

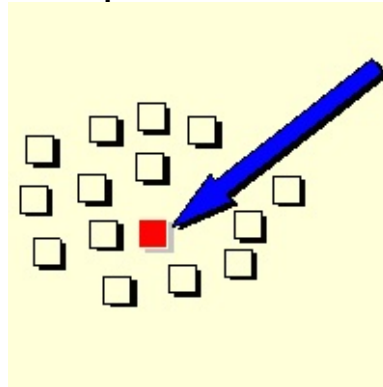
# A data model for BS8723 thesaurus structure

---

Leonard Will

Willpower Information

[www.willpowerinfo.co.uk](http://www.willpowerinfo.co.uk)



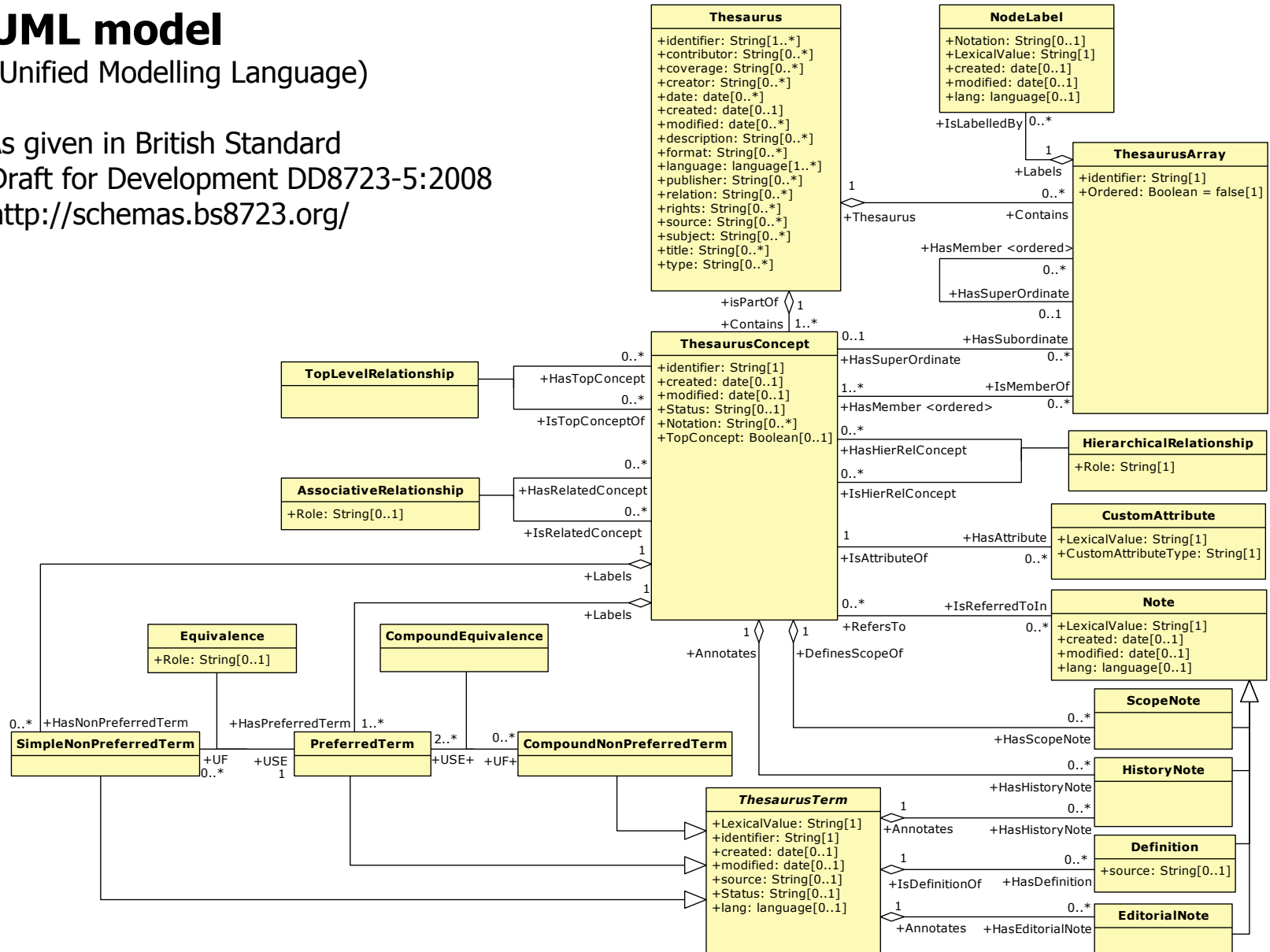
# UML model

(Unified Modelling Language)

