A Theory of Scope

Lars Marius Garshol

Bouvet ASA, Oslo, Norway,
larsga@bouvet.no, http://www.ontopia.net

Abstract. This paper describes an interpretation of scope based on the past few years of research on the subject. Based on this it defines two mathematical operators for scope filtering which correspond to common use cases for scope.

1 Introduction

Scope is a feature of Topic Maps used to represent the qualification of a statement. That is, the scope represents the context of validity for the statement. Any statement in Topic Maps is qualified using a set of topics, possibly empty, and this is called the scope of the statement. The empty scope is known as *the unconstrained scope*, and statements in this scope are considered to have unlimited validity within the context of a particular topic map.

Scope has been present in Topic Maps from the publication of the initial Topic Maps standard [ISO13250], but there has never been any very definite specification of its interpretation. This section presents some of the key problems that have been identified, and describes the current state of the research into these problems.

1.1 The problem of context

The first, and most serious, problem with scope was that its interpretation was widely felt to be underspecified. For example, the original ISO standard defined scope as follows:

> The extent of the validity of a topic characteristic assignment [...]: the context in which a name or an occurrence is assigned to a given topic, and the context in which topics are related through associations.

The issue is what "extent of validity" and "context" really means. XTM 1.0 [XTM1.0] uses essentially the same definition, and so does not provide much additional information. The problem is perhaps best approached by examining the ways in which scope is used. Below follows a list of different uses of scope gathered from specifications, papers, tutorials, and other writings on the subject over the years:

– A name or occurrence might be valid in a particular language (multilinguality).

– A statement may derive from a particular source (provenance).
– A statement may be true according to a particular authority (opinion).
– An occurrence might be intended for a particular audience (say, "technician").
– A statement may only be true in a certain time period.
– A statement may have been inferred, instead of being actually present in the source data.

In all of these cases statements are seen as not being universally valid, but rather valid only in some context. There is also an expectation that statements in different contexts may contradict one another. Scope thus provides the ability to produce different, and potentially conflicting, views of the same subject area, with a scope being a definition of in what views a statement should be included.

Thus, the scope of a statement specifies in what contexts the statement is to be included, but the scope is not itself a context. Instead, a context is equivalent to one particular view of the topic map. It is also clear that each context is a subset of the full topic map. The question that remains is by what means the desired view of the topic map is specified. This takes us to the next problem in the interpretation of scope.

1.2 The multi-topic problem

Since a scope is a set of topics, it is not a given how to interpret a scope containing more than one topic. Are statements in this scope valid when all of the topics apply, or is it enough for one of them to apply? That is, if a statement appears in the scope $\{a, b\}$, does this mean that it is true only when both a and b apply, or also when only a or b apply?

The original Topic Maps standard made it clear (in clause 3.16) that the intended answer was "or". This was later viewed by the editors as a mistake. The next version of the standard, XTM 1.0, reversed the decision to make the answer "and". (XTM 1.0 is inconsistent on this point, and also stated that the interpretation was undefined.)

After the publication of these two standards there was a long phase where the model behind the two syntaxes was defined, and as part of this process substantial work was done on the interpretation of scope, primarily by Marc de Graauw. This work focused on the consequences of the "and"/"or" choice, and a number of important conclusions emerged from this work.

More and less restrictive scopes The first observation was that if scope is interpreted as being an "or" expression then the number of contexts in which a statement is valid increases as more topics are added to its scope. That is, a statement in scope a is valid in fewer contexts than a statement in scope $\{a, b\}$, given that the latter would apply whether a or b were in the context, whereas the former would apply only in the first case.

Obviously, the opposite applies when the interpretation is "and". In this case if a statement has the scope a and b it is not enough for only a to apply; b also has to apply. So as topics are added to the scope of the statement, the number of contexts in which it applies decreases.

