is a substrate where:

- There is at least one *SCTID* that is not substrate concept
- Every *attributeId* must be a descendant of *attribute_concept*
- Every *refsetId* must be a descendant of *refset_concept*
- *relationship* attributes must be *attributeId*s
- *refset* domains must be *refsetId*s

strict_substrate

$$
\begin{array}{|l}
\hline
Substrate \\
\hline
concepts \subset SCTID \\
attributeIds = descendants\ attribute\_concept \\
\forall\, r : relationships \bullet r.a \in attributeIds \\
\\
refsetIds = descendants\ refset\_concept \\
\mathrm{dom}\ refsets \subseteq refsetIds \\
\hline
\end{array}
$$

2.2.3 Permissive Substrate

A permissive substrate is a substrate where every query will return some result – all *SCTID's* are considered valid.

This includes the following rules:

1. Every possible *SCTID* is a substrate concept, attribute and a valid refset
2. The refset function will return a (possibly empty) set of results for any refuted

permissive_substrate

$$
\begin{array}{|l}
\hline
Substrate \\
\hline
concepts = SCTID \land attributeIds = concepts \land refsetIds = concepts \\
\hline
\end{array}
$$

3 SCTIDS or Error Return

The result of applying a query against a substrate is either a (possibly empty) set of SCTID's or an *ERROR*. An *ERROR* occurs when:

- The interpretation of a conceptId is not a substrate *concept*
- The interpretation of a relationship attribute is not a substrate *attributeId*
- The interpretation of a reset is not a substrate *refsetId*

$$ERROR ::= unknownConceptReference \mid unknownAttribute \mid unknownRefsetId$$

$$Sctids\_or\_Error ::= ok\langle\langle \mathbb{P}\, SCTID \rangle\rangle \mid error\langle\langle ERROR \rangle\rangle$$

4 Interpretation of Expression Constraints

This section defines the interpretation of all language constructs that are interpreted in terms of other language constructs. Each interpretation that follows begins with a simplified version of the language construct in the specification. It then formally specifies the constructs that are used in the interpretation, followed by the interpretation itself. We start with the definition of *expressionConstraint*, which, once interpreted, returns either a set of SCTIDs or an error condition.

4.1 expressionConstraint

```
expressionConstraint = ws ( refinedExpressionConstraint / unrefinedExpressionConstraint )
ws
```

`expressionConstraint` takes either a `refinedExpressionConstraint` or `unrefinedExpressionConstraint` and returns its interpretation as either a set of SCTIDs or an error condition.

expressionConstraint ::=
expcons_refined⟨⟨*refinedExpressionConstraint*⟩⟩ |
expcons_unrefined⟨⟨*unrefinedExpressionConstraint*⟩⟩

---

i_expressionConstraint :
Substrate → *expressionConstraint* → *Sctids_or_Error*

∀ *ss* : *Substrate*; *ec* : *expressionConstraint* • *i_expressionConstraint ss ec* =
if *ec* ∈ ran *expcons_refined*
then *i_refinedExpressionConstraint ss* (*expcons_refined*~ *ec*)
else *i_unrefinedExpressionConstraint ss* (*expcons_unrefined*~ *ec*)

4.1.1 unrefinedExpressionConstraint

The interpretation of an `unrefinedExpressionConstraint` is either the interpretation of a `compoundExpressionConstraint` or a `simpleExpressionConstraint`

```
unrefinedExpressionConstraint = compoundExpressionConstraint / simpleExpressionCon-
straint
```

unrefinedExpressionConstraint ::=
unrefined_compound⟨⟨*compoundExpressionConstraint*⟩⟩ |
unrefined_simple⟨⟨*simpleExpressionConstraint*⟩⟩

$$
\begin{array}{l}
i\_unrefinedExpressionConstraint : \\
\qquad Substrate \rightarrow unrefinedExpressionConstraint \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ uec : unrefinedExpressionConstraint \bullet \\
i\_unrefinedExpressionConstraint\ ss\ uec = \\
\textbf{if}\ ucec \in \operatorname{ran} unrefined\_compound \\
\quad \textbf{then}\ i\_compoundExpressionConstraint\ ss\ (unrefined\_compound^{\sim} uec) \\
\textbf{else}\ i\_simpleExpressionConstraint\ ss\ (unrefined\_simple^{\sim} uec)
\end{array}
$$

4.1.2 refinedExpressionConstraint

```
refinedExpressionConstraint = unrefinedExpressionConstraint ws ":" ws refinement / "(" ws
refinedExpressionConstraint ws ")"
```

The interpretation of `refinedExpressionConstraint` is the intersection of the interpretation of the `unrefinedExpressionConstraint` and the `refinement`, both of which return a set of SCTID's or an error. The second production defines `refinedExpressionConstraint` in terms of itself and has no impact on the results.

$$
refinedExpressionConstraint == unrefinedExpressionConstraint \times refinement
$$

$$
\begin{array}{l}
i\_refinedExpressionConstraint : \\
\qquad Substrate \rightarrow refinedExpressionConstraint \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ rec : refinedExpressionConstraint \bullet \\
i\_refinedExpressionConstraint\ ss\ rec = \\
\quad intersect\ (i\_unrefinedExpressionConstraint\ ss\ (first\ rec))(i\_refinement\ ss\ (second\ rec))
\end{array}
$$

4.1.3 simpleExpressionConstraint

The interpretation of `simpleExpressionConstraint` is the application of an optional constraint operator to the interpretation of `focusConcept`, which returns a set of SCTID's or an error. The interpretation of an error is the error.

```
simpleExpressionConstraint = [constraintOperator ws] focusConcept
```

$$
simpleExpressionConstraint == constraintOperator[0\,..\,1] \times focusConcept
$$

$$
\begin{array}{|l}
\hline
i\_simpleExpressionConstraint : \\
\qquad Substrate \rightarrow simpleExpressionConstraint \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ sec : simpleExpressionConstraint \bullet \\
i\_simpleExpressionConstraint\ ss\ sec = \\
\qquad i\_constraintOperator\ ss\ (first\ sec)\ (i\_focusConcept\ ss\ (second\ sec)) \\
\hline
\end{array}
$$

4.1.4 compoundExpressionConstraint

The interpretation of a *compoundExpressionConstraint* is the interpretation of its corresponding component.

```
compoundExpressionConstraint = conjunctionExpressionConstraint / disjunctionExpression-
Constraint / exclusionExpressionConstraint / "(" ws compoundExpressionConstraint ws ")"
```

$$
\begin{array}{l}
compoundExpressionConstraint ::= \\
\qquad compound\_conj\langle\langle conjunctionExpressionConstraint\rangle\rangle \mid \\
\qquad compound\_disj\langle\langle disjunctionExpressionConstraint\rangle\rangle \mid \\
\qquad compound\_excl\langle\langle exclusionExpressionConstraint\rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_compoundExpressionConstraint : \\
\qquad Substrate \rightarrow compoundExpressionConstraint \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ cec : compoundExpressionConstraint \bullet \\
i\_compoundExpressionConstraint\ ss\ cec = \\
\textbf{if}\ cec \in \mathrm{ran}\ compound\_conj \\
\quad \textbf{then}\ i\_conjunctionExpressionConstraint\ ss\ (compound\_conj^{\sim}\ cec) \\
\textbf{else if}\ cec \in \mathrm{ran}\ compound\_disj \\
\quad \textbf{then}\ i\_disjunctionExpressionConstraint\ ss\ (compound\_disj^{\sim}\ cec) \\
\textbf{else}\ i\_exclusionExpressionConstraint\ ss\ (compound\_excl^{\sim}\ cec) \\
\hline
\end{array}
$$

