Lars Johnsen  

**Topic Maps**  
From Information to Discourse Architecture  

**Abstract**  
  
Topic Maps is a standards-based technology and model for organizing and integrating digital information in a range of applications and domains. Drawing on notions adapted from current discourse theory, this article focuses on the communicative, or explanatory, potential of topic maps. It is demonstrated that topic maps may be structured in ways that are “text-like” in character and, therefore, conducive to more expository or discursive forms of machine-readable information architecture. More specifically, it is exemplified how a certain measure of “texture”, i.e. textual cohesion and coherence, may be built into topic maps. Further, it is argued that the capability to represent and organize discourse structure may prove useful, if not essential, in systems and services associated with the emerging Socio-Semantic Web. As an example, it is illustrated how topic maps may be put to use within an area such as distributed semantic micro-blogging.  

**Topic Maps as Information Architecture**  
  
Topic Maps (in upper case) is a standards-based technology for connecting knowledge structures to information resources. Topic maps (in lower case) are concrete manifestations of this technology: digital collections of topics representing and connecting things, or “subjects”, in some universe of discourse (persons, events, concepts, documents, web pages, etc.). Topic maps are often described as a kind of superimposed semantic metadata layer for indexing (often dispersed and heterogeneous) information resources but topic maps may in fact realize a number of organization schemes ranging from simple taxonomies to semantically rich ontologies (Garshol 2004). Topic maps are supported by various query languages, exchange formats and development and publication tools and are increasingly used in web sites, knowledge portals, content management systems and social bookmarking services (Lachica & Karabeg 2008; Garshol 2008; Pepper 2010).  
  
Examples of real world applications are the city of Bergen’s portal (www.bergen.kommune.no), VIMU, a website on Danish-German border
history (www.vimu.info) or fuzzzy.com, a social bookmarking site (fuzzzy.com).

In a topic map, topics may be given one or more alts; they may be categorized in types, subtypes and instances and they may be connected to internal content (information within the topic map itself – descriptions, data values, etc.) as well as external content (resources outside the topic map itself – web pages, files, etc.). Topics may be related in typed associations in which they are assigned semantic roles and they may be linked to external descriptors, for instance Wikipedia entries, to make their meaning, or subject identity, more transparent.

These external descriptors, also known as subject indicators, are accessed through subject identifiers, usually URL’s. Subject identifiers are central to the Topic Maps paradigm because they facilitate the merging of topics sharing one or more subject identifiers, and hence the integration of disparate topic maps. The use of stable, publicly available subject indicators and identifiers, so-called PSI’s and PSID’s, is strongly recommended in the Topic Maps community as the key to more reliable data integration and knowledge federation (Pepper 2003).

For example, in a knowledge portal of ancient Roman history, the subjects of “Brutus” and “Caesar” might be introduced as topics, alted and categorized as “Roman” (in itself a topic) using the “is-a” association. The topic of “Roman” might in turn be defined to be a subtype of “person” by means of the “a-kind-of” association. Some internal content could be attached to the topic of Brutus to describe him, for instance the years of his birth and death. A pointer to a picture of him, i.e. content external to the topic map, might also be provided to enrich the description. And to indicate more precisely which Brutus is being referred to, his full alt “Marcus Junius Brutus” could be given along with a subject identifier pointing to his entry in Wikipedia. Finally, an association might state that Brutus killed Caesar with a dagger. In this association Caesar would play the semantic role of “victim”, Brutus the role of “murderer” and the dagger the role of “instrument”.

In such a history portal, the information structure would be centered around domain entities or so-called non-addressable subjects (Brutus, Caesar, being Roman, dagger, etc.). However, topic maps may also be employed to create information architecture comprising digital content items, or addressable subjects. For example, a topic map might be employed in an e-learning system to organize distributed learning resources on the web. Here the individual topics would represent digital “learning objects” like articles, video lectures or slides. The topics would, therefore, not have subject identifiers but subject locators, pointers to the actual content items, in effect their web
Linking of topics representing both addressable and non-addressable subjects is also possible in a topic map. This may, for instance, be the case in a topic maps-based bookmarking service where web users can collectively store bookmarks and links. Users and bookmarks would constitute non-addressable subjects, while the actual web content items being referred to would be addressable ones.