As given in British Standard

Draft for Development DD8723-5:2008

<http://schemas.bs8723.org/>

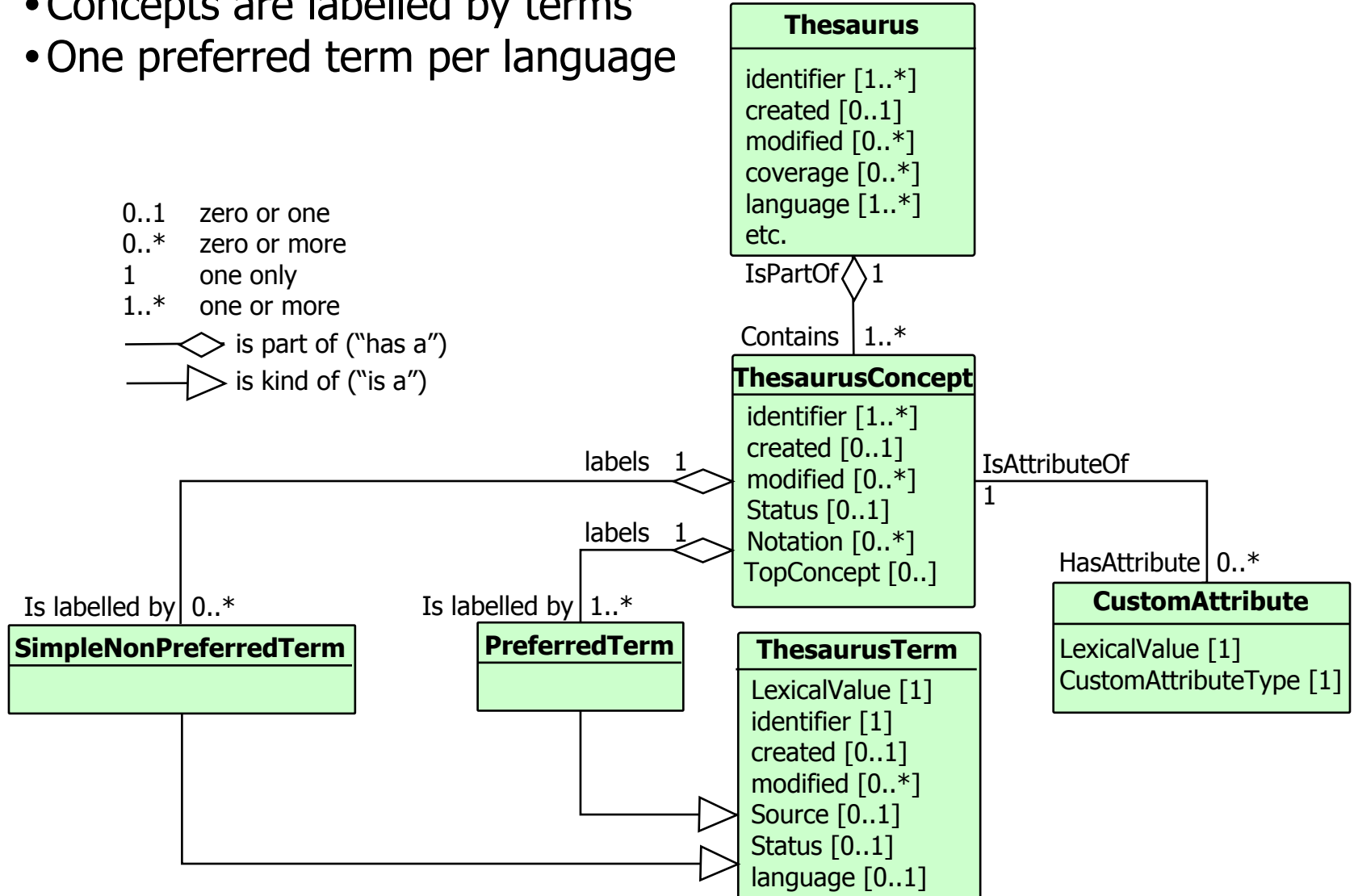


# A thesaurus is made up of concepts

- Concepts are labelled by terms