The unconstrained scope Both the original Topic Maps standard and the XTM 1.0 specification had a concept of "the unconstrained scope", meaning the scope which claims that a statement is valid in all contexts. The original Topic Maps standard defined this as "the scope comprised of all topics in the topic map". XTM 1.0, however, was silent on the representation of this scope.

However, the arguments in the previous section make it clear that using the "and" interpretation of scope the least restrictive scope is the one which contains no topics at all. In other words, the representation of the unconstrained scope must be the empty set.

Similarly, in the "or" interpretation of scope, the least restrictive scope must be the biggest possible set of topics. This is clearly what ISO 13250:2000 intended by its definition, although other maximal sets of topics are imaginable.

Duplication of statements A related question is when a statement can be considered to imply another. Imagine the same statement appearing in scopes a and $\{a, b\}$. In this case, can one of the statements be considered to imply the other?

It turns out that if the scope interpretation is "or" then clearly a is implied by a or b. In fact, in the "or" interpretation any scope containing n topics is equivalent to n statements each with one of the n topics as scope. In other words, in the "or" interpretation of scope the set representation of scope is just a shorthand for repeating the statement in different single-topic scopes.

In the "and" interpretation the situation is clearly different, in that a and b is obviously implied by a. In fact, in the "and" interpretation there is no shorthand for saying "this is valid in situations X and Y"; the only way to do it is to have two separate statements for each case.

So which is it? The TMDM [ISO13250-2] deliberately chose the "and" interpretation as being the most useful interpretation, and also the one that caused the fewest internal problems for the interpretation and representation of the standard. In the "or" interpretation it is not really clear why scope has to be a set, nor is it clear how to represent the unconstrained scope in implementations.

At the same time, Piotr Kaminsky [Kaminsky02] independently made the observation that XTM 1.0 (and the TMDM model based on it) have a built-in assumption that the correct interpretation of scope is "and". In XTM 1.0 and TMDM variant names are alternative forms of a topic name which can be used instead of the topic name in some specific context. In other words, variant names have a more restricted validity than topic names do. This is achieved by letting them inherit the scope of the parent topic name, and requiring that they add at least one topic to this scope. But this assumes that adding topics to a scope makes it more restricted, which again assumes that the interpretation of scope is "and".

There is in other words every reason to assume that “and” was the correct choice.

Structured scope A third alternative to making a simple choice between “and” or “or” interpretations of a set of topics was to make the representation of scope more complex than just a simple set. This solution has been known as “structured scope” in the community, and several proposals have been put forward [deGraauw02] [Ahmed03].

The most complete proposal was developed at an ISO meeting in Montral in 2003 by Kal Ahmed, Ann Wrightson, and Dmitry Bogachev. It essentially turned scope into a set of nested sets, where each set was annotated with an “and” or “or” marker to define the interpretation. An extension to the XTM syntax was also proposed, as follows:

```
<occurrence>
  <scope compositor="AND">
    <topicRef xlink:href="english"/>
    <scope compositor="OR">
      <topicRef xlink:href="beginner"/>
      <topicRef xlink:href="intermediate"/>
    </scope>
  </scope>
  <resourceRef xlink:href="...."/>
</occurrence>
```

There are many equivalent ways to express the same scope in this representation, but there are well-known efficient algorithms for normalizing such expressions (that is, rewriting them to a canonical form). Thus duplicate suppression remains possible, even with this more complex representation of scope.

In the end, however, none of these proposals were adopted, for the following reasons:

- It required extensions to XTM 1.0, which the community in general wanted to avoid where possible.
- Structured scope is expensive to represent, for example in relational databases, and expensive to query.
- The TMDM model is already very complex, making the committee reluctant to complicate it further.
- Scope was at that time (and is still) rarely used, making it seem unlikely that there was any pressing need for this extension.