Generally, topic maps used for information architecture may be visualized in the following way (fig. 1):

As Fig. 1 shows, an information architecture based on topic maps may be said to have (at least) two layers: a knowledge layer representing the objects in the domain being described and a content layer holding information about these objects. A third layer, an identity layer, may also be recognized as a separate component if the topic map contains references to subject indicators.

**Topic Maps as Exposition Space**

Topic Maps provides a flexible model for information architecture as it
allows concepts and content alike to be organized in hierarchical as well as associative structures. However, it may be argued that information in topic maps may also be structured in ways that are more discursive or “text-like” in character and, therefore, it is hoped, conducive to more expository, explanatory or possibly rhetorical forms of machine-readable information architecture. In other words, topic maps may not only be designed and utilized as tools for organizing concepts and content and providing statements about them but also, to some extent, for organizing those statements in meaningful and coherent structures. Topic maps may even, as will be exemplified below, be employed to organize and integrate certain forms of distributed discourse as these increasingly occur in social media and similar places on the web.

In fact, a new model, or metaphor, may be introduced for thinking about the construction of “expository topic maps”. Here topic maps are conceived of as a kind of three-dimensional information space built along three axes. The first axis represents the act or result of sorting out which lets topic maps authors classify subjects into types, subtypes and instances. This axis may be said to go from general to specific. The second axis constitutes the act or result of describing subjects, or, in Topic Maps parlance, assigning characteristics to topics (i.e. adding content to topics or relating topics in associations).
These descriptions may range from lean to rich. The third axis embodies the act or result of detailing, or enriching, the descriptions, adding new layers of detail to already existing information structures in the topic map. Thus a description in a topic map may either be relatively shallow or deep depending on how much it has been elaborated upon. In addition to these axes which make up the three dimensions of the information space, the notion of contextualization is introduced to capture the act or result of scoping information in the topic map. Scoping is the assignment of valid contexts to selected statements or descriptions in a topic map and may, to a certain extent, be likened to the application of metadata.

Detailing within topic maps primarily takes place through reification, the “act of making a topic represent the subject of another topic map construct in the same topic map” (Garshol & Moore 2008). The mechanism of reification is important in the Topic Maps paradigm as it allows topic map creators to “attach additional information to topic map constructs” (ibid.) or in Topic Maps terminology – to assign characteristics to reified topics.

So, for example, a picture attached to the topic of Brutus as external content in the history portal may itself be reified as an addressable subject and described in further detail. And the “killing association” relating Brutus, Caesar and the dagger could be reified and alt’d “The Assassination of Caesar”. It might even get its own subject identifier and subject indicator: http://en.wikipedia.org/wiki/Assassination_of_Julius_Caesar.

In one sense, detailing (through reification) is really creating discourse in topic maps, i.e. a way of “talking” about the world and not just storing organized information about it. Interestingly, the term discourse does in fact occur in the Topic Maps Data Model, the standard defining key concepts in Topic Maps. Here subjects are described as “anything about which the creator of a topic map chooses to discourse” (Garshol & Moore 2008) and topics are defined as representing “subjects of discourse” (ibid.).

Below an attempt is made to explore the relationship between topic map structure and discourse and more narrowly between topic map structure and texture that which makes sentences and clauses “hang together” in discourse. It will be demonstrated that reification, and the subsequent assignment of characteristics to topics and scopes to characteristics, provide a flexible and useful vehicle for building a certain measure of texture into topic maps.

The Notions of Discourse and Texture

The notion of texture is recognized in most theories of discourse even
though terminology varies somewhat. In Renkema (2009), for example, texture is described as an inherent property of discourse ensuring the “information flow” that we normally associate with running text or fluent speech. Texture covers the notions of cohesion, linguistic signals or markers of connectedness, and coherence, semantic unity as perceived by readers or listeners. So, most verbal or written texts, ranging from entire books to extremely brief and compact messages like tweets on Twitter, the social networking and micro-blogging service, have texture.

