DITA Specialization

By: Zarella L. Rendon zrendon@ptc.com

Abstract

Darwin Information Typing Architecture (DITA) is an OASIS standard, based on XML, for authoring and delivering information. DITA has become increasingly popular with organizations looking for a quick entry using a schema specifically designed to support structured documentation, such as technical manuals, reference guides, or any other information that is structural. For an enterprise new to XML, DITA is a quick, and often cheaper, way to get started, because it provides basic building blocks in the form of modular (or topic) authoring structures.

Modularity and simplicity of structure are two major reasons that organizations adopt DITA for their XML implementations. The ability to ramp up and to learn the tag set quickly are features that attract new users, and the fact that there is built-in support for formatting with existing tools make DITA a cost–effective entry point. In many cases, topic based authoring and object reuse is the organization's goal, and out-of-the-box DITA works well in their environment. However, if the necessity for an alternative structure arises, DITA provides a way to support changes to the application. Specialization, which is the reason for the "Darwin" in DITA, provides the ability to tailor the application of XML to the needs of the organization.

What is Specialization?

Specialization is basically the process of "extending" DITA by creating new structures, by adding new elements to existing structures, or by removing elements. The process involves using an existing element in DITA as a model for creating a new element that is more applicable to the implementation.

Specialization makes it possible to create new, more restrictive structure types, yet retain the style and output support already in place for the objects that are specialized. This is due to the fact that any specialized element retains the knowledge of its base origin, even if the element is derived from another specialized element.

Specializations that create new element structures are called *structural specializations*, and specializations that add new elements or attributes to an existing structure are called *domain specializations*.

You can also combine existing specializations in new ways by creating new "customized" document type shells, without creating new specializations at all.

Reasons for Specializing

Many organizations adopt DITA in order to impose a standard look and feel for publications, limiting free-form expression of output style by limiting the authoring process. Strict adherence to base DITA topics also limits the amount of time and effort needed to modify and maintain output stylesheet and transformations. However, because a DITA topic is the most general type of document structure, the need to specialize becomes apparent when the system is put to the test with real documents.

The ability to add new elements to an existing domain provides a measure of familiarity with element names that the user will appreciate and be more comfortable with when learning a new authoring approach.

In some cases, a simple name change of an element within a topic allows an author to more easily understand the required input, and also gives the stylesheet a point of reference for a different output style. For example, if a specific output style is required for a new kind of paragraph, let's say a legal paragraph, where the content of the new paragraph is always indented and italic indented and italic, we can define a specialization of the "p" element called legal-para that will be allowed anywhere a normal paragraph is found, yet allow a different output style from the normal paragraph. This type of specialization is called a *domain* specialization, and is placed in a module that is referenced by domain.

In other cases, a new structure may be required, for example a warranty topic that might retain the same body structure of the general topic, but removes extraneous elements, and allows a new output style to be applied to the warranty as a whole. This type of specialization is called a *structural,* or *Information Type* specialization, because it is defining a new information type. New information types can still take advantage of general ancestor style inheritance if needed, and call also pull in predefined elements from the generic topic types.

What is Involved in Specialization?

Specialization involves creating new modules for the entity or module files associated with a base DITA structure, and then creating a new document type shell that combines the new specialization files with existing files, resulting in a new document type. Understanding of DTD or Schema structure is required before any modification can be made, and should not be attempted without a thorough analysis and specification design of the new specializations.

The best way to start a new specialization is to start from an existing specialization that is based on the same structure you want to create, copy the .ent, .mod, or .xsd files, modify and add elements as required for your new specialization, and delete anything that you don't need. Once the modular specializations are done, you create a new document type that incorporates the new modules that are created for the specialized elements.

There are two special attributes that must be used when making a specialization in DITA, *class* and *domains*. These attributes are signals that help the processor understand what an element is specialized from and where to apply a specialized element.

- The *class* attribute maps a single element back to the element it was specialized from, and can have more than one iteration of specialization, until the final attribute maps to the original general element. For example, a copyright-para can be specialized from a legal-para, which in turn is specialized from the "p" element.
- The *domains* attribute allows a user to specify external modules that will become "active" in the domain where they are referenced. This allows you to create a separate module and use it only in the domain where you want the new element or attribute to appear. For example, if you only want the legal-para to show up in a reference type topic, you can include that domain in the attribute definition for reference, but not concept or topic, etc.

It should be noted that new specializations must be as restrictive as or more restrictive than the types upon which they are based. For example, you can't remove required elements or attributes and you can't add new elements except those that are based on existing elements. There are other restrictions based on the rules for specialization that can be found in the specification for DITA, or other resources on the Web.

Processing Specialized Content

Because of the inheritance properties of specialization, a new element or topic can take advantage of preexisting stylesheet and transformations by walking back up to the general ancestor and applying the same processing code. However, specialization of the structure also allows for new code to be written to apply specifically to a specialized element, which in most cases was the reason for the specialization to begin with.

In some cases, an organization may accept DITA content from another group or outside organization that has implemented a specialization. When this happens, the organization is faced with a decision to either adopt the same specialization and apply modified styles or transforms, or to let the inheritance properties apply without changing their code. A third option would be to *generalize* the specialization (apply a reverse transformation) so that the specialized elements revert to their original general forms. This need for the ability to generalize is a driving factor in the reason for the strictness of specialization rules. If a modification to the application is made without consideration to the proper rules for specialization, the specialization becomes invalid and generalization would be impossible.

Summary

As organizations moving into structured content by adopting DITA for their enterprise-wide solution start to analyze and map their existing content into the topic structure, they become aware of the need for minor domain changes, and in some cases, major structural changes to fit their authoring, processing, and output requirements. Out-of-the-box DITA serves to give the organization a quick entry and existing toolset for producing output, but when the need for change arises, the built-in specialization ability of DITA solves the issue and provides a migration platform for moving from generic to specialized content models. Modular authoring, content and code reuse, simplicity of structure, and ease of entry are key features of DITA, but specialization is the real power behind the standard.

References

OASIS Darwin Information Typing Architecture(DITA) Architectural Specification v1.0

URL: http://docs.oasis-open.org/dita/v1.0/archspec/ditaspecialization.html

Priestley, M., Schell, D.A., *Specialization in DITA: Technology, Process, & Policy*

URL: http://xml.coverpages.org/priestleySIGDOC2002-DITA.pdf

Rendon, Zarella, *DITA Specialization,* scheduled for publication in early 2010

URL: http://xmlpress.net/publications/dita-specialization/