1.3 Scope and inferencing

The relationship between scope and inferencing has not so far received much attention, but it is clear that the two are closely related. A few examples will suffice to demonstrate this. Consider the following topic map (in LTM syntax):

```
tm:type-instance(i : tm:instance, t : tm:type)
tm:supertype-subtype(s : tm:supertype, t : tm:subtype)
```

Here it is clearly safe to assume that i is an instance of s. But what about in this case?

```
tm:type-instance(i : tm:instance, t : tm:type) / a
tm:supertype-subtype(s : tm:supertype, t : tm:subtype) / a
```

It does not appear safe to assume that i is an instance of s in the unconstrained scope, but that this is valid in the scope of a seems unproblematic. And the following case?

```
tm:type-instance(i : tm:instance, t : tm:type) / a
tm:supertype-subtype(s : tm:supertype, t : tm:subtype) / b
```

Here it does not seem permissible to assume anything, except perhaps that i is an instance of s in the scope $\{a, b\}$. The reasoning is that under the "and" interpretation this is the minimal context in which both of the statements required for the inference are valid.

1.4 Scope and reification

It has often been pointed out that anything which can be expressed using scope can also be expressed by means of reification. For example, if one says

```
[norway = "Norge" / norwegian]
```

this can also be expressed using reification:

```
[norway = "Norge" ~ no-no-name]
dc:language(no-no-name : dcc:subject, norwegian : dcc:value)
```

In both cases we are stating that the name is Norwegian, but in one case we are using scope, and in another we are using reification. It could be argued that the reification approach is more explicit, and it could also be argued that the scope approach is simpler and more light-weight.

The main difference between the two is that the Topic Maps technologies have better support for the scope approach (in constraints and querying) and less good support for the reification approach.

1.5 Scope and the identity of statements

It is worth noting that in the TMDM the scope property is included when statements are compared for equality. This means that the following topic map contains two statements:

```
[norway = "Norge" / norwegian
        = "Norge" / swedish]
```

This also means that a topic reifying one of these names will specifically reify either the Swedish or the Norwegian name, and not both. In other words: scope is part of the identity of a statement.

1.6 Other related work

This section has reviewed much work that is directly related to this article, but there is also a body less directly related work.

An early paper on scope in Topic Maps was [Pepper01], perhaps best known for introducing the notion of "axes of scope", which is still relevant, but tangential to the theory presented in this paper. It also presented some early examples of filtering operators.

Much work on context has also been performed in the Semantic Web community, where "named graphs" has been proposed as one solution for RDF. Important papers from this work are [**?**] and [Carroll05].

There is also a rich body of work on context outside semantic technologies, such as the theory f

2 The theory of scope

The theory of scope is formulated on the TMRM [ISO13250-5] representation of Topic Maps currently being defined by ISO. The representation is not entirely finalized yet, but is close to its final form, and updating the theory to conform to the final representation, if necessary at all, should be easy.

The theory consists of two operators which filter the topic map to leave only those statements which are true in a given context. In this paper they are based on the TMRM representation of Topic Maps, but they are not bound to it, however, and could be redefined based on any Topic Maps representation.

2.1 The TMRM mapping of the TMDM

An understanding of this representation is essential in order to follow this section, and so this section gives a brief recapitulation of the mapping.

In the mapping, topics are represented by TMRM proxies of the form shown below. The set of all topic proxies is denoted $\mathcal{T}$.

```
{ <subject-identifier, ...>,
  <subject-locator, ...>,
  <item-identifier, ...> }
```

All statements take the form shown below. The set of all statement proxies is denoted $\mathcal{S}$.

```
{ <type, ...topic proxy...>,
  <scope, ...scope proxy...>,
  <roletype-1, ...topic proxy...>,
  <roletype-2, ...topic proxy...>,
  ...
  <roletype-n, ...topic proxy...> }
```

Scope proxies are of the following form:

```
{ <member, ...topic proxy...>,
  <member, ...topic proxy...>,
  ... }
```

Two operators from the TMRM are used in this paper. The first is $k \leftarrow_M v$, which returns all proxies which contain $\langle k, v\rangle$. The second is $p \rightarrow k$, which returns all values for the key k in the proxy p.