Central to Renkema’s analysis of texture, and discourse structure in general, are two general principles, the discursive principle and the dialogic principle, and three types of discourse relations, i.e. links between sentences or clauses, altly conjunction, adjunction and interjunction.

Put somewhat simply, the discursive principle states that discourse may be seen as a kind of expansion of sentence or clause constituents into new sentences and clauses. The principle may be illustrated by examples like these:

a. Then Brutus jumped on Caesar with a dagger. He had been waiting ...

b. Then Brutus jumped on Caesar with a dagger. The weapon had been hidden ...

c. Then Brutus jumped on Caesar with a dagger. The attack was sudden and fierce ...

d. Then Brutus jumped on Caesar with a dagger. Brutus loved Caesar. But he loved Rome more.

The examples show that a speaker or writer may choose to expand almost any constituent in a clause or sentence. In (a) and (d) it is the agent of the action, in grammatical terms the subject, that is being elaborated on, in (b) it is the instrument used in the action while in (c) it is the entire action itself which becomes the subject of further discourse.

The dialogic principle views discourse and text as a kind of imaginary dialogue between the speaker or writer and the addressee. In this way, continuations like the ones in (a) to (d) may be interpreted as a speaker’s or writer’s responses to an addressee’s hypothetical questions or requests. In (a) the question might be formulated something like “How did Brutus succeed in staging the attack on Caesar?” and in (b) it might be “How was it possible for Brutus to produce a dagger in those circumstances?” A relevant request
by the addressee in (c) might be “Please describe the attack” while one might imagine a question like “OK, so Brutus loved Caesar. But why did he want to kill him then?” to follow the second sentence in (d).

As for the three types of relations between discourse segments, conjunction comprises the “tangible” relations that link discourse spans, the glue, so to speak, between sentences and clauses. Highly frequent examples of conjunction are anaphora, the use of pronouns, and lexical cohesion, the use of semantically related words such as synonyms or hyponyms. Anaphora is exemplified by the link between “Brutus” and “He” in (a) while (b) establishes lexical cohesion through the close semantic relationship between “dagger” and “weapon”.

Adjunction, on the other hand, is the set of semantic relations which may be identified between clauses and sentences. These may, according to Renkema, be divided into three main types, altly elaboration, enhancement, and extension (which may further be divided into a number of subclasses):

- Elaboration refers to cases where a speaker or writer wishes to provide more information on an entity (thing, concept, person, etc.) already introduced. Instances of this type are examples a) and b) above. Here “Brutus” and “a dagger”, already mentioned, are elaborated upon.

- Enhancement comprises discourse relations in which additional information is provided not only about one or more entities but about an entire situation or state of affairs. An example of this is c) in which the whole event of Brutus attacking Caesar with a dagger is described in more detail. In general, enhancement serves to indicate semantic aspects such as the manner, time or cause of an event or incident.

- Extension refers to cases where two events or state of affairs are related, contrasted or sequenced as is the case in d) where a comparison of two states is clearly intended.

While adjunction relations provide detail about entities, events or state of affairs in the domain of discourse, relations belonging to the class of interjunction are connections that carry some kind of communicative intent aimed at affecting the addressee’s beliefs, views or knowledge in a particular way. Take two examples like:
e. Brutus killed Caesar. That was terrible thing to do.

f. Brutus plotted against Caesar. He lied to him. And finally he killed him with a dagger. Brutus was indeed no nobleman.