The signature below is used because the definition of `compountExpressionConstraint` is recursive

$$
\begin{array}{|l}
\hline
i\_compoundExpressionConstraint' : \\
\qquad Substrate \rightarrow compoundExpressionConstraint \rightarrow Sctids\_or\_Error \\
\hline
\end{array}
$$

4.1.5 conjunctionExpressionConstraint

`conjunctionExpressionConstraint` is interpreted the conjunction (intersection) of the interpretation of two or more `subExpressionConstraints`/ The `conjunction` aspect is ignored because there is no other choice

conjunctionExpressionConstraint = subExpressionConstraint 1*(ws conjunction ws subExpressionConstraint)

$$conjunctionExpressionConstraint == subExpressionConstraint \times \text{seq}_1(subExpressionConstraint)$$

Apply the intersection operator to the interpretation of each subExpressionConstraint

$$\begin{array}{|l} i\_conjunctionExpressionConstraint : \\ \quad Substrate \rightarrow conjunctionExpressionConstraint \rightarrow Sctids\_or\_Error \\ \hline \forall\, ss : Substrate;\ cecr : conjunctionExpressionConstraint \bullet \\ i\_conjunctionExpressionConstraint\ ss\ cecr = \\ \quad applyToSequence\ ss\ i\_subExpressionConstraint\ intersect\ cecr \end{array}$$

4.1.6 disjunctionExpressionConstraint

`disjunctionExpressionConstraint` is interpreted the disjunction (union) of the interpretation of two or more `subExpressionConstraints`. The `disjunction` element is ignored because there is no other choice.

disjunctionExpressionConstraint = subExpressionConstraint 1*(ws disjunction ws subExpressionConstraint)

$$disjunctionExpressionConstraint == subExpressionConstraint \times \text{seq}_1(subExpressionConstraint)$$

Apply the union operator to the interpretation of each subExpressionConstraint

$$\begin{array}{|l} i\_disjunctionExpressionConstraint : \\ \quad Substrate \rightarrow disjunctionExpressionConstraint \rightarrow Sctids\_or\_Error \\ \hline \forall\, ss : Substrate;\ decr : disjunctionExpressionConstraint \bullet \\ i\_disjunctionExpressionConstraint\ ss\ decr = \\ \quad applyToSequence\ ss\ i\_subExpressionConstraint\ union\ decr \end{array}$$

4.1.7 exclusionExpressionConstraint

The interpretation `exclusionExpressionConstraint` removes the interpretation of the second `exclusionExpressionConstraint` from the interpretation of the first. Errors are propagated.

exclusionExpressionConstraint = subExpressionConstraint ws exclusion ws subExpression-Constraint

$$exclusionExpressionConstraint == subExpressionConstraint \times subExpressionConstraint$$

$$\begin{array}{|l}
\hline
i_exclusionExpressionConstraint : \\
\quad Substrate \rightarrow exclusionExpressionConstraint \rightarrow Sctids_or_Error \\
\hline
\forall\, ss : Substrate;\ ecr : exclusionExpressionConstraint \bullet \\
i_exclusionExpressionConstraint\ ss\ ecr = \\
\quad minus\ (i_subExpressionConstraint\ ss\ (first\ ecr))(i_subExpressionConstraint\ ss\ (second\ ecr)) \\
\hline
\end{array}$$

4.1.8 subExpressionConstraint

subExpressionConstraint is interpreted as the interpretation of either a *simpleExpressionConstraint* or a *compoundExpressionConstraint*

subExpressionConstraint = simpleExpressionConstraint / "(" ws (compoundExpressionConstraint / refinedExpressionConstraint) ws ")"

$$\begin{array}{l}
subExpressionConstraint ::= \\
\quad subExpr_simple\langle\langle simpleExpressionConstraint\rangle\rangle \mid \\
\quad subExpr_compound\langle\langle compoundExpressionConstraint\rangle\rangle \mid \\
\quad subExpr_refined\langle\langle refinedExpressionConstraint\rangle\rangle
\end{array}$$

$$\begin{array}{|l}
\hline
i_subExpressionConstraint : \\
\quad Substrate \rightarrow subExpressionConstraint \rightarrow Sctids_or_Error \\
\hline
\forall\, ss : Substrate;\ sec : subExpressionConstraint \bullet \\
i_subExpressionConstraint\ ss\ sec = \\
\textbf{if}\ sec \in \mathrm{ran}\ subExpr_simple \\
\quad \textbf{then}\ i_simpleExpressionConstraint\ ss\ (subExpr_simple^{\sim} sec) \\
\textbf{else if}\ sec \in \mathrm{ran}\ subExpr_compound \\
\quad \textbf{then}\ i_compoundExpressionConstraint'\ ss\ (subExpr_compound^{\sim} sec) \\
\textbf{else}\ i_refinedExpressionConstraint\ ss\ (subExpr_refined^{\sim} sec) \\
\hline
\end{array}$$

4.2 refinement

The interpretation of `refinement` is the interpretation of the `subRefinement`, `conjunctionGroup` or `disjunctionGroup`

refinement = subRefinement / conjunctionGroup / disjunctionGroup

$$
\begin{array}{l}
refinement ::= \\
\quad refine\_subrefine\langle\langle subrefinement\rangle\rangle \mid \\
\quad refine\_conjg\langle\langle conjunctionGroup\rangle\rangle \mid \\
\quad refine\_disjg\langle\langle disjunctionGroup\rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_refinement : Substrate \to refinement \to Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ rfnment : refinement \bullet i\_refinement = \\
\textbf{if } refnment \in \operatorname{ran} refine\_subrefine \\
\quad \textbf{then } i\_subrefinement\ ss\ (refine\_subrefine^{\sim} refnment) \\
\textbf{else if } refnment \in \operatorname{ran} refine\_conjg \\
\quad \textbf{then } i\_conjunctionGroup\ ss\ (refine\_conjg^{\sim} refnment) \\
\textbf{else } i\_disjunctionGroup\ ss\ (refine\_disjg^{\sim} refnment) \\
\hline
\end{array}
$$

4.2.1 conjunctionGroup

conjunctionGroup = subRefinement 1*(conjunction subRefinement)

$$
\begin{array}{l}
conjunctionGroup == \\
\qquad subRefinement \times \operatorname{seq}_1(subRefinement)
\end{array}
$$

Apply the intersect operator to the interpretation of each subRefinement

$$
\begin{array}{|l}
\hline
i\_conjunctionGroup : \\
\quad Substrate \to conjunctionGroup \to Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ conjg : conjunctionGroup \bullet \\
i\_conjunctionGroup\ ss\ conjg = \\
\quad applyToSequence\ ss\ i\_subRefinement\ intersect\ conjg \\
\hline
\end{array}
$$

4.2.2 disjunctionGroup

disjunctionGroup = subRefinement 1*(disjunction subRefinement)

$$
\begin{array}{l}
disjunctionGroup == \\
\qquad subRefinement \times \operatorname{seq}_1(subRefinement)
\end{array}
$$

Apply the union operator to the interpretation of each subRefinement

$$
\begin{array}{|l}
\hline
i\_disjunctionGroup : \\
\quad Substrate \rightarrow disjunctionGroup \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ disjg : disjunctionGroup \bullet \\
i\_disjunctionGroup\ ss\ disjg = \\
\quad applyToSequence\ ss\ i\_subRefinement\ union\ disjg \\
\hline
\end{array}
$$