- One preferred term per language

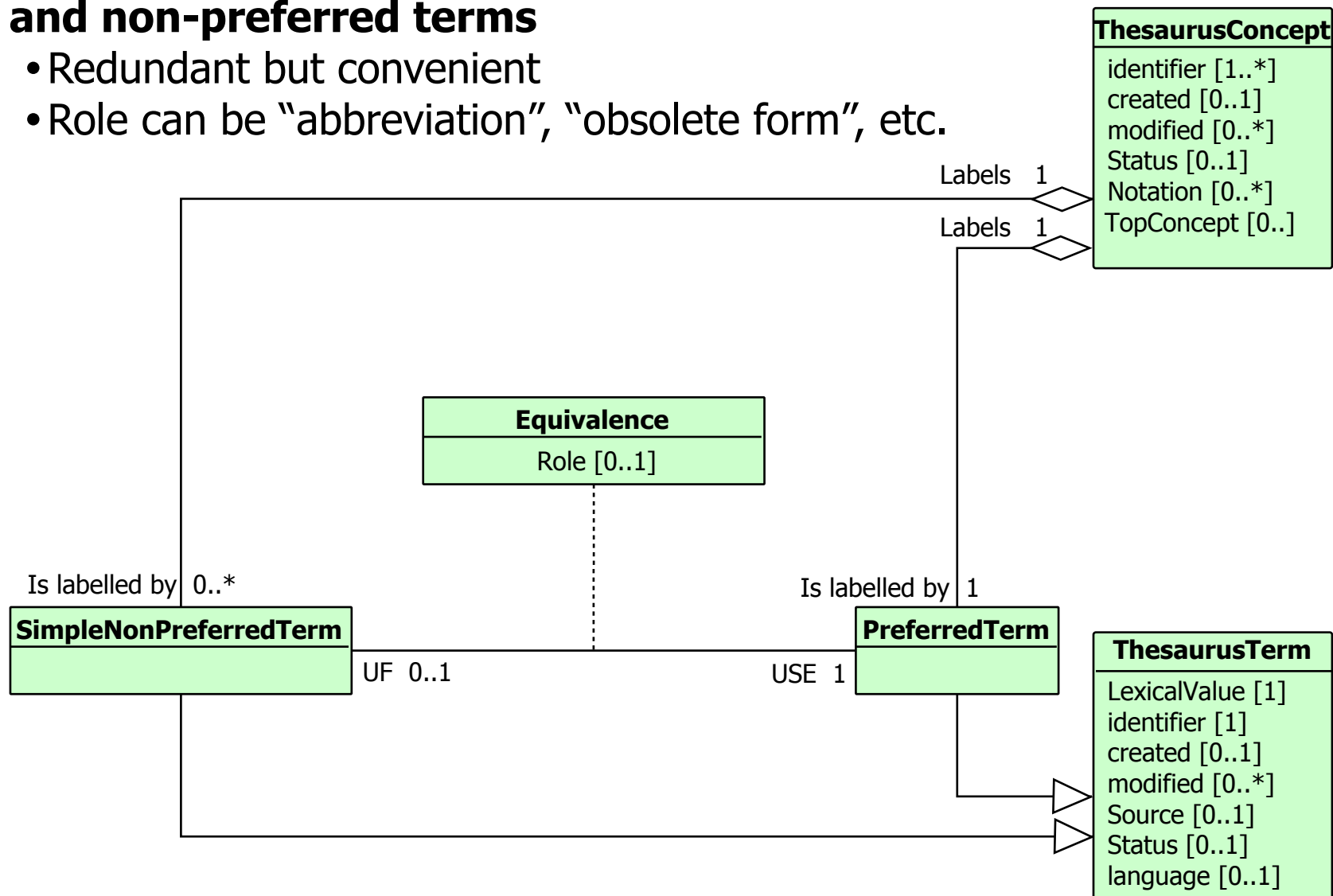
0..1 zero or one  
 0..\* zero or more  
 1 one only  
 1..\* one or more

—◇ is part of ("has a")  
 —▷ is kind of ("is a")