In example (e) the second sentence is not just an informational enhancement of the propositional content of the first but an attempt on the part of the speaker or writer to convince or influence the addressee. And in (f) the first three sentences function as a kind of evidence provided by the sender to make his or her message in the last sentence more credible. Renkema (2009, p. 53) himself sums up the difference between conjunction, adjunction and interjunction in the following way:

- Conjunction: linking form to form
- Adjunction: linking information to information
- Interjunction: linking addresser to addressee

Discourse, Texture and Topic Maps

At a very general level one may claim that reification in topic maps, and the subsequent assignment of scopes to characteristics and characteristics to topics, serve to “expand or elaborate” (discursive principle) and to answer hypothetical questions or meet requests for more information (dialogic principle). More interestingly, perhaps, it may be argued that reification and assignments of characteristics and scope provide a vehicle for building texture into topic maps by allowing topic maps authors to create what we might, adapting Renkema’s original concepts slightly, call conjunction, adjunction and interjunction. This in effect means that topic map creators not only have the means available to organize concepts (and content) and construct “semantic” statements about them but also combine these statements in meaningful text-like structures.

Some examples may serve to illustrate the manifestation of conjunction, adjunction and interjunction in topic maps: Reification facilitates conjunction, the linking of form to form, or cohesion in a topic map. Consider again the association stating that Caesar was murdered by Brutus with a dagger. The dagger, playing the role of instrument in the association, may be reified as a topic entitled “the murder weapon” or words to that effect. This new topic may be taken up at a later stage, described in more detail, or connected to other topics in the topic map. For instance, a link
may be established to the topic representing the person who provided the dagger or concealed it. The important thing is, however, that the topic of “the murder weapon” can always be traced back to the topic map construct from which it originated, the association denoting the killing of Caesar. In an information architecture context this may in itself prove useful because a broader, and possibly more useful, search term is now supplied and attached, albeit indirectly, to the central statement about Caesar’s murder.

The reification of the dagger used in the killing of Caesar and its subsequent use in a new association also illustrates adjunction (linking information to information), and more precisely elaboration, in a topic map. Here a role playing topic is reified and elaborated upon elsewhere. But elaboration is not the only form of adjunction which may be created through reification in a topic map. For instance, the reified association of Brutus killing Caesar with a dagger may serve as a role playing topic in an association stating that “Marcus Antonius hated Brutus because Brutus killed Caesar with a dagger.” In this instance, the second association functions as the cause or reason of the first. In discourse terms, this is an example of enhancement: one association frames, as it were, the propositional content of another. The last form of adjunction, extension, may be produced by reifying two (or more) associations and combining them in a new association. For example, in order to represent “Brutus loved Caesar. But he loved Rome more” one may have to create an association called “comparison” (or something similar) in which the first reified association is given the role of “less” and the second the role of “more”: Brutus loved Caesar and he loved Rome but in comparison his love for Caesar was less than his love of Rome.

While reification is a useful mechanism for creating information flows within topic maps based on conjunction and adjunction, it does not seem to be entirely adequate for encoding interjunction, the linking of addressee to addressee. For this purpose, scope seems more useful. As noted above, scope is applied in topic maps to indicate contexts in which topic characteristics are deemed to be valid. But scope may also be employed to “qualify” or “colour” a certain statement. Thus, scope may not only indicate if a specific association is actually true or false but also the extent to which it is thought to be credible, likely, possible or imaginable. Or it may express the topic map creator’s own attitude towards it: good, better-than-average, acceptable or bad? In other words, scope may capture the kind of meanings in discourse that are typically expressed through grammatical categories like mood (modal verbs) or sentence adverbials (“undoubtedly”, “probably”, “surprisingly”, etc.). How exactly one should encode a bit of discourse like “Brutus killed Caesar. That was terrible thing to do” may be debated but one method would be to simply scope the assertion “Brutus killed Caesar” with a topic connoting “terribleness”.

JOURNAL OF INFORMATION ARCHITECTURE
SPRING 2010, VOL 2 ISS 1 — Trajectories

13
Texture and Exposition

As for the visual “3D” model of topic maps presented above, texture is an aspect of detailing and therefore related to the third dimension of the model: texture adds depth to information structure and descriptions in topic maps, and hence more generally in shared information spaces. In elaborating, framing or juxtaposing statements more light may simply be shed on their informational content.