For convenience we also define the function $\sigma : \mathcal{S} \rightarrow 2^{\mathcal{T}}$, which given a statement produces its scope, as follows:

$$\sigma(s) = s \rightarrow \texttt{scope} \rightarrow \texttt{member} \tag{1}$$

2.2 Statements and contexts

One assumption underlying the theory is that scope is not part of the statement, but a qualification of the statement. Consider the following statement (in LTM):

```
part-of(norway : part, denmark : whole) / eighteenth-century
```

The meaning of this is that if your context includes the eighteenth century, you can assume that Norway is part of Denmark. If it does not, you can not make that assumption.

So once a topic map has been filtered (for example using the two operators) to include only those statements which can be assumed to be true in a particular context, scope can effectively be ignored. All statements in the filtered topic map have passed the filter, and are therefore considered true. Whether, and how, the statements are qualified is then inconsequential.

2.3 Belief

In order to model belief we define the function $b(M, s)$ where is a topic map and s is a set of proxies representing the topics in which we believe. The function returns a subset of the topic map containing only the statements which are considered true in this context.

The b function must have the following properties:

- For all models M it must be the case that $b(M, \mathcal{T}) = M$. That is, if we believe everything, then all statements in the map must also be believed.
- Similarly, if we believe nothing, only statements in the unconstrained scope must be accepted. That is, for all maps M it must be the case that

$$b(M, \emptyset) \cap \mathcal{S} = \texttt{scope} \leftarrow_M \emptyset \tag{2}$$

Given the semantics of the scope set, it is clear that if a statement q has a and b in its scope, both a and b must be believed for q to be believed. That is, statements are believed iff their scopes contain no topics which we do not believe. This leads directly to the following formulation of b:

$$b(M, s) = \{q \in M | q \notin \mathcal{S} \vee \sigma(q) \subset s\} \tag{3}$$

It is easy to see that this formulation of b has the properties stated above.

2.4 Disbelief

In order to model disbelief we define the function $d(M, s)$, which produces the subset of M which could still potentially be true, but leaves out the statements which we do not believe. d should have the properties that:

- when there is nothing we disbelieve all statements in the topic map should be believed. That is, $d(M, \emptyset) = M$.
- Similarly, if we disbelieve everything, that should be the same as believing nothing, ie: $d(M, \mathcal{T}) = b(M, \emptyset)$.

The formulation of the function is easy, as it simply removes statements whose scopes contain a topic we disbelieve:

$$d(M, s) = \{q \in M | q \notin \mathcal{S} \vee \sigma(q) \cap s = \emptyset\} \tag{4}$$

Again it is easy to see that this formulation has the desired properties.

2.5 Semantics of the operators

More and less restrictive scopes The reasoning given above in section 1.2 on page 3 shows that the validity of a scope becomes narrower as more topics are added to it. The b and d operators need to respect this in order to accurately reflect the semantics of scope. In other words, given two statements $q, q' \in M$, if $\sigma(q) \subset \sigma(q')$ there cannot be any context c for which $q' \in b(M, c)$ but not $q \in b(M, c)$.

It's easy to see that $q' \in b(M, c)$ implies that $\sigma(q') \subset c$. Given that we know $\sigma(q) \subset \sigma(q')$, it follows that $\sigma(q) \subset c$, and therefore $q \in b(M, c)$. In other words, the b operator does honour the semantics.

A similar argument applies to d, in that assuming the same relationship between q and q' there cannot be any c for which $q' \in d(M, c)$ but not $q \in d(M, c)$. Here $q' \in d(M, c)$ implies that $\sigma(q') \cap c = \emptyset$. Given that q's scope is a subset of q''s, it's clear that the claim holds here, too.

