# Topic Maps: Introduction and Allegro


## **Dr. Hans Holger Rath**, *Director Consulting*

**STEP Electronic Publishing Solutions GmbH**
*Consulting*
Technologiepark Würzburg-Rimpar
Pavillon 7
D-97222
Rimpar
Germany
Tel: +49.9365.8062.0
Fax: +49.9365.8062.66
Email: consulting@step.de
WWW: http://www.step.de/


### Biography:

Dr. Hans Holger Rath has been director of STEP's Consulting department since April 1998. He started at STEP in April 1996 as senior consultant/project manager. Before he joined STEP he was head of the Document Computing department at ZGDV (Computer Graphics Center, Darmstadt, Germany). Dr. Rath studied computer science in Karlsruhe 1984-1990 and graduated at the TU Darmstadt with the doctoral thesis "Literate Specifying of Hypermedia Documents" in 1996. He was involved in DTD developments for DIN (Deutsches Institut für Normung e.V. – German Standards Institute) and ISO (International Organization for Standardization) and cooperates very closely with publishing houses, aircraft industry and telecommunication industry. All in all he has more than nine years experience in information architectures and related topics. Since 1998 he has represented Germany in ISO/IEC JTC 1/SC 34 – the ISO committee standardizing SGML, HyTime, DSSSL, Topic Maps etc.


## **Steve Pepper**, *Senior Information Architect*

**STEP Infotek A.S**
Gjerdrums vei 12
N-0486 Oslo
Norway
Tel: +47 22 02 16 87
Fax: +47 22 02 16 81
Email: pepper@infotek.no
WWW: http://www.infotek.no


### Biography:

Steve Pepper is the Senior Information Architect at STEP Infotek, a company in the STEP group (based in Germany, Hungary, Norway, Poland, and the UK) that specialises in information reengineering.

Originally trained as a typographer, Steve has worked with structured information since 1988 and participated in a wide range of SGML-related projects in both the public and private sectors in Norway and abroad. He is currently heavily involved in the implementation of document management solutions for leading European reference works publishers and is a principal architect of the Reference Works Module functionality of STEP's flagship product, SigmaLink.

Steve represents Norway on ISO/IEC JTC 1/SC 34, where he convenes WG 3 (Information association) with responsibility for HyTime, Topic maps, ISO-HTML, ISMID (International Standard Metafile for Interactive Documents) and SMDL (Standard Music Description Language). He has played an active role in the development of the new topic map standard (ISO/IEC 13250) and currently represents SC 34 on the “ISO Collaborative PDML (Product Data Markup Language)” project.

A frequent speaker at SGML events around the world, he is the author and maintainer of the popular *Whirlwind Guide to SGML and XML tools*, which is freely available on the internet at `http://www.infotek.no/sgmltool/`, and co-author (with Charles Goldfarb and Chet Ensign) of the *SGML Buyer's Guide*, a comprehensive guide to choosing SGML and XML products and services.

---


**Abstract:**

*The new ISO standard ISO/IEC 13250 Topic Maps defines a model and architecture for the semantic structuring of link networks. Dubbed the “GPS of the information universe”, topic maps will become the solution for organizing and navigating large and continuously growing information pools, and provide a “bridge” between the domains of knowledge representation and information management.*

*This paper is divided into two parts: The first part, “Introduction”, describes the roots, concepts, and possible applications of the topic map standard. The second part, “Allegro”, presents several technical issues which are of great interest when applying topic maps to real world applications. The main focus of the paper is the introduction of “topic map templates” – a semi-official term coined by the standards' committee for a concept that the authors argue is a necessary but as yet unstandardized addition to the basic model.*


---

# 1. Overview

The ISO (International Organization for Standardization) committee JTC 1/SC 34/WG 3 *Information Technology – Document Description and Processing Languages – Information Association* standardized ISO/IEC 13250 Topic Maps [ISO13250] in the summer of 1999. Formally speaking, the ISO standard defines a model and interchange syntax for Topic Maps. The initial ideas – which date back to the early 1990's – related to the desire to model intelligent electronic indexes in order to be able to merge them automatically. But during several years of gestation, the topic map model has developed into something much more powerful that is no longer restricted to simply modelling indexes.

A topic map annotates and provides organising principles for large sets of information resources. It builds a structured semantic link network above those resources. The network allows easy and selective navigation to the requested information. Topic maps are the "GPS (Global Positioning System) of the information universe". Searching in a topic map can be compared to searching in knowledge structures. In fact, topic maps are a base technology for knowledge representation and knowledge management.

The basic concepts of the standard are topics, occurrences of topics, and relationships ("associations") between topics. Section NUTSHELL gives an extensive overview.

The editors of the standard, together with the other members of ISO JTC1/SC34/WG3 (the authors are among those "other members"), defined a well-considered and implementable set of concepts. But first prototypes of practical applications show that there are a number of issues that are not covered by the standard. This was only to be expected since the working group considered it more important to publish a base standard immediately than to delay publication in order to add further refinements. Section MISS discusses some of the concepts that the standard does not cover explicitly and explains why they are important for practical applications.

SGML and XML have DTDs defining classes of instances, but topic maps as currently specified do not have an equivalent construct. The standards working group has recognised this need and coined the term *topic map template* for the "declarative part" of a map. Section TEMPL explains what makes up a template.

Three other additional concepts are also discussed:

- a taxonomy of the basic properties of topic associations (section TRANS ),
- class (or type) hierarchies and how they can be exploited in topic map software (section CLASS ), and
- consistency checking and validity constraints for topic maps (CONSI ).

The section CONCL summarizes the paper and gives an outlook on further topic map developments.

## 2. INTRODUCTION: Topic maps, indexes, and knowledge management

The original motivation for topic maps dates back to the early days of the Davenport group in 1991 and the need to be able to merge indexes. The key insight, as Steve Newcomb, one of the original prime movers, explains

> *was that indexes, if they have any self-consistency at all, conform to models of the structure of the knowledge available in the materials that they index. But the models are implicit, and they are nowhere to be found! If such models could be captured formally, then they could guide and greatly facilitate the process of merging modelled indexes together.*

The scope was later broadened to include other forms of navigational aid – the electronic equivalents of not only printed indexes, but also tables of contents, glossaries, thesauri, cross references, etc. Common to all these applications is the attempt to provide access to information based on a model of the knowledge it contains. At the heart of that model lies the concept of the topic.

After several years of development under the auspices of CApH (Conventions for the Application of HyTime), topic maps were taken on as a work item by SC 34, the ISO committee responsible for SGML, HyTime, and related standards, and a new ISO standard (ISO/IEC 13250) was finalized in 1999 under the editorship of Michel Biezunski, Martin Bryan, and Steve Newcomb.

The topic map standard as it finally emerged defines both an abstract data model and an SGML-based serialization syntax[1] for representing knowledge structures and linking them to information resources. In order to provide maximum flexibility, the standard interchange representation is actually defined in terms of an *SGML architecture*, or “meta document type”, as specified in the HyTime standard ([ISO10744]). A topic map in its interchange form is therefore an SGML (or XML) document (or set of documents) in which different element types, derived from a base set of architectural forms, are used to represent topics, occurrences of topics, and relationships (or “associations”) between topics.

The key concepts, then, are:

- topic (and topic type)
- topic occurrence (and occurrence role)
- topic association (and association type)

Other concepts which extend the expressive power of the topic map model are those of:

- scope
- public subject
- facets

The following sections describe each of these in turn.

## 2.1. Topics and their occurrences

First of all, what is a topic?

### 2.1.1. Topics and topic types

A **topic**, in its most generic sense, can be any “thing” whatsoever – a person, an entity, a concept, really *anything* – regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever.

You can't get much more general than that!

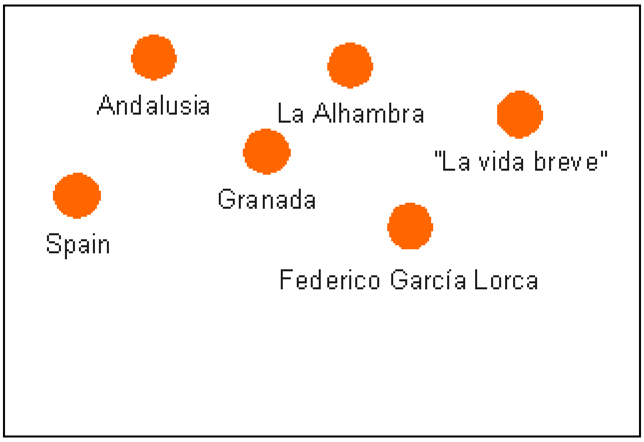
In fact, this is almost word for word how the topic map standard defines **subject**, the term used for the abstraction that the topic itself stands in for.