# Equivalence relationship between preferred and non-preferred terms

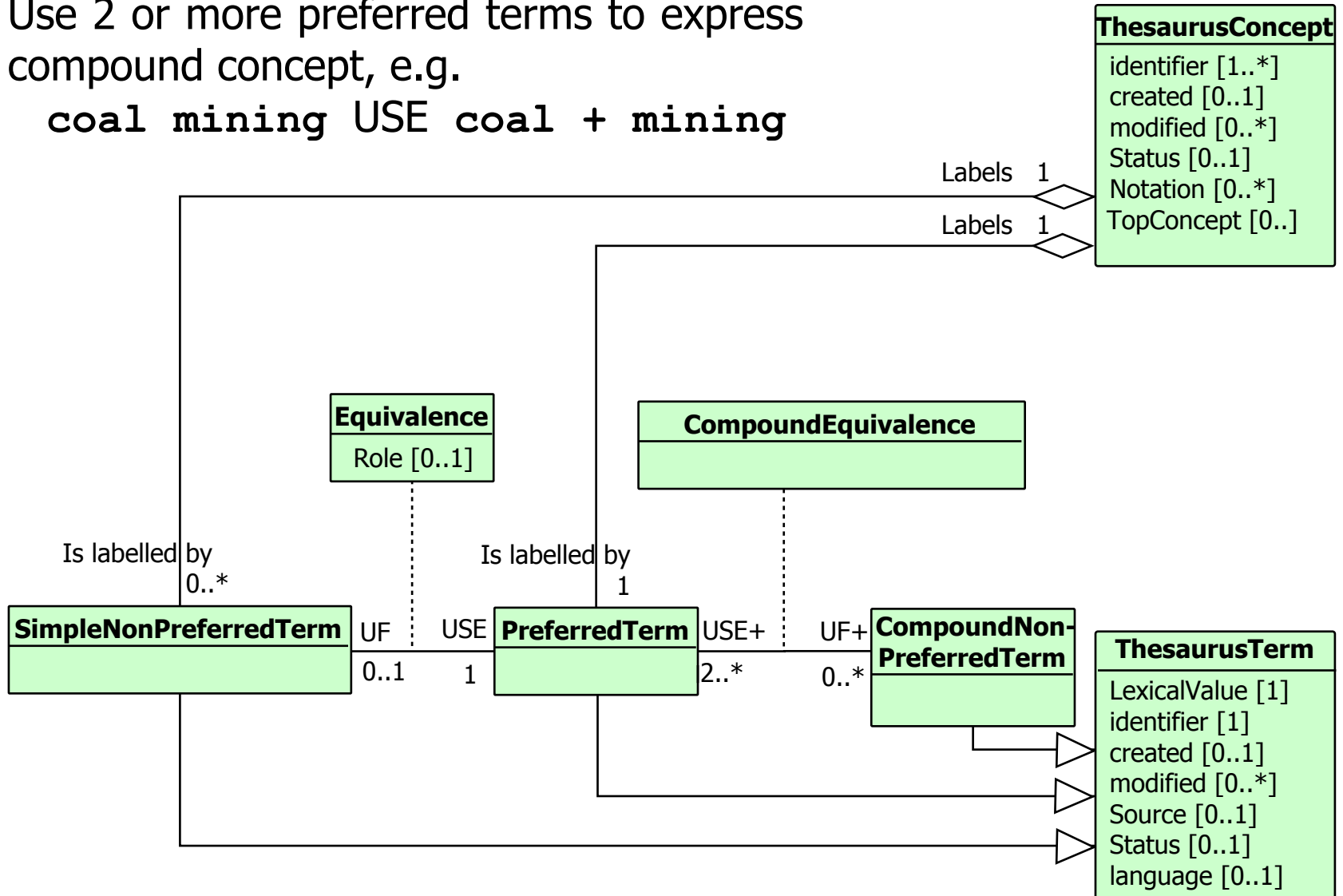
- Redundant but convenient
- Role can be "abbreviation", "obsolete form", etc.



