A Theory of Scope

Lars Marius Garshol

<larsga@bouvet.no>

TMRA 2007

2007-10-11

Why do we need a theory of scope?

- **Scope is defined in the TMDM**
  - but: the job is only half done
- **The following is not defined anywhere**
  - formal semantics of scope
  - how scope interacts with inferencing
  - how scope interacts with constraints in schemas
  - what scope operators we need for TMQL
  - ...
- **This talk aims to take a first step towards solving this**

Some background

Scope - a quick review

- **Scope applies to all *statements* in topic maps**
  - associations, occurrences, topic names, variant names
- **Scope is a set of topics**
  - possibly empty
- **Scope *qualifies* a statement**
  - that is, the scope defines the context in which the statement is considered to be valid
- **Scope enables conflicting views**
  - there is an expectation that statements in different scopes may conflict

Applications of scope

- **Multilinguality**
  - stating that a name or occurrence is in a particular language
- **Provenance**
  - giving the source of a particular statement
- **Opinion**
  - stating that a statement is true according to a particular authority
- **Time**
  - stating that a statement is true in a particular time period only
- **Audience**
  - stating that a statement is suitable for a particular audience
- **Filtering**
  - stating that a statement is inferred, and not in the base data

The AND/OR problem

- **Given**
  - statement @ (a, b)
- **When is it valid?**
  - in context a and context b? (OR)
  - only in context (a, b)? (AND)
- **The answer has a number of consequences...**
  - ISO 13250:2001 the answer is OR
  - XTM 1.0 the answer is undefined
  - TMDM the answer is AND

Restriction

- **Let's start with a single statement**
  - statement @ a
- **If we now add "b", is it valid in fewer or more contexts?**
  - if we choose AND, the answer is fewer
    - now b becomes required, in addition to the original a
    - so under AND adding topics narrows the scope
  - if we choose OR, the answer is more
    - before the statement only applied in a, now it also applies in b
    - under OR adding topics widens the scope

The unconstrained scope

- **This is defined as the scope used for statements that are universally valid**
- **But how is it to be represented?**
  - under OR it must be the set of all topics
    - given that adding topics widens the scope, the biggest is the widest
  - under AND it must be the empty set
    - given that removing topics widens the scope, the smallest is the widest

Duplication of statements

- **Under the OR interpretation**
  - statement @ (a, b, c) is equivalent to
  - statement @ a, statement @ b, statement @ c
  - this means that multi-topic scopes are not supported...
- **Under the AND interpretation**
  - statement @ (a, b, c) is implied by
  - statement @ (a, b)
  - if a statement is the same in different scopes, it has to be repeated

Why choose AND?

- **There are many reasons**
  - simpler representation of the empty scope
  - multi-topic scopes become much easier
  - ...
- **Variant names assume AND scope**
  - variant names inherit the scope of the topic name they belong to
  - this is done because they apply more narrowly than the topic name
  - this implies AND semantics
  - the AND choice was in other words built into XTM 1.0

The constraint problem

- **If the schema says**
  - every topic of type X must have exactly 1 occurrence of type Y
- **does this mean**
  - exactly 1 irrespective of scope?
  - or exactly 1 in each scope?
- **Do we need to be able to say**
  - which of the two we mean?
  - what the set of possible scopes is?

The theory

Basis of theory

- **In the paper the theory is formulated on TMRM**
  - using a particular TMDM mapping
  - this mapping is not published anywhere (yet)
- **This makes it tricky to present the theory here**
  - will simplify in this talk by ignoring how the topic map is actually represented
  - the paper has the full details

Three operators

- **Belief b(M, c)**
  - input: a topic map M, a set of *believed* topics c
  - output: a topic map where all statements we don't believe are removed
- **Disbelief d(M, c)**
  - input: a topic map M, a set of *disbelieved* topics c
  - output: a topic map where all statements we don't believe are removed
- ***Preference projection p(M, <)***
  - *input: a topic map M, and a preference relation between scopes <*
  - *output: a topic map where the non-preferred versions of conflicting statements have been removed*

Belief

- **What it does**
  - b(M, c) removes all statements whose scopes contain a topic not in c
- **If you believe everything, nothing is removed**
  - b(M, {all topics}) = M
- **If you believe nothing, only universally valid statements remain**
  - b(M, Ø) retains only statements in the unconstrained scope

Disbelief

- **What it does**
  - d(M, c) removes all statements whose scopes contain a topic in c
- **If you disbelieve nothing, you believe everything**
  - d(M, Ø) = b(M, {all topics}) = M
- **If you disbelieve everything, only universally valid statements remain**
  - d(M, {all topics}) = b(M, Ø)

Respecting the semantics

- **Given**
  - two statements *s* and *s'* where *scope(s)* $\subset$ *scope(s')*
- **no c exists such that**
  - s' in b(M, c), but s not in b(M, c)
- **The same is true of d(M, c)**

Formal semantics

- **Given a statement *s*, what other statements must be true?**
- **Basically, all statements**
  - that are equal to *s* in everything except the scope, and
  - whose scope is a superset of *scope(s)*
- **This might be added to the TMDM-TMRM mapping**

Inferencing

- **Given**
  - i instance-of t @ a
  - t subtype-of s @ b
- **we can infer**
  - i instance-of s @(a, b)
- **Rationale**
  - b(M, c) will never produce a topic map with the conclusion without one of the assumptions
  - the same applies to d(M, c)

Applying the theory

Multilinguality

- **Used to make a topic map support multiple languages**
  - norwegian - “Norsk” @ norwegian
    - “Norwegian” @ english
- **Requirement**
  - must be able to filter topic map by language
- **Solution**
  - d(M, {all other languages})

Provenance

- **Can be represented in topic maps using scope**
  - use a topic representing each data source
  - add that topic to the scope of each statement from a source
- **Various operations are conceivable**
  - show topic map according to source: b(M, {source})
  - remove data from untrusted sources: d(M, {untrusted sources})

Opinion

- **Example: my topic map about scripts and languages**
  - different script experts hold different, partially conflicting views
  - for example, experts use different classification systems
  - they also disagree on when a particular script was used, what other script it was derived from, etc
- **Solution**
  - scope statements by expert
  - use b(M, {expert}) to see topic map according to a single expert
  - (alternative: d(M, {all other experts})

Audience

- **Information resources scoped by audience**
  - end-user, technician, manager
  - doesn't matter if resources are modelled with occurrences or associations
- **Filter for audience using**
  - b(M, {end-user}) or
  - d(M, {technician, manager})

Time

- **Examples of use**
  - languages written in different scripts at different times (Soviet era, colonial era...)
  - topic map of conference series (people's affiliations etc change)
  - ...
- **Solution**
  - scope by era
  - b(M, {era})
  - d(M, {all other eras})

Filtering

- **If all inferred statements have the *inferred* topic in their scopes, this is easy**
  - d(M, { inferred })

Consequences

TMQL

- **TMQL currently has**
  - a syntactic shorthand for the b(M, c) operator,
  - but d(M, c) can also be expressed
- **Should there be a shorthand for disbelief?**
- **Should it possible to filter the topic map globally for the entire query?**
  - select ... from ... where ... believing foo, bar

TMCL: Solving the constraint problem

- **Should cardinality constraints ignore scope or be per scope?**
- **Checking the use cases we find:**
  - multilingual — per scope (but not for all statements)
  - provenance — ignore scope, perhaps
  - opinion — per scope (not all statements)
  - time — per scope (not all statements)
  - audience — ignore scope (doesn't really matter)
  - filtering — ignore scope (doesn't really matter)
- **Should this be taken into account in TMCL?**