We might think of a “subject” as corresponding to what Plato called an *idea*. A topic, on the other hand, is like the shadow that the idea casts on the wall of Plato's cave: It is an object within a topic map that represents a subject. In the words of the standard: “The invisible heart of every topic link is the subject that its author had in mind when it was created. In some sense, a topic reifies a subject...”

Strictly speaking, the term “topic” refers to the element in the topic map document (the **topic link**) that represents the *subject* being referred to. However, in this article it will often be used more loosely to denote both of these things together. Whenever there is a need to distinguish between the two, we will use the terms “topic link” and “subject”.

So, in the context of an *encyclopaedia*, a topic might represent subjects such as “Spain”, “Andalusia”, “Granada”, “La Alhambra”, the poet “Federico García Lorca”, or a piece of music by Manuel de Falla: that is, anything that might have an entry in the encyclopaedia – but also much else besides.

*Fig. 1. Topics*



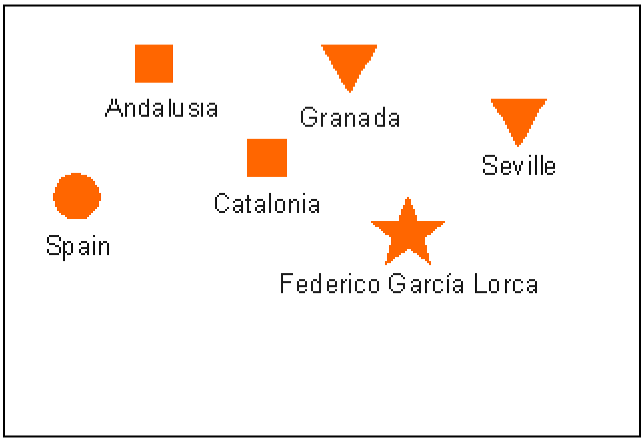

Any individual topic is an instance of zero or more **topic types**.

Thus, Spain would be a topic of type “country”, Andalusia a topic of type “region”, Granada and Sevilla topics of type “city”, García Lorca a topic of types “poet” and “playwright”, etc. In other words, topic types represent a typical *class-instance* relationship.

Exactly what one chooses to regard as topics in any particular application will vary according to the needs of the application, the nature of the information, and the uses to which the topic map will be put: In a *thesaurus*, topics would represent terms, meanings, and domains; in *software documentation* they might be functions, variables, objects, and methods; in *legal publishing*, laws, cases, courts, concepts, and commentators; in *technical documentation*, components, suppliers, procedures, error conditions, etc.

Topic types are themselves defined *as topics* by the standard. You must explicitly declare “country”, “city”, “poet”, etc. as topics in your topic map if you want to use them as types (in which case you will be able to say more about them using the topic map model itself).

*Fig. 2. Topic types*




Topics have three kinds of characteristics: names, occurrences, and roles in associations.

### 2.1.2. Topic names

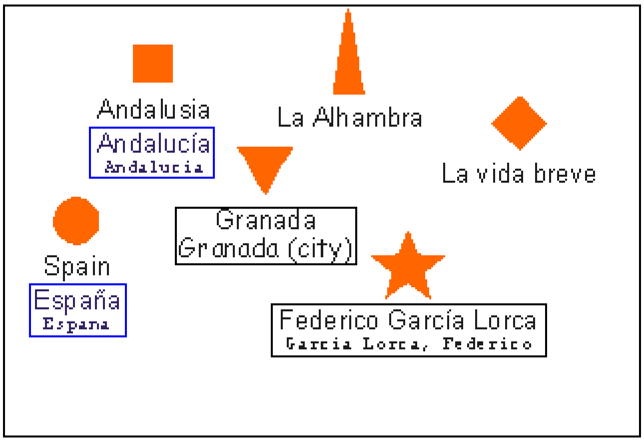
Normally topics have explicit names, since that makes them easier to talk about.[2] However, topics don't *always* have names: A simple cross reference, such as “see page 97”, is considered to be a link to a topic that has no (explicit) name.

Names exist in all shapes and forms: as formal names, symbolic names, nicknames, pet names, everyday names, login names, etc. The topic map standard doesn't pretend to try to enumerate and cover them all. Instead, it recognises the need for some forms of name (that have particularly important and universally understood semantics) to be defined in a standardised way, in order for applications to be able to do something meaningful with them, and at the same time the need for complete freedom and extensibility to be able to define application-specific name types.

The standard therefore provides an element form for **topic name**, which it allows to occur zero or more times for any given topic, and to consist of one or more of the following types of name:

- *base name* (required)
- *display name* (optional)
- *sort name* (optional)

*Fig. 3. Topic names*




The ability to be able to specify more than one topic name can be used to indicate the use of different names in different contexts or *scopes* (about which more later), such as language, style, domain, geographical area, historical period, etc. A corollary of this feature is the *topic naming constraint*, which states that no two subjects can have exactly the same name in the same scope.

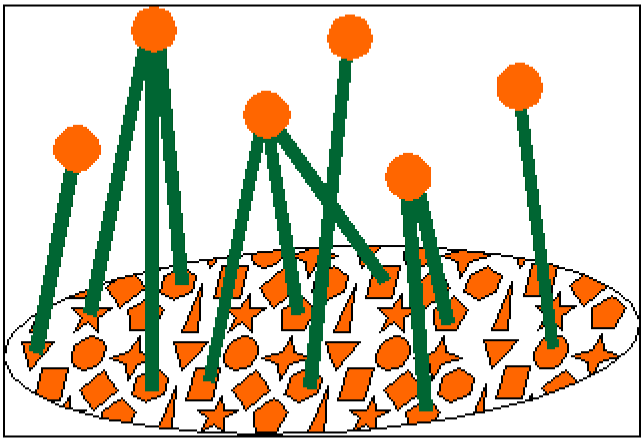
### 2.1.3. Occurrences and occurrence roles

A topic may be linked to one or more information resources that are deemed to be relevant to the topic in some way. Such resources are called **occurrences** of the topic.

An occurrence could be a monograph devoted to a particular topic, for example, or an article about the topic in an encyclopaedia; it could be a picture or video depicting the topic, a simple mention of the topic in the context of something else, a commentary on the topic (if the topic were a law, say), or any of a host of other forms in which an information resource might have some relevance to the subject in question.

Such occurrences are generally outside the topic map document itself (although some of them could be inside it), and they are "pointed at" using whatever mechanisms the system supports, typically HyTime addressing or XPointers.

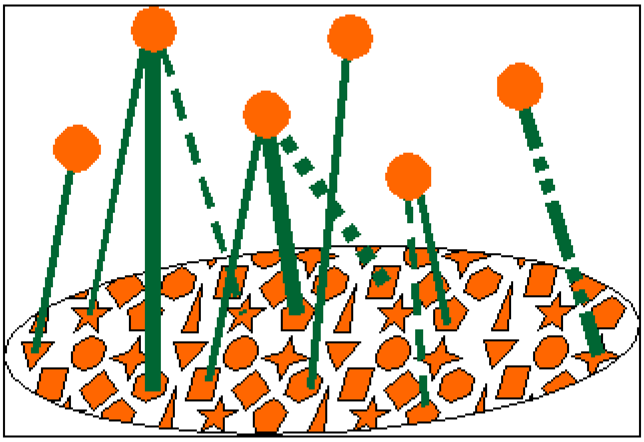
*Fig. 4. Occurrences*



An important point to note here is the *separation into two layers* of the topics and their occurrences. This separation is one of the clues to the power of topic maps and we shall return to it later.

Occurrences, as we have already seen, may be of any number of different types (we gave the examples of “monograph”, “article”, “illustration”, “mention” and “commentary” above). Such distinctions are supported in the standard by the concept of the **occurrence role**.

*Fig. 5. Occurrence roles*



As with topic types, occurrence roles are really topics, and you can therefore use the facilities of topic maps to say useful things about them (such as their names, and the relationships they partake in).

### 2.1.4. Indexes and glossaries

As described so far, topics and occurrences provide a model for explicitly stating which subjects a pool of information pertains to and how. That is basically what an index also does:

```
Andalusia          17, 77
Catalonia              72
Granada                49
Seville                22
```

But topic maps offer more. Through the concept of occurrence roles, they generalise and extend the conventions used to distinguish different kinds of references from one another. For example, in a conventional index, the page number "77" in the example above might have been set in italic, in order to indicate an illustration as opposed to a textual reference.

Some books contain more than one index (index of names, index of places, etc.). Topic types provide the same facility, but extend it in several directions to enable the creation of multiple, dynamic, user-controlled indexes organised as taxonomic hierarchies.