4.2.3 subRefinement

The interpretation of a `subRefinement` is the interpretation of the corresponding `attributeSet`, `attributeGroup` or `refinement`.

```
subRefinement = attributeSet / attributeGroup / "(" ws refinement ws ")?
```

$$
\begin{array}{l}
subRefinement ::= \\
\quad subrefine\_attset\langle\langle attributeSet\rangle\rangle \mid \\
\quad subrefine\_attgroup\langle\langle attributeGroup\rangle\rangle \mid \\
\quad subrefine\_refinement\langle\langle refinement\rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_subRefinement : \\
\quad Substrate \rightarrow subRefinement \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ subrefine : subRefinement \bullet \\
i\_subRefinement\ ss\ subrefine = \\
\textbf{if}\ subrefine \in \mathrm{ran}\ subrefine\_attset \\
\quad \textbf{then}\ i\_attributeSet\ ss\ (subrefine\_attset^{\sim}\ subrefine) \\
\textbf{else if}\ subrefine \in \mathrm{ran}\ subrefine\_attgroup \\
\quad \textbf{then}\ i\_attributeGroup\ ss\ (subrefine\_attgroup^{\sim}\ subrefine) \\
\textbf{else}\ i\_refinement\ ss\ (subrefine\_refinement^{\sim}\ subrefine) \\
\hline
\end{array}
$$

4.3 attributeSet

```
attributeSet = subAttributeSet / conjunctionAttributeSet / disjunctionAttributeSet
```

$$
\begin{array}{l}
attributeSet ::= \\
\quad attset\_subattset\langle\langle subAttributeSet\rangle\rangle \mid \\
\quad attset\_conjattset\langle\langle conjunctionAttributeSet\rangle\rangle \mid \\
\quad attset\_disjattset\langle\langle disjunctionAttributeSet\rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_attributeSet : \\
\qquad Substrate \rightarrow attributeSet \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ attset : attributeSet \bullet \\
i\_attributeSet\ ss\ attset = \\
\textbf{if}\ attset \in \operatorname{ran} attset\_subattset \\
\quad \textbf{then}\ i\_subAttributeSet\ ss\ (attset\_subattset^{\sim} attset) \\
\textbf{else if}\ attset \in \operatorname{ran} attset\_conjattset \\
\quad \textbf{then}\ i\_conjunctionAttributeSet\ ss\ (attset\_conjattset^{\sim} attset) \\
\textbf{else}\ i\_disjunctionAttributeSet\ ss\ (attset\_disjattset^{\sim} attset) \\
\hline
\end{array}
$$

4.3.1 conjunctionAttributeSet

conjunctionAttributeSet = subAttributeSet 1*(conjunction subAttributeSet)

$$
conjunctionAttributeSet == \\
\qquad subAttributeSet \times \operatorname{seq}_1(subAttributeSet)
$$

Apply the intersect operator to the interpretation of each subAttributeSet

$$
\begin{array}{|l}
\hline
i\_conjunctionAttributeSet : \\
\quad Substrate \rightarrow conjunctionAttributeSet \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ conjaset : conjunctionAttributeSet \bullet \\
i\_conjunctionAttributeSet\ ss\ conjaset = \\
\quad applyToSequence\ ss\ i\_subAttributeSet\ intersect\ conjaset \\
\hline
\end{array}
$$

4.3.2 disjunctionAttributeSet

disjunctionAttributeSet = subAttributeSet 1*(disjunction subAttributeSet)

$$
disjunctionAttributeSet == \\
\qquad subAttributeSet \times \operatorname{seq}_1(subAttributeSet)
$$

Apply the union operator to the interpretation of each subAttributeSet

$$
\begin{array}{|l}
\hline
i\_disjunctionAttributeSet : \\
\quad Substrate \rightarrow disjunctionAttributeSet \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ disjaset : disjunctionAttributeSet \bullet \\
i\_disjunctionAttributeSet\ ss\ disjaset = \\
\quad applyToSequence\ ss\ i\_subAttributeSet\ union\ disjaset \\
\hline
\end{array}
$$

4.3.3 subAttributeSet

```
subAttributeSet = attribute / "(" ws attributeSet ws ")"
```

$$
\begin{array}{l}
subAttributeSet ::= \\
\quad subaset\_attribute\langle\langle attribute \rangle\rangle \mid \\
\quad subaset\_attset\langle\langle attributeSet \rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_subAttributeSet : \\
\qquad Substrate \rightarrow subAttributeSet \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ subaset : subAttributeSet \bullet \\
i\_subAttributeSet\ ss\ subaset = \\
\textbf{if}\ subaset \in \operatorname{ran} subaset\_attribute \\
\quad \textbf{then}\ i\_attribute\ ss\ (subaset\_attribute^{\sim} subaset) \\
\textbf{else}\ i\_attributeSet\ ss\ (subaset\_attset^{\sim} subaset) \\
\hline
\end{array}
$$

4.4 attributeGroup

```
attributeGroup = [cardinality ws] "{" ws attributeSet ws "}"
```

4.5 attribute

```
attribute = [cardinality ws] [reverseFlag ws] ws attributeName ws (concreteComparisonOper-
ator ws concreteValue / expressionComparisonOperator ws expressionConstraintValue )
cardinality = "[" nonNegativeIntegerValue to (nonNegativeIntegerValue / many) "]"
```

$$
\begin{array}{l}
attribute ::= \\
\quad attrib\_conc\langle\langle concreteAttribute \rangle\rangle \mid \\
\quad attrib\_expr\langle\langle expressionAttribute \rangle\rangle
\end{array}
$$

$$
\begin{array}{|l}
\hline
i\_attribute : \\
\qquad Substrate \rightarrow attribute \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ att : attribute \bullet \\
i\_attribute\ ss\ att = \\
\textbf{if}\ att \in \operatorname{ran} attrib\_conc \\
\quad \textbf{then}\ i\_concreteAttribute\ ss\ (attrib\_conc^{\sim} att) \\
\textbf{else}\ i\_expressionAttribute\ ss\ (attrib\_expr^{\sim} att) \\
\hline
\end{array}
$$

For the sake of simplicity, we separate out the components of the concrete and expression constraints.

$unlimitedNat ::= num\langle\langle\mathbb{N}\rangle\rangle \mid many$
$cardinality == \mathbb{N} \times unlimitedNat$
$[reverseFlag]$

4.5.1 expressionAttribute

```
expressionComparisonOperator = "=" / "!=" / "<>"
```

$expressionComparisonOperator ::= eco\_eq \mid eco\_neq$

expressionAttribute

$card : cardinality[0 \mathinner{..} 1]$
$reverse : reverseFlag[0 \mathinner{..} 1]$
$name : attributeName$
$operator : expressionComparisonOperator$
$value : expressionConstraintValue$

4.5.2 concreteAttribute

```
concreteComparisonOperator = "=" / "!=" / "<>" / "<=" / "<" / ">=" / ">"
concreteValue = QM stringValue QM / "#" numericValue
stringValue = 1*(anyNonEscapedChar / escapedChar)
numericValue = decimalValue / integerValue
```

$concreteComparisonOperator ::=$
$\quad cco\_eq \mid cco\_neq \mid cco\_leq \mid ccl\_lt \mid cco\_geq \mid cco\_gt$
$concreteValue ::= stringValue \mid integerValue \mid decimalValue$

concreteAttribute

$card : cardinality[0 \mathinner{..} 1]$
$name : attributeName$
$operator : concreteComparisonOperator$
$value : concreteValue$

The interpretation of a `concreteAttribute` selects the set of quads in the substrate that have an attribute in the set of attributes determined by the interpretation of `attributeName` having *CONCRETEVALUE* targets that meet the supplied comparison rules.

$$
\begin{array}{|l}
\hline
i\_concreteAttribute : \\
\quad Substrate \rightarrow concreteAttribute \rightarrow Quads\_or\_Error \\
\hline
\forall\, ss : Substrate;\ ca : concreteAttribute \bullet \\
i\_concreteAttribute\ ss\ ca = \\
(\textbf{let}\ attids == i\_attributeName\ ss\ ca.name \bullet \\
\quad i\_concreteAttributeConstraint\ ss\ attids\ ca.operator\ ca.value) \\
\hline
\end{array}
$$

The interpretation of an attribute. `attributeOperator` and `attributeName` determines the set of possible attributes in the substrate relationship table. `reverseFlag` and `expressionConstraintValue` determine the set of candidate targets (if `reverseFlag` is absent) or `source_direction` (if `reverseFlag` is present).