Arguably, texture may be added to topic maps for three major reasons: The first is to explain specific (otherwise unrelated) statements or “facts” in a topic map: when, where, how and why does some event take place? For instance, why, when, how and why did Brutus kill Caesar with a dagger? The second is to foreground topics. In the example above, the topic of “the murder weapon” is brought to the fore as the need arises to detail the role of the dagger in the killing of Caesar. The third is to create information chains or pathways through topic map space. Usually, a topic map is an unordered set of topics and statements about these topics. And normally no topics or statements have a higher status, or significance, than others. But by connecting topics and statements in textual chains, pathways may be constructed for users to follow, highlighting certain topics or topic clusters.

This corresponds somewhat to the notion of trails in hypertext systems and allows the topic map author to be more focused on certain aspects or features in the topic map or to select portions of topic map content that must be read or seen in a specified order [1].

In more general terms, texture provides means for conveying information in topic maps, rather than just organizing it, in ways that are amenable to specialized software.

Organizing Discourse on the Web Using Topic Maps

So far, focus in this article has implicitly been on the possibility of adding texture to topic maps primarily through the reification of material within those topic maps. Still, it is important to bear in mind that discourse structure may also be imposed on content external to the topic map itself. Since a topic map can represent addressable subjects such as documents, photos, video clips, or web pages (and their parts), it may also describe the way these are communicatively connected, so to speak. For example, a topic map may make explicit that a specific paragraph, section or chapter in a document functions as a textual introduction to, or explanation or interpretation of,
In principle, there are no constraints on what content items can be “discourse-linked” in topic maps: a blog entry may be seen as evidence for claims made in an online newspaper article or vice versa. Likewise, there are no restrictions, a priori, on the types of discourse relations that may be said to exist between specific information resources. This means that a topic map author is free to label the discourse relationship between, say, two tweets on Twitter in any way he or she wishes to. Because of this flexibility, topic maps have the potential of becoming a kind of (personal) Web 2.0 tool for charting or managing distributed web communication in various forms.

The fact that discourse relations may link addressable as well as non-addressable subjects also makes possible the construction of “mixed discourse” in topic maps. A mixed discourse may be understood as the conjunction of (joined) topic map statements, typically about non-addressable subjects, with external content using a discourse relation. An example of a minimal piece of mixed discourse would be the case where the reified topic map assertion “Brutus killed Caesar with a dagger” is linked, via a background relation, to a topic representing a textual account of the political climate in Rome prior to the murder of Caesar. Mixed discourse in other words facilitates the integration of a machine-readable semantic knowledge base with communication artefacts meant for human consumption.

Actually, it may be argued that the capability to model and represent discourse-like information and discourse-like information flows, will be useful, if not essential, in Semantic Web applications but perhaps even more so in systems associated with what has come to be known as the Socio-Semantic Web where human inputs and efforts are seen as significant contributions to the creation of semantically tagged information. The reason is simply that humans normally prefer to express themselves and communicate through discourse.

Streams, Semantic Micro-blogging, and Topic Maps

One area where the encoding of texture might be of value is within the field of micro-blogging where the call for more “semantic” services and applications is being heard with increasing frequency. Several people have come up with interesting ideas, concepts and approaches for designing and implementing more semantic micro-blogging systems. Jeff Sayre, for instance, has proposed an approach to distributed, semantic micro-blogging based on a reinterpretation of Nova Spivack’s concept of the Stream (Sayre 2010; Spivack 2009). In Spivack’s original work the stream is a metaphor
for the current web of data streams (blogs, feeds, etc.) generated by users and software that are built on top of the “old” web of sites, web pages and hyperlinks. What characterizes streams, according to Spivack, is that they:

- Are information flows centred around a particular topic
- Change often
- Can be accessed and consumed independently of their user-interface
- Are often linked by “acts of communication” (i.e. they often constitute comments, ratings, approval, etc.)

In Sayre’s reinterpretation of the concept, a stream denotes “the flow of ideas from a given individual. A Stream is thus a monologue that contributes to a greater conversation.” A stream is made up of small components, so-called drops, embodying one idea or statement. In the world of microblogging, a drop may be said to be the same as a posting. Further, streams may confluence into rivers and rivers may confluence into an entire ocean. In Twitter parlance, a drop equals a tweet, an individual’s musings are a stream and the output of the people one follows is a river. To this set of water metaphors Sayre adds the concept of channel which is defined as “drops that are grouped under a specific subtopic to form substream categories.” Thus a channel on Twitter might be somebody’s tweets on politics and another channel his tweets on his favourite football team.