Glossaries can also be implemented using just the bare bones of the topic map standard that has been described so far. After all, a glossary is nothing more than a set of topic definitions, ordered by topic name:

```
España, see Spain
...
Spain: Constitutional monarchy in southern Europe...
```

The definitions are just one particular kind of occurrence (those that play the role of "definition"). With a topic map it is easy to create and maintain much more complex glossaries than this; for example, ones that use different kinds of definitions (perhaps suited to different kinds of users).

## 2.2. Topic associations

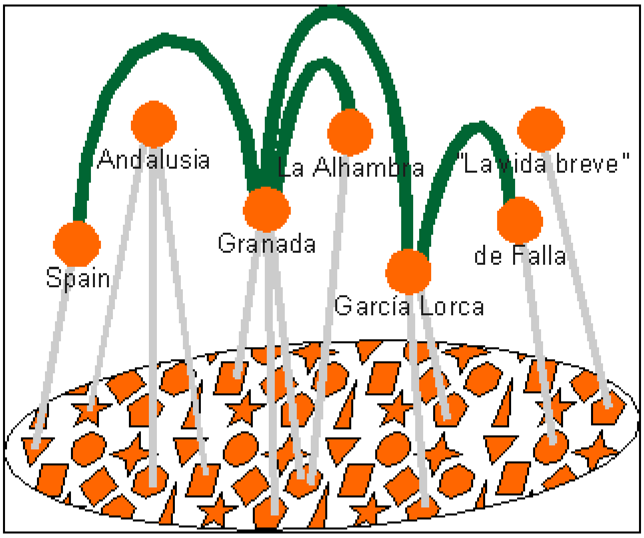
Up to now, all the constructs that have been discussed have had to do with topics as the basic organising principle for information. The concepts of "topic", "topic type", "name", "occurrence" and "occurrence role" allow us to organise our information resources according to topic, and to create simple indexes, but not much more.[3]

The really interesting thing, however, is to be able to describe *relationships* between topics, and for this the topic map standard provides a construct called the **topic association**.

A topic association is (formally) a link element that asserts a relationship between two or more topics. Examples might be as follows:

- "Andalusia *is in* Spain"
- "La Alhambra *is in* Granada"
- "García Lorca was *born in* Granada"
- "*La vida breve* was *written by* Manuel de Falla"
- "Lorca *collaborated with* de Falla"

*Fig. 6. Topic associations*



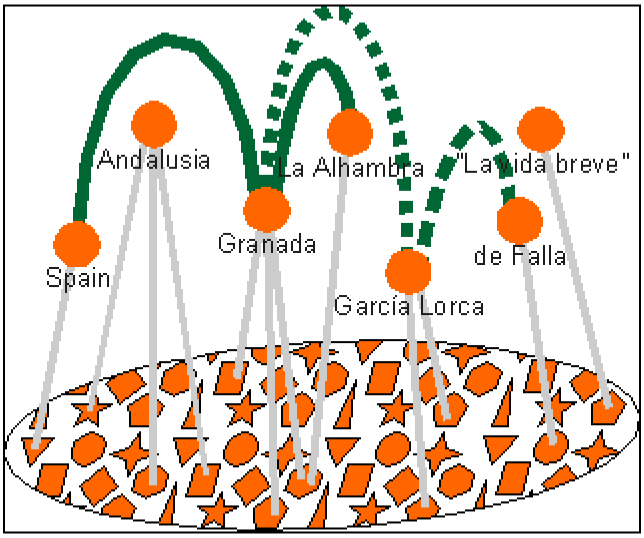

### 2.2.1. Association types

Just as topics can be grouped according to type (country, city, poet, etc.) and occurrences according to role (mention, article, commentary, etc.), so too can associations between topics be grouped according to their type. The **association type** for the relationships mentioned above are `is_in` (or geographical containment), `born_in`, `written_by`, and `collaborated_with`. As with most other constructs in the topic map standard, association types are themselves defined in terms of topics.

The ability to do typing of topic associations greatly increases the expressive power of the topic map, making it possible to group together the set of topics that have the same relationship to any given topic. This is of great importance in providing intuitive and user-friendly interfaces for navigating large pools of information.

It should be noted that topic types are regarded as a special (i.e. syntactically privileged) kind of association type; the semantics of a topic having a type (for example, of Granada being a city) could quite easily be expressed through an association (of type "instance-of") between the topic "Granada" and the topic "city". The reason for having a special construct for this kind of association is the same as the reason for having special constructs for certain kinds of names (indeed, for having a special construct for names at all): The semantics are so general and universal that it is useful to standardise them in order to maximise interoperability between systems that support topic maps.

*Fig. 7. Association types*



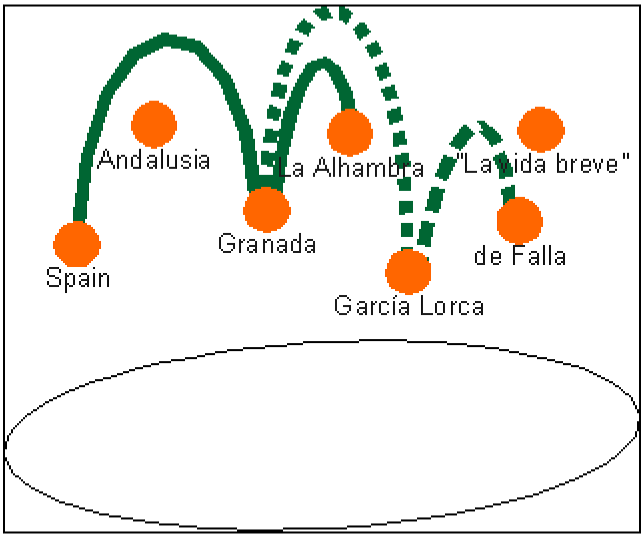

It is also important to note that while both topic associations and normal cross references are hyperlinks, they are very different creatures: In a cross reference, the anchors (or end points) of the hyperlink occur *within the information resources* (although the link itself might be outside them); with topic associations, we are talking about links (between topics) that are *completely independent* of whatever information resources may or may not exist or be considered as occurrences of those topics.

Why is this important?

Because it means that topic maps are information assets in their own right, irrespective of whether they are actually connected to any information resources or not. The knowledge that Granada is in Andalusia, that *La vida breve* was written by de Falla and is set in Granada, etc. etc. is useful and valuable, whether or not we have information resources that actually pertain to any of these topics.

Also, because of the separation between the information resources and the topic map, the same topic map can be overlaid on different pools of information, just as different topic maps can be overlaid on the same pool of information to provide different "views" to different users. Furthermore, this separation provides the potential to be able to interchange topic maps among publishers and to merge one or more topic maps.[4]

*Fig. 8. Topic maps as portable semantic networks*




### 2.2.2. Association roles

Each topic that participates in an association has a corresponding **association role** which states the role played by the topic in the association. In the case of the relationship "García Lorca was born in Granada", expressed by the association between García Lorca and Granada, those roles might be "person" and "birthplace"; for "*La vida breve* was written by Manuel de Falla" they might be "opera" and "composer". It will come as no surprise now to learn that also association roles are regarded as topics in the topic map standard!

Another aspect of topic associations that is worth noting, is that they are not one-way. The `born_in` relationship between García Lorca and Granada implies what might be called a `fostered_by` relationship between the province and the poet ("Granada fostered García Lorca"), and the `written_by` relationship between *La vida breve* and de Falla is also a `composed` relationship between the composer and his opera ("de Falla composed *La vida breve*").

Sometimes associations are "symmetrical", in the sense that the nature of the relationship is the same whichever way you look at it. For example, the corollary of "Lorca collaborated with de Falla" would (presumably) be that "de Falla collaborated with Lorca". Sometimes the anchor roles in such symmetrical relationships are the same (as in this case: "collaborator" and "collaborator"), sometimes they are different (as in the case of the "husband" and "wife" roles in a "married-to" relationship).

Other association types, such as those that express superclass/subclass and some part/whole (meronymy/holonymy) relationships,[5] are transitive: If we say that Lorca is a poet, and that a poet is a writer, we have implicitly said that Lorca is a writer. Similarly, by asserting that Granada is in Andalusia, and that Andalusia is in Spain, we have automatically asserted that Granada is in Spain and any topic map-aware search engine should be able to draw the necessary conclusions without the need for making the assertion explicitly.[6]

### 2.2.3. Thesauri, semantic networks, and knowledge management

The addition of typed associations to the basic topic paradigm enables topic maps to be able to model thesauri and other networks of information and knowledge.