`cardinality` determines the minimum and maximum matches. In all cases, only a subset of the sources (targets if `reverseFlag` is present) in the substrate relationship table will be returned in the interpretation.

4.6 AttributeSubject

i_attributeSubject interprets the constraintOperator, if any in the context of the attribute name and interprets it as a set of *ATTRIBUTE*s

$$attributeOperator == \{descendantOrSelfOf, descendantOf\}$$
$$attributeSubject == attributeOperator[0\,..\,1] \times attributeName$$

$$
\begin{array}{|l}
\hline
i\_attributeSubject : Substrate \pfun attributeSubject \pfun Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ as : attributeSubject \bullet i\_attributeSubject\ ss\ as = \\
\quad i\_constraintOperator\ ss\ (first\ as)\ (i\_attributeName\ ss\ (second\ as)) \\
\hline
\end{array}
$$

4.7 Attribute

`attribute` consists of an optional `cardinality` and an `attributeConstraint`.

$$unlimitedNat ::= num\langle\langle \mathbb{N} \rangle\rangle \mid many$$
$$cardinality == \mathbb{N} \times unlimitedNat$$

$$
\begin{array}{|l}
attribute \\
\hline
card : cardinality[0\,..\,1] \\
attw : attributeConstraint \\
\hline
\end{array}
$$

attributeConstraint is either an attribute expression constraint or a concrete value constraint.

$$attributeConstraint ::= aec\langle\langle attributeExpressionConstraint \rangle\rangle \mid acvc\langle\langle attributeConcreteValueConstraint \rangle\rangle$$

`attributeExpressionEonstraint` is a combination of a source constraint (target if reverse flag is true) and an expression constraint value whose interpretation yields a set of SCTID's that filter the target (source if reverse flag).

$[reverseFlag]$

$$
\begin{array}{|l}
\hline
attributeExpressionConstraint \\
a : attributeSubject \\
rf : reverseFlag[0 \,..\, 1] \\
cs : expressionConstraintValue \\
\hline
\end{array}
$$

A concrete value constraint is the combination of a source constraint and a comparison operator/concrete value whose interpretation yields a set of source ids. The intersection of the source and concrete value interpretation is the interpretation of `attributeConcreteValueConstraint`

$comparisonOperator ::= eq \mid neq \mid gt \mid ge \mid lt \mid le$

$$
\begin{array}{|l}
\hline
attributeConcreteValueConstraint \\
a : attributeSubject \\
op : comparisonOperator \\
v : concreteValue \\
\hline
\end{array}
$$

The interpretation of an attribute is the interpretation of the cardinality applied against the underlying attribute constraint, producing a set of SCTID's or an error condition

The interpretation of an attribute constraint is the interpretation of either the attribute expression constraint or the concrete expression constraint that produces a set of Quads or an error condition.

The actual interpretations if attribute expression and concrete value are in Section 5

$$
\begin{array}{l}
i\_attribute : \\
\quad Substrate \pfun attribute \pfun Sctids\_or\_Error \\
i\_attributeConstraint : \\
\quad Substrate \pfun attributeConstraint \pfun Quads\_or\_Error \\
i\_attributeExpressionConstraint : \\
\quad Substrate \pfun attributeExpressionConstraint \pfun Quads\_or\_Error \\
i\_attributeConcreteValueConstraint : \\
\quad Substrate \pfun attributeConcreteValueConstraint \pfun Quads\_or\_Error \\
\hline
\forall\, ss : Substrate;\ a : attribute \bullet \\
\quad i\_attribute\ ss\ a = i\_cardinality\ a.card\ (i\_attributeConstraint\ ss\ a.attw) \\
\forall\, ss : Substrate;\ aw : attributeConstraint \bullet i\_attributeConstraint\ ss\ aw = \\
\quad \textbf{if}\ aec^{\sim} aw \in attributeExpressionConstraint \\
\quad\quad \textbf{then}\ i\_attributeExpressionConstraint\ ss\ (aec^{\sim} aw) \\
\quad \textbf{else}\ i\_attributeConcreteValueConstraint\ ss\ (acvc^{\sim} aw) \\
\forall\, ss : Substrate;\ aec : attributeExpressionConstraint;\ eca : expressionConstraintArgs \mid \\
\quad eca.atts = i\_attributeSubject\ ss\ aec.a \land \\
\quad eca.rf = aec.rf \land \\
\quad eca.subjOrTarg = i\_expressionConstraintValue\ ss\ aec.cs \bullet \\
\quad i\_attributeExpressionConstraint\ ss\ aec = i\_attributeExpression\ ss\ eca \\
\forall\, ss : Substrate;\ awc : attributeConcreteValueConstraint;\ aca : concreteConstraintArgs \mid \\
\quad aca.atts = i\_attributeSubject\ ss\ awc.a \land \\
\quad aca.op = awc.op \land \\
\quad aca.t = awc.v \bullet \\
\quad i\_attributeConcreteValueConstraint\ ss\ awc = i\_concreteAttributeConstraint\ ss\ aca
\end{array}
$$

4.8 AttributeSet and AttributeGroup

An `attributeGroup` is an `attributeSet`. An `attributeSet` consists of a sequence of one or more attribute constraints joined by `binaryOperator`s.

$$
\begin{array}{l}
binaryOperator ::= conjunction \mid disjunction \mid exclusion \\
attributeGroup == attributeSet \\
attributeSet == attribute \times \mathrm{seq}(binaryOperator \times attribute)
\end{array}
$$

$$
\begin{array}{l}
i\_attributeGroup : Substrate \pfun attributeGroup \pfun Sctids\_or\_Error \\
i\_attributeSet : Substrate \pfun attributeSet \pfun Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ ag : attributeGroup \bullet i\_attributeGroup\ ss\ ag = \\
\quad i\_attributeSet\ ss\ ag \\
\forall\, ss : Substrate;\ a : attributeSet \bullet i\_attributeSet\ ss\ a = \\
\quad idgroups\_to\_sctids\ (evalCmpndAtt\ ss\ (i\_groupedAttribute\ ss\ (first\ a))\ (second\ a))
\end{array}
$$

4.9 Compound attribute evaluation

The left-to-right evaluation of `attributeSet`. The interpretation takes a *lhs* as a function from *SCTID* to *GROUP* and a *rhs* which is sequence of operator/attribute tuples and recursively interprets the *lhs* and interpretation of the head of the *rhs*.