# Compound equivalence

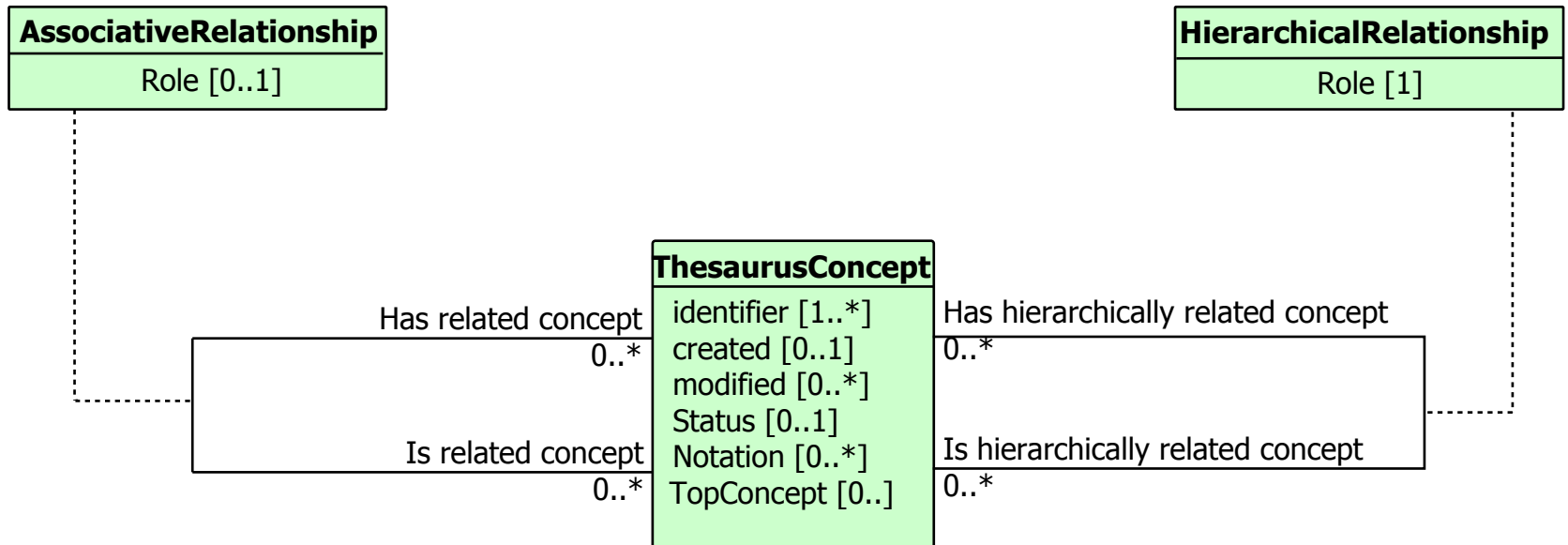
Use 2 or more preferred terms to express compound concept, e.g.