A thesaurus is a network of interrelated terms (along with their definitions, examples, etc.) within a particular domain. There exist various standards for thesauri ([Z3919], [ISO5964], [ISO2788]) that predefine relationship types such as "broader term", "narrower term", "used for", and "related term", all of which correspond directly to association types in a topic map. Other thesaurus constructs, such as "source", "definition", and "scope note" would be modelled as occurrence roles in a topic map.

One advantage of applying the topic map model to thesauri is that it becomes possible to create hierarchies of association types that extend the thesaurus schema without deviating from accepted standards (for example, by subclassing "used for" as "synonymous for", "abbreviation for", and "acronym for"). Further advantages would be gained from using the facilities for scoping, filtering and merging described in the next section.

"Semantic networks", "associative networks" and "knowledge" (or "conceptual") "maps" are terms used within the fields of semantics and artificial intelligence to describe various models for representing knowledge structures within a computer. Many of these already correspond closely to the topic/association model. Adding the topic/occurrence axis provides a means for "bridging the gap" between knowledge representation and the field of information management.

“Knowledge management” is one of today's buzzwords and a term that often involves not a little marketing hype. For the big consulting companies, knowledge management is essentially about new business management techniques designed to address the fact that people (and the expertise they possess) are the primary assets in an increasingly knowledge-based economy. Others equate knowledge management with information management (especially some vendors of information management tools, who are only too happy to slap a new label on their boxes).

But knowledge is fundamentally different from information: the difference is that between knowing a thing versus simply having information about it. And if, as one writer claims ([RUG97]) “knowledge management covers three main knowledge activities: generation, codification, and transfer”, then topic maps can be regarded as the standard for codification that is the necessary prerequisite for the development of tools that assist in the generation and transfer of knowledge.

## 2.3. Additional concepts

### 2.3.1. Scope

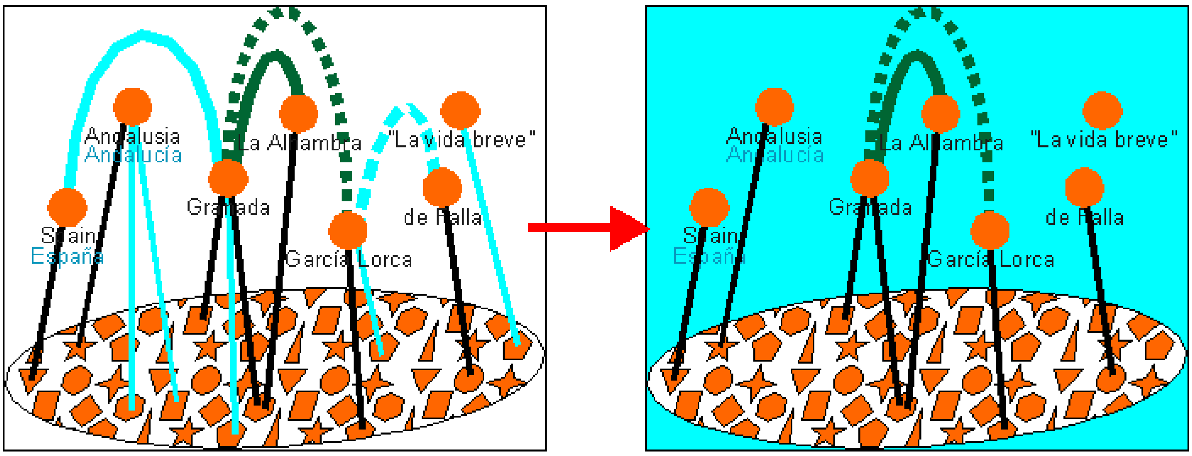
From the preceding discussion we see that topics can have various characteristics assigned to them: they can have *names*, they might have *occurrences*, and for every association in which they partake, they have a *role*. These different kinds of assertions that can be made about a topic are collectively known as **topic characteristics**.

In the topic map standard, any assignment of a characteristic to a topic, be it a name, an occurrence or a role, is considered to be valid within certain limits, which may or may not be specified explicitly. The limit of validity of such an assignment is called its **scope**, and scope – as you might expect – is defined in terms of topics.

For example, when I refer to “Granada”, it is clear that I am referring to the city in Spain. Or is it? How can someone know that I am not talking about the town of the same name in Nicaragua, or the song by Agustín Lara that Carreras sang in the first Three Tenors concert? Presumably because of the context set by my use of examples so far in this paper.

With topic maps, there is machinery for specifying that kind of scope explicitly, and also for handling situations (for example, when merging topic maps) in which the use of implicit scoping might otherwise lead to errors or ambiguities.

*Fig. 9. Scoping topic names, occurrences and associations*




Scopes applied to topic characteristics

The effect of applying scope

One part of this machinery, is the concept of the **theme**, which is defined as “a member of the set of topics used to specify a scope”. In other words, a theme is a topic that is used to limit the validity of a set of assignments. So, in a topic map where the scope was set in terms of the themes “Spain” and “popular music”, the name “Granada” could be unambiguously used to denote the song referred to above.

### 2.3.2. Public subject

Sometimes the same subject is represented by more than one topic link. This can be the case when two topic maps are merged. In such a situation it is necessary to have some way of establishing the identity between seemingly disparate topics. For example, if reference works publishers from Norway, Poland and Germany were to merge their topic maps, there would be a need to be able to assert that the topics “Spania”, “Hiszpania” and “Spanien” all refer to the same subject.

The concept that enables this is that of **public subject**, and the mechanism used is an attribute (the **identity attribute**) on the topic element. This attribute addresses an resource which identifies the subject in question as unambiguously as possible. That resource could be some official, publicly available document (for example, the ISO standard that defines 2- and 3-letter country codes), or it could simply be a definitional description within (or outside) one of the topic maps.

Any two topics that reference the same subject by means of their identity attributes are considered to be semantically equivalent to a single topic that has the union of the characteristics (the names, occurrences and associations) of both topics. In the topic map grove, a single topic node results from combining the characteristics of the two topics.[7]

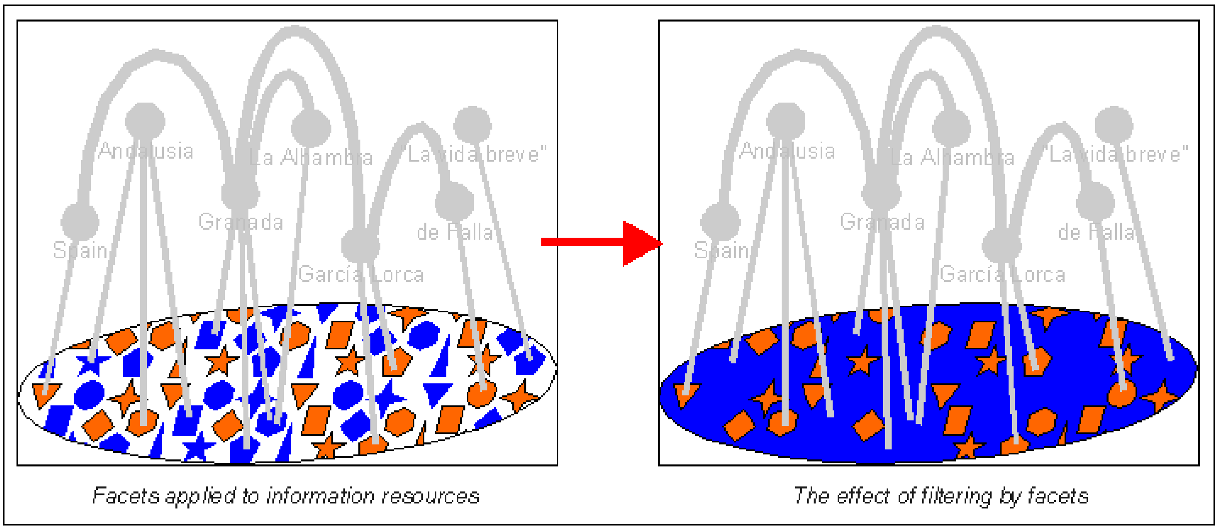
### 2.3.3. Facets

The final feature of the topic map standard to be considered in this introduction is the concept of the **facet**.

Facets basically provide a mechanism for assigning property-value pairs to information resources. A facet is simply a property; its values are called **facet values**. Facets are typically used for supplying the kind of metadata that might otherwise have been provided by SGML or XML attributes. This could include properties such as "language", "security", "applicability", "user profile", etc. Facets could also cover the kinds of properties used in faceted classification systems (hence the name); for example, typical facets within the domain of medicine might be "disease", "therapy" and "age group".[8]

Once such properties have been assigned, they can be used to create query filters producing restricted subsets of resources, for example those whose language is "Spanish" and user profile is "secondary school student". This provides a complement to scoping; whereas the latter can be seen as a filtering mechanism that is based on *properties of the topics*, facets provide for filtering based on *properties of the information resources themselves*.