$$
\begin{array}{l}
\textit{evalCmpndAtt} : \\
\qquad \textit{Substrate} \to \textit{IDGroups} \to \mathrm{seq}(\textit{binaryOperator} \times \textit{attribute}) \to \textit{IDGroups} \\
\\
\textit{gintersect}, \textit{gunion}, \textit{gminus}, \textit{gfirstError} : \\
\qquad \textit{IDGroups} \to \textit{IDGroups} \to \textit{IDGroups} \\
\hline
\forall\, ss : \textit{Substrate};\ lhs : \textit{IDGroups};\ rhs : \mathrm{seq}(\textit{binaryOperator} \times \textit{attribute}) \bullet \\
\textit{evalCmpndAtt}\ ss\ lhs\ rhs = \\
\textbf{if}\ rhs = \langle\rangle \\
\quad \textbf{then}\ lhs \\
\textbf{else if}\ \textit{first}\,(\textit{head}\ rhs) = \textit{conjunction} \\
\textbf{then}\ \textit{evalCmpndAtt}\ ss\,(\textit{gintersect}\ lhs(\textit{i\_groupedAttribute}\ ss\,(\textit{second}\,(\textit{head}\ rhs))))\,(\textit{tail}\ rhs) \\
\textbf{else if}\ \textit{first}\,(\textit{head}\ rhs) = \textit{disjunction} \\
\textbf{then}\ \textit{evalCmpndAtt}\ ss\,(\textit{gunion}\ lhs(\textit{i\_groupedAttribute}\ ss\,(\textit{second}\,(\textit{head}\ rhs))))\,(\textit{tail}\ rhs) \\
\textbf{else}\ \textit{evalCmpndAtt}\ ss\,(\textit{gminus}\ lhs(\textit{i\_groupedAttribute}\ ss\,(\textit{second}\,(\textit{head}\ rhs))))\,(\textit{tail}\ rhs) \\
\\
\forall\, a, b, r : \textit{IDGroups} \mid \\
\qquad r = \textbf{if}\ \textit{gerror}^{\sim} a \in \textit{ERROR} \lor \textit{gerror}^{\sim} b \in \textit{ERROR}\ \textbf{then}\ \textit{gfirstError}\ a\ b \\
\qquad \textbf{else}\ \textit{id\_groups}\,(\textit{id\_groups}^{\sim} a \cap \textit{id\_groups}^{\sim} b) \bullet \\
\qquad \textit{gintersect}\ a\ b = r \\
\\
\forall\, a, b, r : \textit{IDGroups} \mid \\
\qquad r = \textbf{if}\ \textit{gerror}^{\sim} a \in \textit{ERROR} \lor \textit{gerror}^{\sim} b \in \textit{ERROR}\ \textbf{then}\ \textit{gfirstError}\ a\ b \\
\qquad \textbf{else}\ \textit{id\_groups}\,(\textit{id\_groups}^{\sim} a \cup \textit{id\_groups}^{\sim} b) \bullet \\
\qquad \textit{gunion}\ a\ b = r \\
\\
\forall\, a, b, r : \textit{IDGroups} \mid \\
\qquad r = \textbf{if}\ \textit{gerror}^{\sim} a \in \textit{ERROR} \lor \textit{gerror}^{\sim} b \in \textit{ERROR}\ \textbf{then}\ \textit{gfirstError}\ a\ b \\
\qquad \textbf{else}\ \textit{id\_groups}\,(\textit{id\_groups}^{\sim} a \setminus \textit{id\_groups}^{\sim} b) \bullet \\
\qquad \textit{gminus}\ a\ b = r
\end{array}
$$

$$
\begin{array}{l}
\textit{i\_groupedAttribute} : \\
\qquad \textit{Substrate} \to \textit{attribute} \to \textit{IDGroups} \\
\hline
\forall\, ss : \textit{Substrate};\ a : \textit{attribute} \bullet \\
\qquad \textit{i\_groupedAttribute}\ ss\ a = \textit{i\_groupCardinality}\,(\textit{i\_attributeConstraint}\ ss\ a.attw)\ a.card
\end{array}
$$

4.10 Group Cardinality

The interpretation of cardinality within a group impose additional constraints:

- $[0\,..\,n]$ – the set of all substrate concept codes that have at least one group (entry) in the substrate relationships and, at most n matching entries in the same group
- $[0\,..\,0]$ – the set of all substrate concept codes that have at least one group (entry) in the substrate relationships and *no* matching entries
- $[1..*]$ – (default) at least one matching entry in the substrate relationships
- $[m_1\,..\,n_1]op[m_2\,..\,n_2]...$ – set of substrate concept codes where there exists at least one group where all conditions are simultaneously true

The interpretation of a grouped cardinality is a function from a set of SCTID's to the groups in which they were qualified.

The algorithm below partitions the input set of Quads by group and validates the cardinality on a per-group basis. Groups that pass are returned

TODO: This assumes that q.t is always type object. It doesn't say what to do if it is concrete **TODO:** the *quads_to_idgroups* function seems to express what is described below more simply

$$
\begin{array}{l}
\hline
i\_groupCardinality : \\
\qquad Quads\_or\_Error \rightarrow cardinality[0\,..\,1] \rightarrow IDGroups \\
\hline
\forall\, quads : Quads\_or\_Error;\ oc : cardinality[0\,..\,1];\ uniqueGroups : \mathbb{P}\, GROUP; \\
\qquad quadsByGroup : GROUP \nrightarrow \mathbb{P}\, Quad \mid \\
\quad uniqueGroups = \{q : quads\_for\ quads \bullet q.g\} \land \\
\quad quadsByGroup = \{g : uniqueGroups;\ q : \mathbb{P}\, Quad \mid \\
\qquad q = \{e : quads\_for\ quads \mid e.g = g\} \bullet g \mapsto (evalCardinality\ oc\ q)\} \bullet \\
i\_groupCardinality\ quads\ oc = \\
\quad id\_groups\,\{sctid : SCTID;\ groups : \mathbb{P}\, GROUP \mid sctid \in \{q : \bigcup(\mathrm{ran}\ quadsByGroup) \bullet \\
\qquad \textbf{if}\ quad\_direction\ quads = source\_direction\ \textbf{then}\ q.s\ \textbf{else}\ object^{\sim} q.t\} \land \\
\quad groups = \{g : \mathrm{dom}\ quadsByGroup \mid (\exists\, q : quadsByGroup\ g \bullet \\
\qquad sctid = \textbf{if}\ quad\_direction\ quads = source\_direction\ \textbf{then}\ q.s\ \textbf{else}\ object^{\sim} q.t)\} \bullet \\
\qquad sctid \mapsto groups\} \\
\hline
\end{array}
$$

4.11 Cardinality

Interpretation: *cardinality* is tested against a set of quads with the following rules:

1. Errors are propagated
2. No cardinality or passing cardinality returns the sources / targets of the set of quads
3. Otherwise return an empty set

$$
\begin{array}{l}
i\_cardinality : \\
\quad cardinality[0\,..\,1] \to Quads\_or\_Error \to Sctids\_or\_Error \\
\hline
\forall\, card : cardinality[0\,..\,1];\ quads : Quads\_or\_Error \bullet \\
i\_cardinality\ card\ quads = \\
\textbf{if}\ quads \in \operatorname{ran} error \\
\quad \textbf{then}\ idgroups\_to\_sctids\,(quads\_to\_idgroups\ quads) \\
\textbf{else} \\
\quad idgroups\_to\_sctids\,(quads\_to\_idgroups\,(quad\_value\,(evalCardinality\ card\,(quads\_for\ quads), quad\_
\end{array}
$$

evalCardinality Evaluate the cardinality of an arbitrary set of type T.