Equivalence A key question for the theory to answer is what sets of statements could be considered equivalent. For example, is

```
part-of(norway : part, denmark : whole) / eighteenth-century
```

equivalent to the following?

```
part-of(norway : part, denmark : whole) / eighteenth-century
part-of(norway : part, denmark : whole) / eighteenth-century lmg
```

In fact, it is. We know that any filtered topic map which contains the second association will also contain the first. Further, in filtered topic maps scopes are ignored (per section 2.2 on the previous page), and so we know the second topic map makes the relationship between Norway and Denmark hold in exactly the same contexts as the first topic map.

Inferencing We are now ready to revisit the question of inferencing from section 1.3 on page 4 in more depth, and show that there are two different ways to approach the problem, and that both give the same result.

The first approach is to use inference rules that ignore scope, and to apply them so that when a set of statements z in the topic map cause another set of statements z' to be inferred, the scope of all statements in z' is set to the union of the scope of all statements in z. This way, it is impossible to use the b and d operators in such a way that a filtered topic map contains an inferred statement (that is, a statement in z') without all assumed statements (that is, all statements in z) also being included.

The second approach is to filter the topic map first, and then to perform inferencing afterwards, using the remaining statements, again ignoring scope in the inferencing process itself. This also ensures that all assumptions necessary for an inferred statement are included in the topic map when the inferred statement is included.

3 Evaluation

It's not enough for the scope operators to be correct, they must also be useful, and the best test of this is whether or not they support the use cases outlined in section 1.1 on page 1.

3.1 Language

This use case is usually handled by representing each language as a topic, and attaching to each name and occurrence a scope showing what language the name/occurrence is valid in.

There are two different ways to solve this with the two operators:

- Believe only the desired language, and no other language. This will return a topic map with only text in the desired language. If no other topics are believed this means that statements in non-language scopes will be filtered out. This can be avoided by also believing other non-language topics as necessary to keep desired statements.
- Disbelieve all languages other than the desired language. This produces a topic map containing only text in the desired language without removing statements in non-language scopes. This may be easier if it is for some reason difficult to know what other scopes to believe.

3.2 Provenance

Provenance is a common name for the use case of tracking what information derives from what source. In Topic Maps this can be done by creating a topic for each source and then adding the source topic to the scope of all statements from each source.

There are several different ways this can be used. One is to believe a particular source (or set of sources) and thus see the topic map according to these sources. Another is to disbelieve a source (or, again, a set of sources) and thus remove suspect information from the topic map.

3.3 Opinion

In some cases there is a lack of agreement about facts in a domain. For example, the Scripts and Languages topic map classifies scripts into six different types of script (alphabet, abjad, abugida, featural, syllabary, logographical), following the terminology of Peter T. Daniels. However, other experts on writing systems, like William Bright, use another classification (alphabet, abjad, alphasyllabary, featural, syllabary, logographical) [Bright00]. Similarly, opinions are divided on which scripts were derived from which other scripts, etc.

This can all be expressed in Topic Maps using a scope for each school of opinion, and scoping disputed statements accordingly, as follows:

```
type-instance(devanagari : instance, abugida : type) / daniels
type-instance(devanagari : instance, alphasyllabary : type) / bright
type-instance(latin : instance, alphabet : type)
```

This again makes it possible to see the topic map according to a particular school of opinion by believing one school of opinion, or disbelieving one or more.

3.4 Audience

Relevant information resources, whether connected to subject topics via associations or occurrences, might be suitable only for particular audiences. Examples of audiences might be technicians, end-users, engineers, or managers.

Here, the goal is to show a member of a particular audience only relevant resources. This is easily done by either believing that audience, or disbelieving all other audiences. The use case is very similar to that of language.

3.5 Time

Some statements are only true in a given time period, such as:

```
part-of(norway : part, denmark : whole) / eighteenth-century
```

Again, believing one time period or disbelieving all other time periods will solve this.