*Fig. 10. Applying facets for filtering*




In a sense, facets are orthogonal to the topic map model itself (except to the extent that both facets and facet values, like most other things in the topic map standard, are regarded as topics). Despite this, facets provide a useful mechanism that complements and significantly extends the power of topic maps.

# 3. ALLEGRO: Templates, topology, and type hierarchies

This section discusses four areas in which the topic map standard could be extended in order to provide more support for the authoring process and improved inferencing facilities during navigation.

## 3.1. The missing pieces: An overview

During the years of its gestation the topic map model changed many times – from an extremely high level of generality to much more specific models designed to be used solely for navigation. The final result is – as most standards – a compromise. The working group believes that it offers an optimal balance between extreme power and flexibility on the one hand, and sufficiently well-defined semantics on the other.

The members of the working group always had in mind that the standard has to be implementable, and they tended towards a more general model because of both implementability and applicability reasons. They knew that first practical applications might uncover concepts which are not explicitly described in the standard, but they felt it was more important to have a base standard approved and published than to delay publication any longer merely to add further refinements. Adapting the standard to the XPointer (or XPath) addressing format – as soon as it becomes a W3C (Word Wide Web Consortium) recommendation – is already on the agenda of the working group.

The STEP Group[9] started investigating topic map applications in autumn 1998 in the context of reference works (especially encyclopedias and dictionaries). Applying topic maps to encyclopedias is quite natural: Topic maps model knowledge structures and lexicons represent large parts of the “knowledge” of society. Thus this application field is a perfect candidate for detecting shortcomings and finding improvements.

### 3.1.1. Separating the declarative part

Topic maps are a well-designed standard for modelling semantic information networks. It defines the basic concepts and almost everything in the map is itself a topic. Even the “objects” declaring a topic map are topics, namely themes, topic types, occurrence role types, association types, and association role types. Having such recursive declarations makes perfect sense when the goals are to limit the concepts to a sensible minimum and make topic maps self-contained and self-documenting.

But the standard does not provide a name or definition for the list of declarative “objects” of a map and this can lead to some confusion: Users often mix up “declarative” topics and “regular” topics during discussions. In addition to that, the different tasks of topic map design, creation, and maintenance are hard to distinguish and to separate. The same is true for user access rights: As long there is no distinction, different rights cannot be assigned to the map.

The separate declarative part could also be used for defining classes of topic maps that share a common set of topics for types with predefined semantics.

The standard therefore stands in need of a formally defined construct that covers the declarative part of a topic map.

### 3.1.2. Applying theoretical background

The most interesting constructs in topic maps as far as representing knowledge structures is concerned are associations. Because these are in fact relations it makes sense to take a look at mathematics and apply some of the theoretical background of relations. Furthermore the scientific fields of linguistics and philosophy may provide additional taxonomies.

The concepts that we find could lead to predefined basic association types and association properties. Neither of these are covered by the standard today, but they could offer much more precise semantics in the maps. The topic map template will be the ideal place to define them.

### 3.1.3. Class-instance relation is not enough

All topics, occurrences, and associations can be seen as instances of classes (types). The classes themselves are expressed as topics.[10]

This class-instance relationship is in fact merely a syntactically privileged association type, as the standard makes clear:

> *The class-instance relationship ... could alternatively be established by a topic association link whose semantic is the relationship between a class and an instance of that class.*

This means that the class-instance relation is an association type predefined by the standard. Any topic map software has to support it as a built-in function, e.g. by displaying the name of the referenced topic as the name of the type.

If we are looking at the class-instance relation from an object oriented view, then there is a justifiable demand for a superclass-subclass relationship as well. However, the standard explicitly declares that such a relationship has to be user-defined. Here are the relevant quotes:

> *The topic relationships established by the types attribute are not superclass-subclass relationships. They are only class-instance relationships.*
>
> *Superclass-subclass relationships between topics can be asserted by topic association links that have been user-defined for that purpose.*

STEP's experiences made with the encyclopedia applications show that the superclass-subclass relationship is a very powerful mechanism for performing inferencing, i.e. deriving implicit information about the current “object”. The implicit information can be used when querying the map or when declaring and/or checking consistency constraints. And because these features should be an integral part of a topic map software a user-defined and therefore application-specific solution is too weak.

### 3.1.4. Questions of consistency

The standard has almost nothing to say on the subject of validation and consistency. The “Conformance” section of the standard focuses on the understanding of the defined constructs, the interchange syntax, and import/export of topic maps. But nothing more, as this excerpt from the standard shows:

> *This International Standard constrains neither the uses to which topic maps can be put, nor the character of the processing that may be applied by a conforming application.*

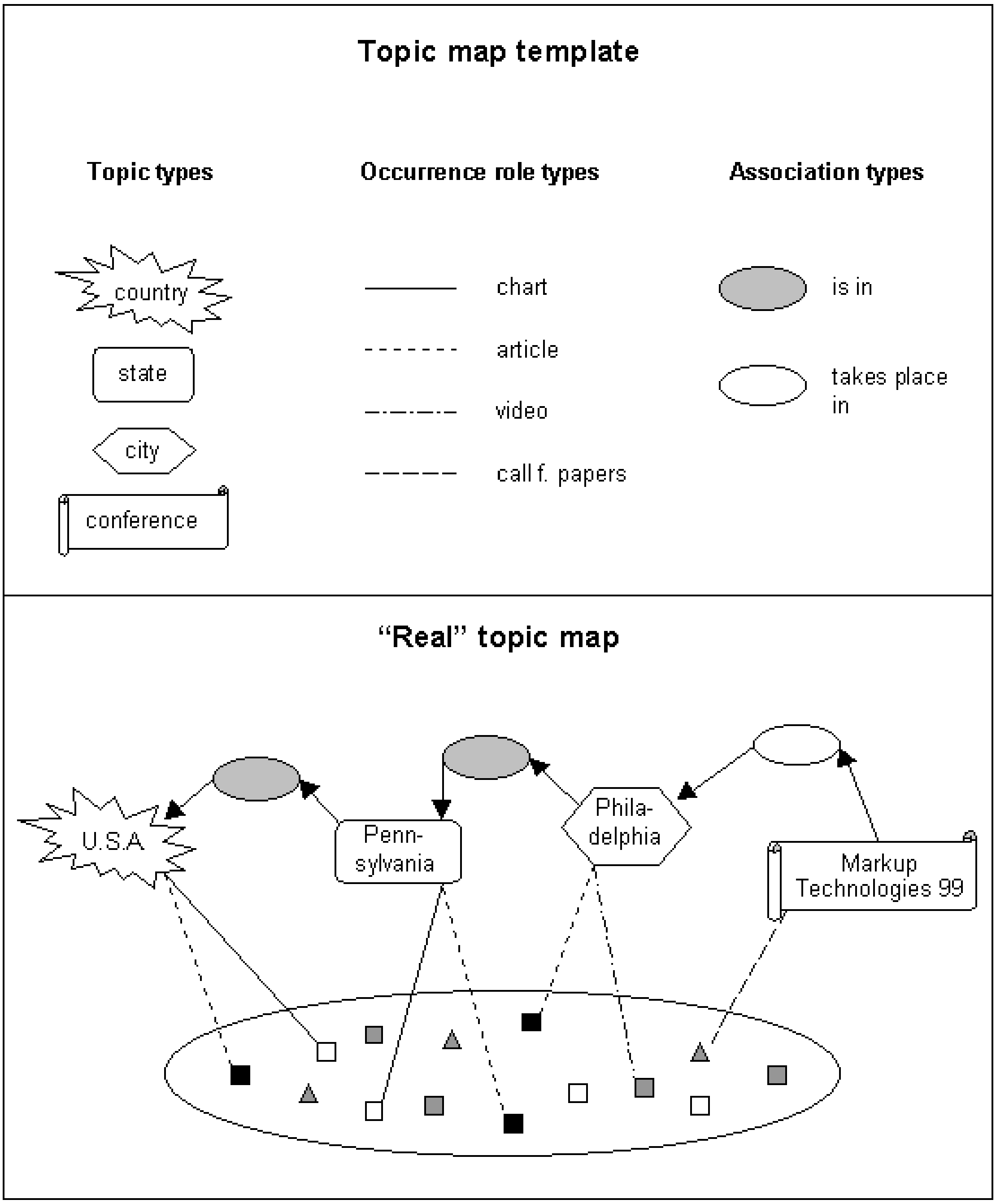
A topic map author (or authoring team) needs system support when developing a map with millions of topics and associations. The question of the consistency of the map becomes a key issue, because it is nearly impossible to check a map of that size manually.