In Sayre’s model, Semantic Web tools and methods are intended to facilitate the distribution, management and integration of drops, channels, streams, rivers and oceans. For instance, at a very basic level semantic tagging would enable users to subscribe to channels rather than streams and to make much more precise searches in rivers and oceans. Sayre mentions RDF, and derived formats like FOAF, as the preferred technology for organizing distributed micro-blogs and their associated networks of users. This is hardly surprising, RDF being the standard most often associated with the Semantic Web (Manola & Miller 2004; Brickley & Miller 2010).

There is no reason, of course, why Topic Maps might not be a sound alternative to RDF for representing, structuring and exposing metadata in micro-blogs. Subject identifiers can be attached to drops, channels, streams and rivers - instead of hashtags - to capture what these are really about; subject locators can identify individual data sets like drops or streams; and merging seems like an appropriate way to implement the confluence.
of streams and rivers. But, as exemplified above, topic maps might also be applied to encode discourse structures within and across drops, channels, streams, and rivers. For instance, a specific drop might not only be marked up in terms of what it is about but also what communicative function it performs in an ongoing discussion in a particular online community: is it a request, an elaboration, a piece of evidence, a rebuttal of a claim made by another blogger, or something entirely different? This sort of information is not only likely to lead to more effective search and retrieval results but also to more effective processing and accurate analyses of data streams and communication patterns.

Obviously, to be able to automatically or semi-automatically process text-like information flows in and across micro-blogging systems, a systematic approach to discourse representation needs to be adopted. Here users cannot be allowed to freely alt and apply discourse relations but have to choose from a predefined set. In Topic Maps terms, an ontology, i.e. a class system, of discourse relations needs to be developed and put into place and preferably be linked to a set of PSI’s to expound their meaning. Once again, Renkema’s work provides a good starting-point as it contains an extensible taxonomy of discourse types including definitions and examples. It is just waiting to be put into a topic map, really.

**Conclusions**

This article demonstrated and exemplified how texture may be built into topic maps in order to make them even more flexible for information and knowledge representation purposes. In more general terms, an attempt has been made to suggest how Topic Maps may function as a means of extending “information architecture” into what might be called “discourse architecture”.

Discourse architecture may be construed as a kind of hybrid between a traditional organizational scheme like a tree or a network classifying, describing and relating entities and their properties and running text in which situations are unfolded, contextualized and evaluated. Defined in this way, discourse architecture, and “discourse topic maps”, may be seen as an effort to enhance findability as well as intelligibility using a common approach or model.

Identifying in what settings and to what uses discourse topic maps may precisely be put has largely been outside the scope of this article apart from the section on semantic micro-blogging. It does seem reasonable to suggest, however, that they may have a role to play in areas where “explanation
is king” such as e-learning or collaborative knowledge creation. In such contexts topic maps should not only be conceived of as access points, or portals, to primary (web) content but as mergeable information products in their own right.

Other areas might include, as already indicated above, Web 2.0 and Web 3.0 scenarios where the focus is on user-generated content and/or automated processing. It goes without saying, of course, that more research is needed to explore and evaluate how topic maps may be designed, developed and applied in concrete information architecture projects and approaches and, more narrowly, what forms of texture might be relevant and valuable for what types of information architecture or domains.

Acknowledgments

The author is greatly indebted to all anonymous reviewers for their valuable comments on this article.

References


Footnotes

[1] The issue of how such information chains or pathways can actually be presented to, and browsed by, users, is not within the scope of this article but the extensive literature on navigation in digital environments, especially in hypermedia and web based ones, points to possible solutions and pitfalls. See for example McKnight, Dillon & Richardson 1993; Dillon 2000; Kalbach 2007; Hinton 2009.


Cite as