**coal mining** USE **coal** + **mining**



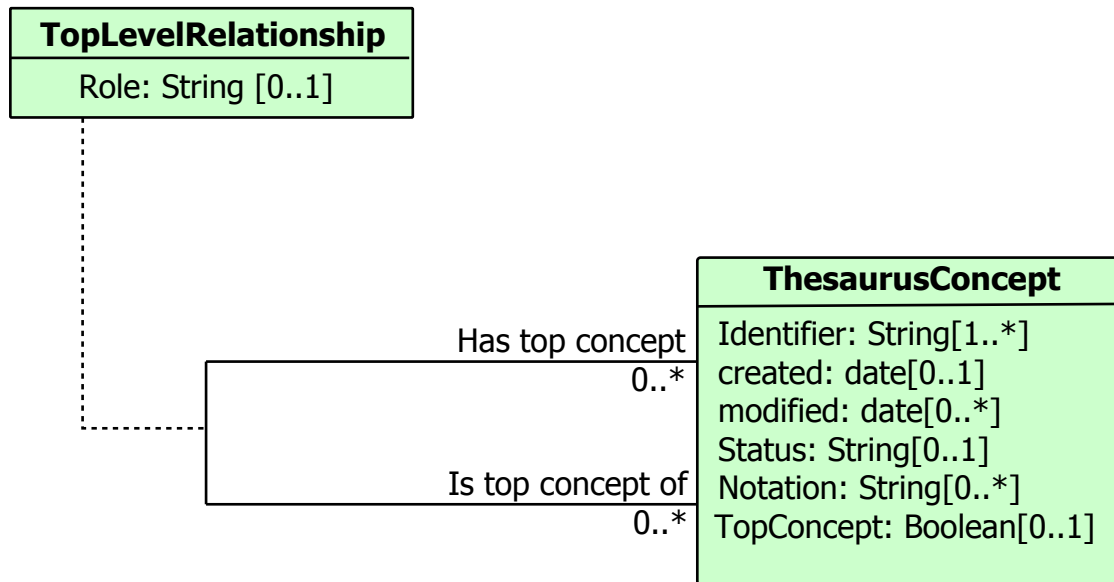
## Relationships between concepts

- Hierarchical relationship may have generalised role, e.g. "BT", "NT" or more specific role, e.g. "BTG", "BTP", "BTI"
- Associative relationship may have generalised role, e.g. "RT" (which need not be specified because the same in both directions), or more specific role, e.g. "CAUSE/EFFECT", "PROCESS/PRODUCT"



## Top level relationship

- Redundant but convenient (could be found by tracing links upward)
- Links a concept with the concept at the top of each hierarchy in which it occurs
- “TopConcept” attribute (Y/N) flags top concepts

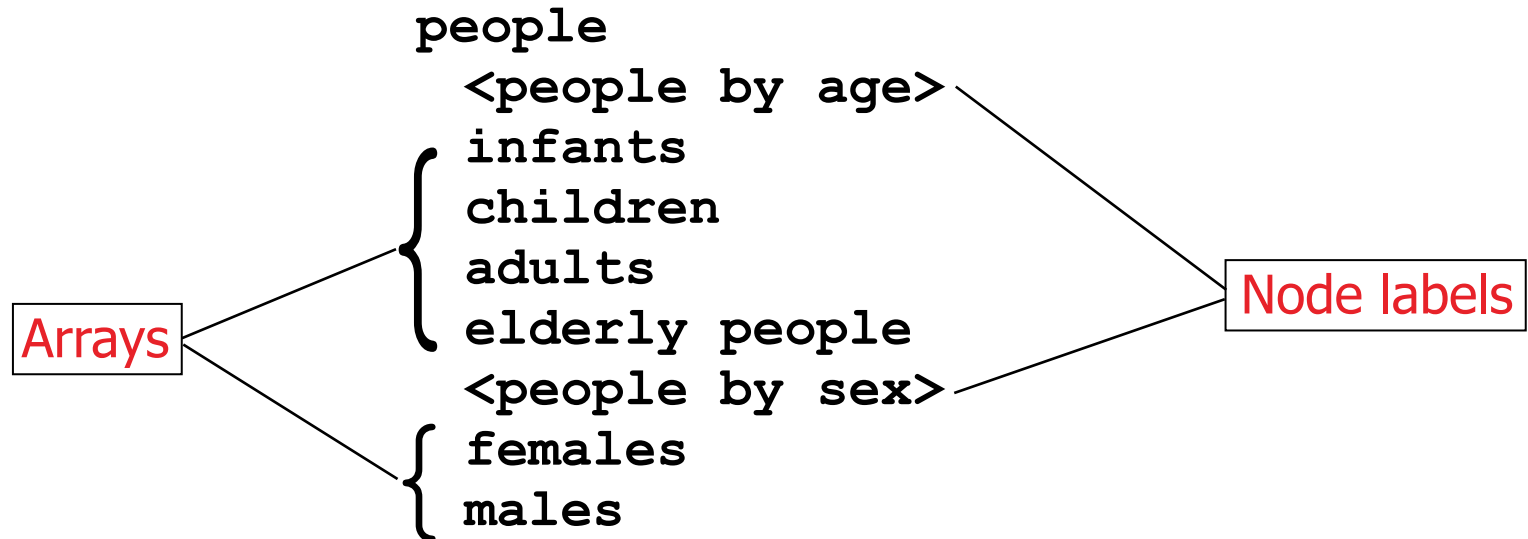


This diagram also shows that all attributes have a “type” specified, e.g. String, date, Boolean.

These have been omitted from other diagrams for clarity.