For that reason we need concepts to declare consistency constraints and to validate that those constraints have been obeyed.

## 3.2. Topic map templates

The ISO working group has already responded to the need to be able to separate the declarative part of a topic map. It coined the term *topic map template* for a topic map that only consists of topics that are declared in order to be used as types in a class of topic maps. At the present time this term is only “semi-official”, since the concept has not yet been refined and added to the standard.

*Topic map template*




### 3.2.1. What is a topic map template?

A topic map template consists of all constructs which have a declarative meaning for the map (see figure F-TEMPL). These are all the topics used as themes and as types for

- other "regular" topics,
- occurrence roles,

- associations,
- association roles,
- facets, and
- facet values.

As we will see later, the class hierarchy information and consistency constraints will also become part of a topic map template.

The topic map designer shall mark the topics in the template for which kind of type they could be used in the "real" map. This can be done by either grouping the topics (see below TEMPL-MOD ) or by assigning attribute values. The latter approach provides more flexibility for marking topics for more than one kind of type.

In any case it is clearly important that the topics of the template can be distinguished somehow from the topics of the topic map instance(s) belonging to the class of topic maps defined by the template, and that the template becomes a "manageable" object with its own (public) identifier, owner, version number, etc.

### 3.2.2. Using templates in topic maps

The topic map template – which is a topic map – can be copied into or referenced by another topic map.

The copied template acts as a starting point for a new map containing all the themes and types which will be extended during the further development of the map.

The referenced template provides the basic themes and types which are used by the referencing map. A referenced template makes use of the merging features of topic maps defined by the standard. Thus more than one template could be referenced. Though the precondition for merging is the existence of carefully worded subject identities.

### 3.2.3. Template modules

It might be meaningful that a template consists of sub-templates to modularize the design. Candidates for template modules are

- clusters of all “typing” topics for the various “objects” as listed above, e.g. all topics which shall be used as topic types,
- the class hierarchy information, or
- the consistency constraints.

But this is only one possibility. How the declarations will be clustered in modules depends to a large degree on the application specific requirements. The only important thing is that the template can easily identified and separated from the real map.

### 3.2.4. Distributing the design and creation tasks

The design and creation of topic maps can now be split up into subtasks because of the availability of templates and template modules. Furthermore, user access rights of user groups as well as roles can be assigned.

The tasks of the designer might be:

- declaration of themes,
- declaration of all topics which are candidates for types,
- marking the topics with the kind(s) of type it is intended for,
- defining the consistency constraints.

The tasks for the editor might be:

- definition of the “real” topics,
- definition of associations between them,
- establishing the occurrence links to the relevant information objects,
- checking the consistency of the map by applying the consistency constraints (this will be an automatic process).

The assignment of facets can be seen as a completely separate task.

### 3.2.5. Role of topic map templates for ISO/IEC 13250

The concept of templates offers the ISO working group the possibility of defining various templates which are specific for different application areas. These templates would contain built-in types (i.e. topics) and association types with predefined semantics which could be supported by “template-conformant” applications.

Such templates could be published as annexes to the standard or as separate standards, as has already been done with SGML DTDs (e.g. ISO 12083).

## 3.3. Association taxonomy

The investigation of the theoretical backgrounds of relations leads us to the domains of mathematics, linguistics, artificial intelligence, and philosophy. All these scientific fields deal with knowledge representation and knowledge structures in one way or another.

We will concentrate on two issues from this broad research area: relations in mathematics (i.e. the abstract properties of associations) and relationship types in artificial intelligence and linguistics (i.e. specific classes of associations).

### 3.3.1. Association properties

The most important relations – in the mathematical sense – are the *binary relations*.[11]

**Definition:** A *binary relation between the sets A and B* is: every subset $R$ of $A \times B$ ($R \subseteq A \times B$).

The properties which are of interest for topic maps are only effective for a restricted kind of relations.

**Definition:** A *binary relation in M* is: a binary relation $R$ with $A = B = M$, thus $R \subseteq M \times M$.

A binary relation is also a binary predicate.

**Definition:** A *predicate (relation) R is fulfilled (true)* for $x \in A$ and $y \in B \Leftrightarrow (x, y) \in R$.

$(x, y) \in R$ can be abbreviated as $xRy$.

Now we can define the properties for relations in $M$.

| **Property of *R*** | **Definition** |
|---|---|
| reflexive | $\forall x \in M: xRx$ |
| symmetric | $\forall x, y \in M: xRy \Rightarrow yRx$ |
| transitive | $\forall x, y, z \in M: xRy \wedge yRz \Rightarrow xRz$ |
| anti-reflexive | $\forall x \in M: \neg xRx$ |
| anti-symmetric | $\forall x, y \in M, x \neq y: xRy \Rightarrow \neg yRx$ |
| connex | $\forall x, y \in M: xRy \vee yRx$ |

Certain combinations of these properties define special classes of relations, of which there are four:

**Definitions:**

- R is an *equivalence relation*: $R$ is reflexive, symmetric, and transitive.
- R is an *partial ordering relation*: $R$ is reflexive, anti-symmetric, and transitive.
- R is a *total order relation*: $R$ is reflexive, anti-symmetric, transitive, and connex.
- R is a *strong order relation*: $R$ is anti-reflexive, anti-symmetric, and transitive.

Some examples of specific relations will serve to illustrate the various properties and classes of relations ($M = \{0, 1, 2, 3, ...\}$).

| | relation examples | | |
|---|---|---|---|
| **Property / class** | **is denominator of** | **is less than equal** | **is less than** |
| reflexive | yes | yes | no |
| symmetric | no | no | no |
| transitive | yes | yes | yes |
| anti-reflexive | no | no | yes |
| anti-symmetric | yes | yes | yes |
| connex | no | yes | no |
| order rel. | yes | yes | no |
| total order rel. | no | yes | no |
| strong order rel. | no | no | yes |

Why is all the theory relevant for topic maps? Let us analyze the association type "geographical_object *is in* geographical_object". It is transitive, anti-reflexive, and anti-symmetric; thus it is a strong order relation. Topic map software that was aware of these facts (i.e. the properties of this prarticular association type) would be capable of automatically deriving implicit knowledge from the map.

An example: From the given associations

- Pennsylvania *is in* USA
- Philadelphia *is in* Pennsylvania
- Pittsburgh *is in* Pennsylvania

the topic map software can derive that

- Philadelphia *is in* USA
- Pittsburgh *is in* USA
- USA *is not in* Pennsylvania
- Philadelphia *is not in* Philadelphia
- etc.

It is obvious that the most informative statements of this example derive from the property of transitivity.

Another example: Let us analyze the association type "street *is parallel to* street". It is reflexive, symmetric, and transitive; thus it is an equivalence relation.

If we have the associations

- Park Avenue *is parallel to* Madison Avenue
- Madison Avenue *is parallel to* Fifth Avenue

then the associations

- Park Avenue *is parallel to* Fifth Avenue
- Fifth Avenue *is parallel to* Madison Avenue
- etc.

can easily be derived. The relevant information comes from the symmetry and again from the transitivity property.

The examples show that a simple set of association properties, i.e. the relation properties introduced above, would give more “knowledge” from the topic map than explicitly coded in it. This means that the map becomes smaller, that the effort creating a map will be minimized, that possible coding errors will be reduced tremendously, and that the inferencing capabilities of the topic map's query engine will be greatly enhanced. Furthermore the consistency checking can make use of the property information, which again improves the quality of the map.

### 3.3.2. Basic association types

The previous section introduced the basic association properties. This section investigates if also basic association types would make sense.

A lot of research has been done in the area of knowledge structures[12]. Some of the research work covers relations in the lexicon [ILE88]. Others investigated the linguistic relations in the semantic of English language [FEL98], [WORDNET]. The results are a summary of relations that express the basics concepts of knowledge representation.

A large class is comprised of the *part-whole* or *holonymy/meronymy* relations. [WCH87] and [CHW88] list six and seven subclasses of holonymy respectively:

- component-object (e.g. branch/tree)
- member-collection (e.g. tree/forest)
- portion-mass (e.g. slice/cake)
- stuff-object (e.g. aluminum/airplane)
- feature-activity (e.g. paying/shopping)
- place-area (e.g. Philadelphia/Pennsylvania)
- phase-process (e.g. adolescence/growing up)

Iris et al [ILE88] reduce this to four basic subclasses:

- functional-part (← phase-process, feature-activity)
- segmented-part (← component-object, place-area)

- collection-member (← member-collection, stuff-object)
- subset (← portion-mass)

According to [ILE88] only *segmented-part* and *subset* exhibit transitivity. Individual *functional-part* or *collection-member* relations could be transitive, but the property does not apply to these classes as a whole.