- If the cardinality isn't supplied ($\#opt\_cardinality = 0$), return the set.
- If the number of elements is greater or equal to the minimum cardinality ($first\,(head\ opt\_cardinality)$) then:
  - If the max cardinality is an integer ($num^{\sim} second\,(head\ opt\_cardinality)$) and it is greater than or equal to the number of elements or:
  - the max cardinality is not specified ($second\,(head\ opt\_cardinality) = many$)

  return the set
- Otherwise return $\emptyset$

$$
\begin{array}{l}
[T] \\
\hline
evalCardinality : cardinality[0\,..\,1] \to \mathbb{P}\ T \to \mathbb{P}\ T \\
\hline
\forall\, opt\_cardinality : cardinality[0\,..\,1];\ t : \mathbb{P}\ T \bullet \\
evalCardinality\ opt\_cardinality\ t = \\
\quad \textbf{if}\ \#opt\_cardinality = 0 \lor \\
\qquad (\#t \geq first\,(head\ opt\_cardinality)) \land \\
\qquad (second\,(head\ opt\_cardinality) = many \lor \\
\qquad num^{\sim}(second\,(head\ opt\_cardinality)) \geq \#t) \\
\quad \textbf{then}\ t \\
\quad \textbf{else}\ \emptyset
\end{array}
$$

5 Substrate Interpretations

This section defines the interpretations that are realized against the substrate.

5.1 attributeName

`attributeName` is the interpretation of a `conceptReference` with the additional caveat that the SCTID(s) have to be substrate *attributeIds*

attributeName = conceptReference

$attributeName == conceptReference$

$$
\begin{array}{l}
i\_attributeName : \\
\quad Substrate \to attributeName \to Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ attName : attributeName \bullet \\
i\_attributeName\ ss\ attName = \\
(\textbf{let}\ attn == i\_conceptReference\ ss\ attName \bullet \\
\quad \textbf{if}\ attn \in \operatorname{ran} error \\
\quad\quad \textbf{then}\ attn \\
\quad \textbf{else if}\ (result\_sctids\ attn) \subseteq ss.attributeIds \\
\quad\quad \textbf{then}\ attn \\
\quad \textbf{else}\ error\ unknownAttributeId)
\end{array}
$$

5.2 attributeExpressionConstraint

`attributeExpressionConstraint` takes a substrate, an optional reverse flag, a set of attribute SCTIDs, an expression operator (equal or not equal) and a set of subject/target SCTIDS (depending on whether reverse flag is present) and returns a collection of quads that match / don't match the entry.

$$
\begin{array}{l}
i\_attributeExpressionConstraint : \\
\quad Substrate \to reverseFlag[0\,..\,1] \to Sctids\_or\_Error \to \\
\quad\quad expressionComparisonOperator \to Sctids\_or\_Error \to Quads\_or\_Error \\
\hline
\forall\, ss : Substrate;\ rf : reverseFlag[0\,..\,1];\ atts : Sctids\_or\_Error; \\
\quad op : expressionComparisonOperator;\ subj\_or\_targets : Sctids\_or\_Error \bullet \\
i\_attributeExpressionConstraint\ ss\ rf\ atts\ op\ subj\_or\_targets = \\
\textbf{if}\ atts \in \operatorname{ran} error \lor subj\_or\_targets \in \operatorname{ran} error \\
\quad \textbf{then}\ qfirstError\{atts, subj\_or\_targets\} \\
\textbf{else if}\ \#args.rf = 0 \land op = eco\_eq\ \textbf{then} \\
\quad quad\_value(\{t : result\_sctids\ subj\_or\_targets; \\
\qquad\quad a : result\_sctids\ atts;\ rels : ss.relationships \mid \\
\qquad\quad object^{\sim} rels.t = t \land rels.a = a \bullet rels\}, source\_direction) \\
\textbf{else if}\ \#args.rf = 1 \land op = eco\_eq\ \textbf{then} \\
\quad quad\_value(\{s : result\_sctids\ subj\_or\_targets;\ a : result\_sctids\ atts;\ rels : ss.relationships \mid rels.s \\
\textbf{else if}\ \#args.rf = 0 \land op = eco\_neq\ \textbf{then} \\
\quad quad\_value(\{t : result\_sctids\ subj\_or\_targets;\ a : result\_sctids\ atts;\ rels : ss.relationships \mid object^{\sim} \\
\textbf{else if}\ \#args.rf = 1 \land op = eco\_neq\ \textbf{then} \\
\quad quad\_value(\{s : result\_sctids\ subj\_or\_targets;\ a : result\_sctids\ atts;\ rels : ss.relationships \mid rels.s
\end{array}
$$

5.3 concreteAttributeConstraint

$$
\begin{array}{l}
i\_concreteAttributeConstraint : \\
\quad Substrate \rightarrow Sctids\_or\_Error \rightarrow concreteComparisonOperator \rightarrow \\
\qquad concreteValue \rightarrow Quads\_or\_Error \\
\hline
\forall\, ss : Substrate;\ atts : Sctids\_or\_error;\ op : concreteComparisonOperator; \\
\qquad val : concreteValue \bullet \\
i\_concreteAttributeConstraint = \\
\textbf{if } atts \in \mathrm{ran}\, error \\
\quad \textbf{then } qerror\,(error^{\sim} atts) \\
\textbf{else } quad\_value\{ss.relationships \mid ss.a \in (result\_sctids\ atts) \wedge \\
\qquad ss.t \in \mathrm{ran}\, concrete \wedge val \in concreteMatch\,(concrete^{\sim} ss.t)\ op\}
\end{array}
$$

$$
\begin{array}{l}
concreteMatch : \\
\quad CONCRETEVALUE \rightarrow comparison \rightarrow concreteValue
\end{array}
$$

Interpretation: Apply the substrate descendants (*descs*) or ancestors (*ancs*) function to a set of SCTID's in the supplied *Sctids_or_Error*. Error conditions are propagated.

$$
\begin{array}{l}
i\_constraintOperator : \\
\quad Substrate \pfun constraintOperator[0\,..\,1] \pfun Sctids\_or\_Error \pfun Sctids\_or\_Error \\
completeFun : (SCTID \pfun \mathbb{P}\, SCTID) \rightarrow SCTID \rightarrow \mathbb{P}\, SCTID \\
\hline
\forall\, ss : Substrate;\ oco : constraintOperator[0\,..\,1];\ subresult : Sctids\_or\_Error \bullet \\
i\_constraintOperator\ ss\ oco\ subresult = \\
\quad \textbf{if } error^{\sim} subresult \in ERROR \vee \#oco = 0 \\
\qquad \textbf{then } subresult \\
\quad \textbf{else if } head\ oco = descendantOrSelfOf \\
\qquad \textbf{then } ok(\bigcup\{id : result\_sctids\ subresult \bullet \\
\qquad\qquad completeFun\ ss.descendants\ id\} \cup result\_sctids\ subresult) \\
\quad \textbf{else if } head\ oco = descendantOf \\
\qquad \textbf{then } ok(\bigcup\{id : result\_sctids\ subresult \bullet \\
\qquad\qquad completeFun\ ss.descendants\ id\ \}) \\
\quad \textbf{else if } head\ oco = ancestorOrSelfOf \\
\qquad \textbf{then } ok(\bigcup\{id : result\_sctids\ subresult \bullet \\
\qquad\qquad completeFun\ ss.ancestors\ id\} \cup result\_sctids\ subresult) \\
\quad \textbf{else } ok(\bigcup\{id : result\_sctids\ subresult \\
\qquad\qquad \bullet\ completeFun\ ss.ancestors\ id\ \}) \\
\forall\, f : (SCTID \pfun \mathbb{P}\, SCTID);\ id : SCTID \bullet completeFun\ f\ id = \\
\quad \textbf{if } id \in \mathrm{dom}\, f \textbf{ then } f\ id \textbf{ else } \emptyset
\end{array}
$$