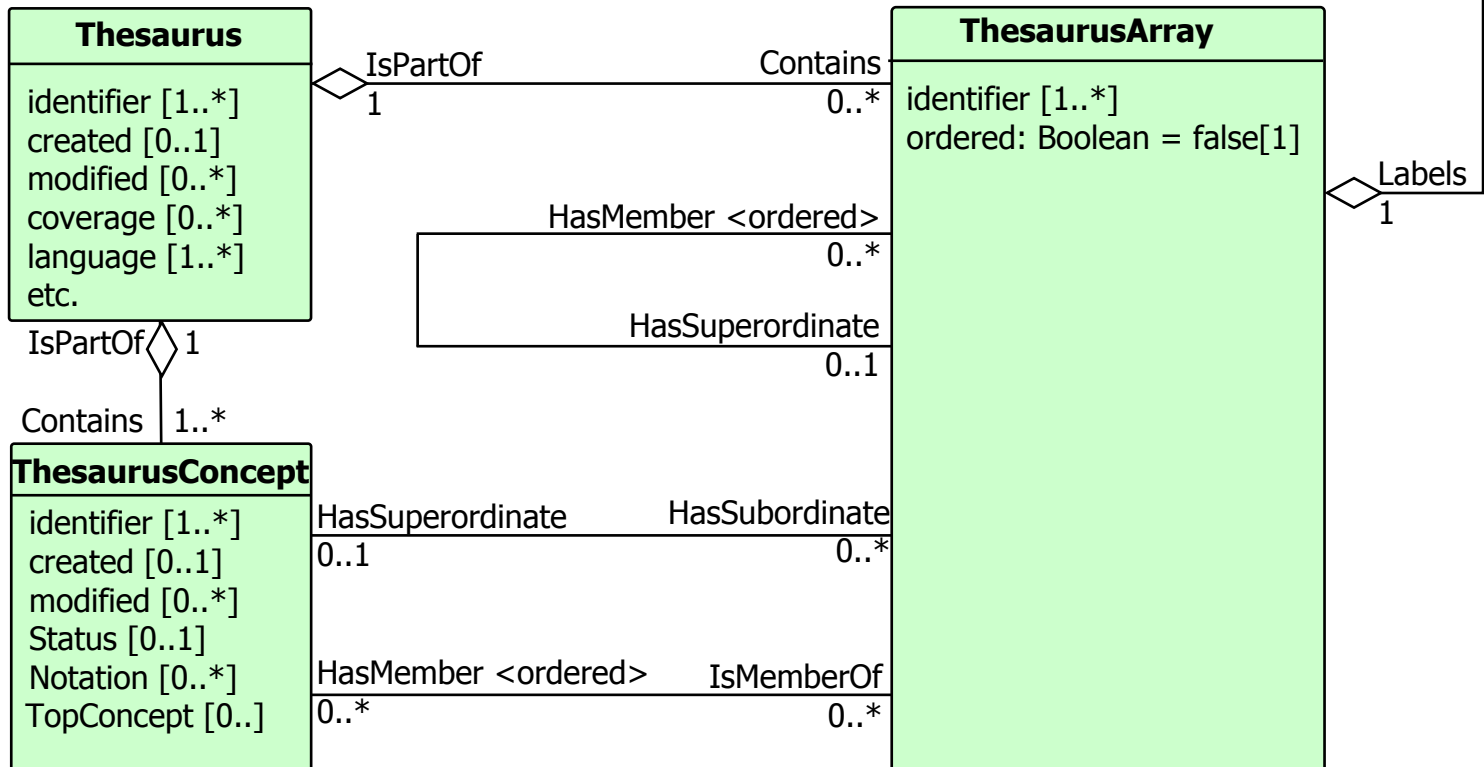
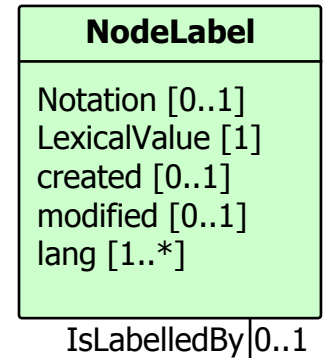
## Thesaurus arrays

- Group sibling concepts; sequence may be ordered
- Arrays can contain concepts or other arrays
- Node labels specify a "characteristic of division", e.g. **<people by age>**
- Super- and sub-ordinate relationship is not BT/NT



# Thesaurus arrays

- Group sibling concepts; sequence may be ordered
- Arrays can contain concepts or other arrays
- Node labels specify a “characteristic of division”, e.g. **<people by age>**
- Super- and sub-ordinate relationship is not BT/NT



# Notes

- Various kinds of notes can refer to concepts or terms
- Concepts may be referred to in notes on other concepts