We can conclude that the *part-whole* class with its subclasses *functional-part*, *segmented-part*, *collection-member*, and *subset* shall be predefined association types – declared in a template.

Some other relevant relationship types are

- synonymy (e.g. equals, identical to),
- similarity (e.g. similar to),
- order (e.g. less than, older than, closer to),
- result-agent (e.g. “object” is caused by “agent”, “artwork” created by “artist”, “painting” painted by “painter”),
- tool-agent (e.g. “tool” is used by “agent”, “instrument” is played by “musician”), and
- strict implication[13] (e.g. “activity 1” implies “activity 2”, ”snoring” implies “sleeping”).

The *synonymy*, *order*, and *strict implication* are transitive relations. *Synonymy* and *similarity* are symmetric. For every *result-agent* and *tool-agent* relation exists an inverse one (“agent” causes “object”, “agent” uses “tool”). Strict implication is non-symmetrical: you can sleep without snoring, but you cannot snore without sleeping! All these relations are candidates to be predefined association types that are declared in a template.

The contributions from linguistics introduce further subclasses for *synonymy* relations (thesauri: [AGB97]) and build a class hierarchies with the *hyponymy* for nouns and the *troponymy* for verbs (dictionaries: [FEL98], [WORDNET]). Both *hyponymy* and *troponymy* represent the “is a” or “is a kind of” relation, which is already covered by the topic type construct. The *synonymy* subclasses seemed to be quite specific, thus there is no need to have them as predefined association types. They are in any case more appropriately handled through the use of multiple topic names.

## 3.4. Class hierarchies

The realisation of the need for class hierarchies stems from STEP's encyclopedia projects. A topic map for a lexicon contains a very large number of topics (typical orders of magnitude are hundreds of thousands or millions) and associations (even more). But most of the topic, association, and occurrence role types can be reduced to a small number of “super-types” – as we have already seen in the previous section.

### 3.4.1. Superclass-subclass

The superclass-subclass relationship of topic types, association types, and occurrence role types go hand in hand, as following examples shows:

- Topic types: (person) → (artist, ...) → (painter, sculptor, writer, poet, composer, ...); (object) → (artwork, ...) → (painting, sculpture, novel, poem, opera, ...)
- Association types and occurrence role types: (object "was caused by" person) → (artwork "was created by" artist) → (opera "was composed by" composer)

### 3.4.2. Class hierarchy and association type properties

The class hierarchies become even more important when the end-user navigates or queries the map. If someone would like to know "Which pieces of music were composed by Germans that were influenced by W.A. Mozart?", it is very likely that this information is not exactly part of the map. But with just a few topics, transitive associations, and a class hierarchy the answer can be found very easily.

The facts of the map:

- **The topic type (class) hierarchies:** person → composer; piece of music → opera; geographical object → country; geographical object → city.
- **The transitive association type:** "geographical object" is in "geographical object".
- **Other association types:** "composer" has composed "piece of music"; "person" was influenced by "person"; "person" was born in "geographical object".
- **The topics:** W.A. Mozart (composer); R. Wagner (composer); L. van Beethoven (composer); Bonn (city); Leipzig (city); Germany (country); Lohengrin (opera).
- **The associations:** Bonn is in Germany; Leipzig is in Germany; L. van Beethoven was born in Bonn; R. Wagner was born in Leipzig; Lohengrin was composed by R. Wagner; R. Wagner was influenced by W.A. Mozart.

The algorithm how the topic map software would find the solution with these facts could work as follows:

- Extraction of the known topics from the query: Germany, W.A. Mozart.
- Extraction of the types of the unknown topics: person ($X$), piece of music ($Y$).
- Extraction of the association types: born in, influenced by, composed by.
- Finding the missing topics using the associations and class hierarchies:

  $X$ is born in Germany (country is also a geographical object) ⇒ $X$ is born in Bonn or Leipzig (both cities are in Germany) ⇒ $X$ is L. van Beethoven or R. Wagner (both composers are also persons);

  $X$ was influenced by W.A. Mozart (composer is also a person) ⇒ R. Wagner was influenced by W.A. Mozart (both composers are also persons) ⇒ $X$ is R. Wagner;

  $Y$ was composed by $X$ ⇒ $Y$ was composed by R. Wagner ⇒ Lohengrin was composed by R. Wagner (opera is also piece of music) ⇒ $Y$ is Lohengrin.

This very simple example shows the power of combining class hierarchies with properties of association types (here transitivity). As already stated above, both class hierarchies and association type properties are the basis for compact topic maps, minimized creation and maintenance efforts, and a reduction of coding errors.

This supports our contention that the concept of class hierarchies should be a predefined association type of topic map template ensuring the correct built-in interpretation by the topic map software.

## 3.5. Validation of consistency

All the previously introduced concepts extend topic maps in ways that increase their expressive power and ease creation and maintenance efforts. In addition to this, the topic map developer wants to have something at hand to help ensure the quality of the map. The information provided by a topic map based on the standard architecture is not enough – the developer asks for validation concepts.

Real life topic maps will consist of millions of topics and associations. Checking a map of such a size manually is clearly impossible, and yet checking is absolutely necessary for both proof-reading and quality assurance. It is obvious that both the designer and the editor need access to an automatic process that can validate a topic map against a set of consistency rules.

The validation is the task of the topic map development environment (e.g. an editorial system). It should be performed permanently or on demand – like structure validation against the DTD in an SGML/XML editor.

The standard has almost nothing to say on the subject of validation and consistency. The “Conformance” section of the standard focuses on the understanding of the defined constructs, the interchange syntax, and import/export of topic maps. But nothing more, as this excerpt from the standard shows:

> *This International Standard constrains neither the uses to which topic maps can be put, nor the character of the processing that may be applied by a conforming application.*

This shows that we have to develop a schema language for the definition of the consistency constraints.

### 3.5.1. Consistency constraints

The topic map standard provides the architectural element types which can be used in a derived DTD (Document Type Definition). However, the degree to which semantics can be modelled in a DTD and through content models is rather limited. A topic map will consist of a large number of “independent” elements which are connected by links and not by element structures.

Consequently a separate schema is needed which contains all the information necessary for the validation process. We call this construct *consistency constraints* or just *constraints*. The constraints are a set of predefined association types declared in the template.

### 3.5.2. What should be validated?

Constraints may be assigned to three potential layers:

- topic map modeling,
- user interface for topic maps, and
- operations on the map.

Here, we focus on the topic map modeling layer.

#### 3.5.2.1. Associations

The most important candidates for validation are the associations. This is obvious because they are the key concept and carry a large number of parameters which might be "misused".

The starting point is the association type. This controls which association role types can be combined. Beside the possible combination(s) the number of the various roles within these combinations might be of interest.

The association role type in turn governs the set of topic types which may be referenced.

It is necessary that the constraint schema brings the association type, the role type, and the topic type into a meaningful combination.

An example:

| | |
|---|---|
| Association type | ***is in*** (geographical containment) |
| Valid association role types | one **containee:** one *container* |
| Valid topic type combinations | **city:** *(country \| state \| county)*<br>**county:** *(state \| country)*<br>**state:** *(country)* |

#### 3.5.2.2. Occurrences

The assignment of the proper information resource types – if type information is provided by the editorial system – to the occurrence role types is also of interest as well as the meaningful combination of topic types and occurrence role types.

An example:

| | |
|---|---|
| Topic type: | *person* |
| Valid occurrence role types: | *biography*, *portrait* |
| Valid resource types for *biography:* | SGML/XML instance with public identifier "-//STEP//DTD biography//EN" |
| Valid resource types for *portrait:* | object types TIFF, GIF, JPEG |

### 3.5.2.3. Scopes

Furthermore the correct use of scopes and especially the combination of different scopes might be checked.

The topic type could restrict the possible scopes for the topics, their topic names, base name, display name, sort name, and their occurrences.[14]

The association types might restrict the meaningful scopes for the associations also. The combination of the meaningful scopes of the association and the referenced topics should be checked also because the association type is closely related to the possible types of the referenced topics.

An example:

| | |
|---|---|
| Themes: | *before Einstein's theory of relativity*, *after Einstein's theory of relativity* |
| Topic types: | *physical law*, *mathematical axiom* |
| Occurrence role types: | *definition* |
| Constraints: | The scope *before Einstein's theory of relativity* might be used for occurrences with role *definition* for topics of type *physical law*; but it must not be used for *definitions* of *mathematical axioms*. |

### 3.5.2.4. Topic names

For reasons of completeness checking of the topic names should also be possible. Topic names might be checked against text patterns or against database entries. The constraints will be governed by the topic type in question.