5.4 FocusConcept

```
focusConcept = [memberOf] conceptReference
```

5.4.1 focusConcept

`focusConcept` is either a simple concept reference or the interpretation of the `memberOf` function applied to a concept reference.

focusConcept ::=
    focusConcept_m⟨⟨*conceptReference*⟩⟩ |
    focusConcept_c⟨⟨*conceptReference*⟩⟩

Interpretation: If `memberOf` is present the interpretation of `focusConcept` is union the interpretation of `memberOf` applied to each element in the interpretation of `conceptReference`. If `memberOf` isn't part of the spec, the interpretation is the interpretation of `conceptReference` itself

$$
\begin{array}{|l}
\hline
i\_focusConcept : Substrate \rightarrow focusConcept \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ fc : focusConcept \bullet \\
i\_focusConcept\ ss\ fc = \\
\quad \textbf{if}\ focusConcept\_c^{\sim} fc \in conceptReference \\
\quad\quad \textbf{then}\ i\_conceptReference\ ss\ (focusConcept\_c^{\sim} fc) \\
\quad \textbf{else}\ i\_memberOf\ ss\ (i\_conceptReference\ ss\ (focusConcept\_m^{\sim} fc)) \\
\hline
\end{array}
$$

5.4.2 memberOf

`memberOf` returns the union of the application of the substrate *refset* function to each of the supplied reference set identifiers. An error is returned if (a) *refsetids* already has an error or (b) one or more of the refset identifiers aren't substrate *refsetIds*

$$
\begin{array}{l}
i\_memberOf : Substrate \rightarrow Sctids\_or\_Error \rightarrow Sctids\_or\_Error \\
i\_refset : Substrate \rightarrow SCTID \rightarrow Sctids\_or\_Error \\
\hline
\forall\, ss : Substrate;\ refsetids : Sctids\_or\_Error \bullet \\
i\_memberOf\ ss\ refsetids = \\
\quad \textbf{if}\ refsetids \in \operatorname{ran} error \\
\quad\quad \textbf{then}\ refsetids \\
\quad \textbf{else}\ bigunion\{sctid : result\_sctids\ refsetids \bullet i\_refset\ ss\ sctid\} \\
\\
\forall\, ss : Substrate;\ sctid : SCTID \bullet \\
i\_refset\ ss\ sctid = \\
\textbf{if}\ sctid \notin ss.refsetIds \\
\quad \textbf{then}\ error\ unknownRefsetId \\
\textbf{else if}\ sctid \in \operatorname{dom} ss.refsetIds \\
\quad \textbf{then}\ ok\ (ss.refsets\ sctid) \\
\textbf{else}\ ok\ \emptyset
\end{array}
$$

5.5 ConceptReferences

5.5.1 conceptId

```
conceptId = sctId
```

Interpretation: `conceptId` is interpreted as *SCTID* that it represents. For our purposes, all `conceptId`s are considered valid, so this is a bijection.

$[conceptId]$

$$
\begin{array}{l}
i\_conceptId : conceptId \rightarrowtail\!\!\!\!\!\rightarrow SCTID
\end{array}
$$

5.5.2 conceptReference

```
conceptReference = conceptId [ "|" Term "|"]
conceptId = sctId
```

Interpretation: `conceptReference` is interpreted as the *set of SCTID*s that are equivalent to the supplied SCTID if it is known concepts, c, in the substrate. If it isn't in the list of known concepts otherwise as the *unknownConceptReference* error. The `Term` part of `conceptReference` is ignored.

$conceptReference == conceptId$

$$i\_conceptReference : Substrate \rightarrow conceptReference \rightarrow Sctids\_or\_Error$$

$$\forall\, ss : Substrate;\ c : conceptReference \bullet i\_conceptReference\ ss\ c =$$
$$(\textbf{let}\ sctid == i\_conceptId\ c \bullet$$
$$\textbf{if}\ sctid \in ss.concepts\ \textbf{then}\ ok\ (ss.equivalent\_concepts\ sctid)$$
$$\textbf{else}\ error\ unknownConceptReference)$$

6 Glue and Helper Functions

This section carries various type transformations and error checking functions

6.1 Types

- **direction** – an indicator whether a collection of quads was determined as source to target (*source_direction*) or target to source (*targets_direction*)
- **Quads_or_Error** – a collection of *Quad*s or an error condition. If it is a collection *Quad*s, it also carries a direction indicator that determines whether it represents a set of sources or targets.
- **IDGroups** – a map from *SCTID*s to the *GROUP* they were in when they passed if successful, otherwise an error indication.

$$direction ::= source\_direction \mid targets\_direction$$
$$Quads\_or\_Error ::= quad\_value\langle\langle \mathbb{P}\ Quad \times direction \rangle\rangle \mid qerror\langle\langle ERROR \rangle\rangle$$

$$IDGroups ::= id\_groups\langle\langle SCTID \pfun \mathbb{P}\ GROUP \rangle\rangle \mid gerror\langle\langle ERROR \rangle\rangle$$

6.2 Result transformations

- **result_sctids** – the set of *SCTID*s in *Sctids_or_Error* or the empty set if there is an error
- **quads_for** – the set of quads in a *Quads_or_Error* or an empty set if there is an error
- **quad_direction** – the direction of a *Quads_or_Error* result. Undefined if error
- **to_idGroups** – the *SCTID* to *GROUP* part of in an id group or an empty map if there is error
- **quads_to_idgroups** – convert a set of quads int a set of id groups using the following rules:
  - If the set of quads has an error, propagate it
  - If the quad direction is *source_direction* (target to source) a list of unique relationship subjects and, for each subjects, the set of different groups it appears as a subject in
  - Otherwise return a list of relationship target sctids and, for each target, the set of different groups it appears as a target in.

- **idgroups_to_sctids** – remove the sctids in an idgroup sans the group identifiers.