3.6 Inferred statements

One possible scope for statements is to mark them as being inferred, as opposed to being actually present in the source data. The operation users are most likely to want in this case is to view the topic map without the inferred statements. Assuming `inferred` is a scoping topic, this is easily handled with $d(M, \{\texttt{inferred}\})$.

4 Conclusion

We have proposed a theory of scope, and shown that it is compatible with the current knowledge about scope in Topic Maps, and that it can handle the known use cases for scope.

More work remains before scope is fully understood, however, especially on the interaction between constraints and scope, but that is beyond the scope of this paper.

References

[deGraauw02] de Graauw, M.; *Structuring scope*; ISO SC34 working document N0347, 2002-11-20, `http://www.marcdegraauw.com/files/structuring_scope.htm` and `http://www.jtc1sc34.org/repository/0347.htm`.

[Ahmed03] Pepper, S.; *ISO/IEC JTC 1/SC34/WG3 Meeting Report (Montreal, 1-4 August 2003)*; ISO SC34 working document N0439, 2003-05-07, `http://www.jtc1sc34.org/repository/0439.htm`.

[ISO13250-5] ISO/IEC CD 13250-5: Topic Maps – Reference Model, 2007-11-24, International Organization for Standardization, Geneva, Switzerland. `http://www.jtc1sc34.org/repository/0939.pdf`.

[ISO13250-2] *ISO/IEC IS 13250-2:2006: Topic Maps – Data Model*, International Organization for Standardization, Geneva, Switzerland, 2006. `http://www.isotopicmaps.org/sam/sam-model/`.

[ISO13250] ISO/IEC IS 13250:2000: *Information Technology - Document Description and Processing Languages - Topic Maps.* International Organization for Standardization, Geneva, Switzerland. `http://www.y12.doe.gov/sgml/sc34/document/0129.pdf`

[XTM1.0] Pepper, S.; Moore, G.; *XML Topic Maps (XTM) 1.0*; 2001-08-06; TopicMaps.Org specification; `http://www.topicmaps.org/xtm/1.0/`.

[Kaminsky02] Kaminsky, P.; *Integrating Information on the Semantic Web Using Partially Ordered Multi Hypersets*; master of science thesis, University of Victoria; 2002; `http://www.ideanest.com/braque/Thesis-web.pdf`.

[Guha04] Guha, R.; McCool, R.; Fikes, R.: *Contexts for the Semantic Web*; In: McIlraith, S. A.; Plexousakis, D.; van Harmelen, F.: The Semantic Web – ISWC 2004. LNCS 3298, Springer (2004).

[Bright00] Bright, W.; *A Matter of Typology: Alphasyllabaries and Abugidas*; Studies in the Linguistic Sciences; Volume 1, Number 1 (Spring 2000); ISSN 0049-2388; University of Illinois at Urbana-Champaign; `http://www.linguistics.uiuc.edu/sls/`

[Pepper01] Peper, S.; Grnmo, G. O.: *Towards a General Theory of Scope*; Proceedings of Extreme Markup 2001; `http://www.ontopia.net/topicmaps/materials/scope.htm`.

[Carroll05] Carroll, J. J.; Bizer, C.; Hayes, P.; Stickler, P.: *Named Graphs, Provenance and Trust*; HP Labs technical report HPL-2004-57R1; 2005-04-19; `http://www.hpl.hp.com/techreports/2004/HPL-2004-57R1.html`.

[Costanza05] Costanza, P.; Hirschfeld, R: *Language Constructs for Context-oriented Programming – An Overview of ContextL*; In Proceedings of the Dynamic Languages Symposium (DLS), co-located with the Conference on Object Oriented Programming Systems Languages and Applications (OOPSLA), San Diego, California, USA, October 18, 2005, ACM DL; http://www.swa.hpi.uni-potsdam.de/publications/media/CostanzaHirschfeld_2005_LanguageConstructsForContextOrientedProgrammingAnOverviewOfContextL_AcmAuthorsVersion.pdf.