An example:

| | |
|---|---|
| Topic types: | *component in assembly group*, *chemical substance* |
| Constraints: | Check base name of topic of type *component* against pattern (regular expression) "P[0-9]+[A-D][E-G][0-5]"; check sort name of *chemical substance* against table "substance names" in chemical database. |

All type combination constraints might limit the number of superclasses and/or subclasses of the affected types.

# 4. Conclusions

The paper provided an extensive introduction to the new standard ISO/IEC 13250 Topic Maps. The standard defines a model and architecture for the semantic structuring of link networks. It can be seen as a base technology for modeling knowledge structures. The standards working group defined topic maps in such a way that a limited but implementable set of core concepts express the necessary semantics.

The STEP Group investigated how topic maps can be applied to reference works and uncovered some concepts which are not made explicit in the standard:

- possibility to separate the declarative part from the “real” map,
- predefined association types and association type properties,
- class hierarchies for types, and
- consistency constraints as input to map validation.

The paper explained these concepts and presented meaningful solutions.

First experiences have shown that the part of a topic map made up by all topics used as themes and types by other "objects" in the map should be clustered somehow. For this purpose the term "topic map template" was coined by the ISO working group. Templates can be used as starting points for new maps or can be used by reference in order to provide all the themes and types the map needs. Standardizing topic map templates will offer base topic maps for specific application areas and could form the basis of semantic application profiles.

We looked at related academic fields like mathematics, linguistics, and philosophy to get some substantial input about relations. The results are a list of association type properties which give important hints to the topic map software and a list of basic association types which could act as built-in superclasses.

The introduction of the superclass-subclass relationship was the logical consequence.

Another technical issue covered by the paper is the validation problem. Topic maps might become rather big with millions of topics, occurrences, and associations. Manual consistency checking will be impossible. All the previously defined concepts open the possibility for sophisticated rule-based validation of topic maps. The proposed consistency constraints are those rules which declare the semantics not expressible with DTDs and which control the validation process.

A couple of examples proved that standardizing the missing concepts as predefined topic map templates will help both the topic map developer and the topic map user. The improvements were presented on a level that they can be used as input to the ISO working group for further discussions.

---


## Acknowledgements

The authors would like to thank their colleagues Geir Ove Grønmo (STEP Infotek, Norway), Rafal Ksiezyk (STEP Poland), and Graham Moore (STEP-DPSL, UK), as well as all the members of STEP's Reference Works Module Club – leading European reference works publishers – for their input and open discussions about topic maps.


---

## Bibliography


**[AiGiBa97]** Aitchison, J., Gilchrist, A., Bawden, D.: *Thesaurus construction and use – a practical manual*, 3rd edition, London: Aslib, 1997.

**[ChHeWi88]** Chaffin, R.; Hermann, D.J.; Winston, M.: *An empirical taxonomy of part-whole relations: Effects of part-whole relation type on relation identification*, Language and Cognitive Process 3, 1988.

**[Fel98]** Fellbaum, C. (ed): *WordNet – An Electronic Lexical Database* , MIT Press, 1998.

**[ISO13250]** International Organization for Standardization, *ISO/IEC 13250, Information technology – SGML Applications – Topic Maps*, Geneva: ISO, [forthcoming].

**[ISO10744]** International Organization for Standardization, *ISO/IEC 10744:1999 Information technology – Hypermedia/Time-based Structuring Language (HyTime)*, Geneva: ISO, 1997.

**[ISO2788]** International Organization for Standardization, *ISO 2788:1986. Guidelines for the establishment and development of monolingual thesauri*, Geneva: ISO, 1986.

**[ISO5964]** International Organization for Standardization, *ISO 5964:1985. Guidelines for the establishment and development of multilingual thesauri*, Geneva: ISO, 1985.

**[IrLiEv88]** Iris, M., Litowitz. B., Evens, M.: *Problems of the part-whole relation*, in: Evens, M. (ed): *Relational models of the lexicon*, Cambridge, 1988.

**[Ksi99]** Ksiezyk, R.: *Trying not to get lost with a Topic Map*, in: Proceedings of XML Europe 99 Conference, GCA, Alexandria, VA, 1999.

**[Meg98]** Megginson, D.: *Structuring XML Documents*, Prentice Hall, 1998.

**[Pep99a]** Pepper, S.: *Euler, Topic Maps, and Revolution*, in: Proceedings of XML Europe 99

Conference, GCA, Alexandria, VA, 1999.

**[Pep99b]** Pepper, S: "Navigating Haystacks, Discovering Needles", *Markup Languages* [forthcoming].

**[Ran67]** Ranganathan, S.R., *Prolegomena to Library Classification*, Bombay: Asia Publishing House, 1967.

**[RaPe00]** Rath, H.H. and Pepper, S.: *Topic maps at work*, to be published in: Goldfarb, C.F., Prescod, P. (eds): *XML Handbook*, 2nd edition, Prentice Hall, 2000.

**[Rat99]** Rath, H.H.: *Technical Issues on Topic Maps*, in: Proceedings of Metastructures 99 Conference, GCA, Alexandria, VA, 1999.

**[RiDu88]** Ringland, G.A. and Duce, D.A.: *Approaches to Knowledge Representation: An Introduction*, Research Studies Press/John Wiley 1988

**[Rug97]** Ruggles, Rudy L., ed. *Knowledge management tools*, Boston: Butterworth-Heinemann, 1997.

**[Str99]** Streich, Robert, "Techniques for managing collections of interrelated text modules", *Markup Languages*, Vol.1:2, Cambridge: MIT Press 1999.

**[Vic60]** Vickery, B. C., *Faceted classification: a guide to construction and use of special schemes*, London: Aslib, 1960.

**[Vic66]** Vickery, B. C., *Faceted classification schemes*, New Brunswick: Rutgers, 1966.

**[WiChHe87]** Winston, M.E., Chaffin R., Hermann, D.: *A taxonomy of part-whole relations*, Cognitive Science 11, 1987.

**[WordNet]** *WordNet – a Lexical Database for English*, Cognitive Science Laboratory, Princeton University, Princeton, NJ, http://www.cogsci.princeton.edu/~wn/.

**[Z3919]** ANSI/NISO, *Z39.19. Guidelines for the construction, format and management of monolingual thesauri*, Bethesda: ANSI/NISO 1993.

[1] An XML-based serialization syntax will be defined once the W3C's recommendations for XML-based linking and addressing have been finalised.

[2] It should be clear that the preceding paragraphs would have been rather more difficult to understand if we hadn't given names to our topics and topic types!

[3] The principle exception to this statement is the topic type, as we shall see shortly.

[4] However, in order to be able to merge topic maps successfully, the additional concepts of *scope* and *public subject* are required. These are discussed below.

[5] For a discussion of the various kinds of part/whole relationship and their properties, see [ILE88].

[6]The current version of the topic map standard does not have “built in” support for expressing transitivity, but this would not prevent applications from providing such capabilities.

[7]Of course, the fact that the identity attributes of two topics are not identical is not sufficient to prove that the topics do not refer to the same subject; the only thing that can be proven is that there *is* identity, not that there *is not* identity.

[8]For a short description of faceted classification see the article by Bob Streich in an earlier issue of this journal ([STR99]). For more detailed expositions, see [RAN67], [VIC60] and [VIC66].

[9]The STEP Group consists of STEP Electronic Publishing Solutions GmbH (Rimpar, Germany), STEP Infotek AS (Oslo, Norway), STEP Electronic Publishing Kft (Budapest, Hungary), STEP Poland Ltd. (Warsaw, Poland), and STEP-DPSL Ltd. (Swindon, UK).

[10]NB: The recursion “a topic has a type which is a topic which has a type” stops if no type is assigned. This is possible because the type is an optional attribute of the topic, occurrence, and association. If the attribute is not specified, the meaning is that the “object” has no more specific type (i.e. belongs to no more specific class) than that of the base class to which it belongs (“topic”, “occurrence”, or “association”, respectively).

[11]N-ary relations and “elementary associations” (in which the number of arguments cannot be further reduced) with more than two arguments are not covered in this paper, because they form a more complex class.

[12]See [RIDU88] for an introduction and extensive bibliography.

[13]Definition of *strict implication*: A proposition $P$ entails a proposition $Q$ ($P \Rightarrow Q$) if and only if there is no conceivable state of affairs that could make $P$ true and $Q$ false.

[14]Because assigning scopes to the topic or the topic name are just shortcuts for assignments to every name or occurrence, the set of scopes of the topic must be a superset of the scopes for the names and occurrences, and the set of scopes of the topic name must be a superset of the scopes for the individual names.