$$
\begin{array}{l}
result\_sctids : Sctids\_or\_Error \to \mathbb{P}\, SCTID \\
quads\_for : Quads\_or\_Error \to \mathbb{P}\, Quad \\
quad\_direction : Quads\_or\_Error \nrightarrow direction \\
to\_idGroups : IDGroups \to SCTID \to \mathbb{P}\, GROUP \\
quads\_to\_idgroups : Quads\_or\_Error \to IDGroups \\
idgroups\_to\_sctids : IDGroups \to \mathbb{P}\, SCTID \\
\hline
\forall\, r : Sctids\_or\_Error \bullet result\_sctids\, r = \\
\quad \textbf{if } error^{\sim} r \in ERROR \textbf{ then } \emptyset \\
\quad \textbf{else } ok^{\sim} r \\
\\
\forall\, q : Quads\_or\_Error \bullet quads\_for\, q = \\
\quad \textbf{if } qerror^{\sim} q \in ERROR \textbf{ then } \emptyset \\
\quad \textbf{else } first\, (quad\_value^{\sim} q) \\
\\
\forall\, q : Quads\_or\_Error \bullet quad\_direction\, q = \\
\quad second\, (quad\_value^{\sim} q) \\
\\
\forall\, g : IDGroups \bullet to\_idGroups\, g = \\
\quad \textbf{if } gerror^{\sim} g \in ERROR \textbf{ then } \emptyset \\
\quad \textbf{else } id\_groups^{\sim} g \\
\\
\forall\, q : Quads\_or\_Error \bullet quads\_to\_idgroups\, q = \\
\quad \textbf{if } qerror^{\sim} q \in ERROR \textbf{ then } gerror(qerror^{\sim} q) \\
\quad \textbf{else if } quad\_direction\, q = source\_direction \\
\qquad \textbf{then } id\_groups\, \{s : SCTID \mid (\exists\, qr : quads\_for\, q \bullet s = qr.s) \bullet \\
\qquad s \mapsto \{qr : quads\_for\, q \bullet qr.g\}\} \\
\quad \textbf{else} \\
\qquad id\_groups\, \{t : SCTID \mid (\exists\, qr : quads\_for\, q \bullet t = object^{\sim} qr.t) \bullet \\
\qquad t \mapsto \{qr : quads\_for\, q \bullet qr.g\}\} \\
\\
\forall\, g : IDGroups \bullet idgroups\_to\_sctids\, g = \\
\quad \textbf{if } gerror^{\sim} g \in ERROR \textbf{ then } \emptyset \\
\quad \textbf{else } \mathrm{dom}(id\_groups^{\sim} g)
\end{array}
$$

Definition of the various functions that are performed on the result type.

- **firstError** – aggregate one or more *Sctids_or_Error* types, at least one of which carries and error and merge them into a single *Sctid_or_Error* instance propagating at least one of the errors (Not fully defined)
- **qfirstError** – convert two *Sctids_or_Error* types, into a *Quads_or_Error* propagating at least one of the errors. (not fully defined)

- **union** – return the union of two *Sctids_or_Error* types, propagating errors if they exist, else returning the union of the SCTID sets.

- **intersect** –return the intersection of two *Sctids_or_Error* types, propagating errors if they exist, else returning the intersection of the SCTID sets.
- **minus** – return the difference of one *Sctids_or_Error* type and a second, propagating errors if they exist, else returning the set of SCTID's in the first set that aren't in the second.
- **bigunion** – return the union of a set of *Sctids_or_Error* types, propagating errors if they exist, else returning the union of all of the SCTID sets.
- **bigintersect** – return the intersection a set of *Sctids_or_Error* types, propagating errors if they exist, else returning the intersection of all of the SCTID sets.

$$
\begin{array}{l}
firstError : \mathbb{P}\, Sctids\_or\_Error \pfun Sctids\_or\_Error \\
qfirstError : \mathbb{P}\, Sctids\_or\_Error \pfun Quads\_or\_Error \\
\\
union, intersect, minus : Sctids\_or\_Error \to Sctids\_or\_Error \to \\
\qquad Sctids\_or\_Error \\
bigunion, bigintersect : \mathbb{P}\, Sctids\_or\_Error \to Sctids\_or\_Error \\
\hline
\forall\, x, y : Sctids\_or\_Error \bullet union\ x\ y = \\
\qquad \textbf{if } x \in \mathrm{ran}\ error \lor y \in \mathrm{ran}\ error \textbf{ then } firstError\ \{x, y\} \\
\qquad \textbf{else } ok\ ((ok^{\sim} x) \cup (ok^{\sim} y)) \\
\\
\forall\, x, y : Sctids\_or\_Error \bullet intersect\ x\ y = \\
\qquad \textbf{if } x \in \mathrm{ran}\ error \lor y \in \mathrm{ran}\ error \textbf{ then } firstError\ \{x, y\} \\
\qquad \textbf{else } ok\ ((ok^{\sim} x) \cap (ok^{\sim} y)) \\
\\
\forall\, x, y : Sctids\_or\_Error \bullet minus\ x\ y = \\
\qquad \textbf{if } x \in \mathrm{ran}\ error \lor y \in \mathrm{ran}\ error \textbf{ then } firstError\ \{x, y\} \\
\qquad \textbf{else } ok\ ((ok^{\sim} x) \setminus (ok^{\sim} y)) \\
\\
\forall\, rs : \mathbb{P}\, Sctids\_or\_Error \bullet bigunion\ rs = \\
\qquad \textbf{if } \exists\, r : rs \bullet r \in \mathrm{ran}\ error \textbf{ then } firstError\ rs \\
\qquad \textbf{else } ok\ (\bigcup \{r : rs \bullet result\_sctids\ r\}) \\
\\
\forall\, rs : \mathbb{P}\, Sctids\_or\_Error \bullet bigintersect\ rs = \\
\qquad \textbf{if } \exists\, r : rs \bullet r \in \mathrm{ran}\ error \textbf{ then } firstError\ rs \\
\qquad \textbf{else } ok\ (\bigcap \{r : rs \bullet result\_sctids\ r\})
\end{array}
$$

7 Appendix 1

Representing optional elements of type T. Representing it as a sequence allows us to determine absence by $\#T = 0$ and the value by $head\ T$.

$$T[0\,..\,1] == \{s : \mathrm{seq}\ T \mid \#s \leq 1\}$$

8 Appendix 2

A generic function that takes:

- A substrate
- A function that takes a substrate, a sequence of type T and returns *Sctids_or_Error* (example: *i_subExpressionConstraint*)
- An operator that takes two *Sctids_or_Error* and returns a combination (example: *union*)
- A structure of the form "$T \times \text{seq}_1\ T$

And returns *Sctids_or_Error*

In the formalization below, *first seq_e* refers to the left hand side of the $T \times \text{seq}_1\ T$ and *second seq_e* to the right hand side. *head*(*second seq_e*) refers to the first element in the sequence and *tail*(*second seq_e*) refers to the remaining elements in the sequence, which may be empty ($\langle\rangle$).

$$
\begin{array}{|l}
[T] \\
\hline
\mathit{applyToSequence} : \mathit{Substrate} \to (\mathit{Substrate} \to T \to \mathit{Sctids\_or\_Error}) \to \\
\quad (\mathit{Sctids\_or\_error} \to \mathit{Sctids\_or\_error} \to \mathit{Sctids\_or\_error}) \to \\
\quad (T \times \text{seq}_1\ T) \to \mathit{Sctids\_or\_Error} \\
\hline
\forall\ \mathit{ss} : \mathit{Substrate};\ f : (\mathit{Substrate} \to T \to \mathit{Sctids\_or\_Error}); \\
\quad \mathit{op} : (\mathit{Sctids\_or\_error} \to \mathit{Sctids\_or\_error} \to \mathit{Sctids\_or\_error}); \\
\quad \mathit{seq\_e} : (T \times \text{seq}_1\ T) \bullet \\
\mathit{applyToSequence}\ \mathit{ss}\ f\ \mathit{op}\ \mathit{seq\_e} = \\
\textbf{if}\ \mathit{tail}(\mathit{second}\ \mathit{seq\_e}) = \langle\rangle\ \textbf{then} \\
\quad \mathit{op}\ (f\ \mathit{ss}\ (\mathit{first}\ \mathit{seq\_e}))(f\ \mathit{ss}\ (\mathit{head}\ (\mathit{second}\ \mathit{seq\_e}))) \\
\textbf{else} \\
\quad \mathit{op}\ (f\ \mathit{ss}\ (\mathit{first}\ \mathit{seq\_e}))(f\ \mathit{ss}\ ((\mathit{head}(\mathit{second}\ \mathit{seq\_e}))(\mathit{tail}(\mathit{second}\ \mathit{seq\_e}))))
\end{array}
$$