DITA Proposed Feature #45b

DITA Proposed Feature #45b

Add page range indexing element

Longer Description

The current indexterm element cannot express the full range of indexing semantics needed for production book indexes. This proposal addresses the ability to express page ranges. Page ranges indicate where the index entry refers to an extended discussion that goes over a number of pages. According to the *Chicago Manual of Style* , "if a text discussion extends over more than one page (...), as it often does, *beginning and ending references have to be given* " (emphasis added). This would typically be manifested as a page range like 34-36. This is distinguished from individual references over consecutive pages (34, 35, 36).

The need to express page ranges is even more urgent in a single-sourcing context. You cannot tell when authoring if the pertinent range would span a page. Page breaks will change with the media type (letter size, pocket edition, large print edition). Inserting an illustration can turn a two paragraph range into a three page span. Index entry ranges will be even more common if the print boook and index pages use paragraph numbers instead of page numbers (e.g., 18.44-46), something quite common in nonfiction documentation. In this proposal, "page range" refers to both page number ranges and paragraph number ranges.

Index page ranges is a fairly common feature in authoring environments. Microsoft Word, FrameMaker and LaTeX offer index page ranges, to cite some of the most well-known applications. For writers moving to DITA, the absence of this feature will be quite jarring.

A page range cannot be expressed using starting and ending element tags: it would be too restrictive. Index entries are supposed to capture "pertinent statements" (c.f. *Chicago Manual of Style*), not structural content. According to the *Microsoft Manual of Style for Technical Publications* , an section range should *not* be in the index if it is listed in the table of contents. Rather, index entries capture content that may be orthogonal to the main content structure. Pertinent content for index entries can overlap or straddle each other or structural boundaries like task steps. The way to achieve this flexibility in expressing index page ranges is to use index "marker" elements within pairs of `indexterm` elements. This will consist of two new elements that will be added into the content model of the `indexterm` element:

- index-range-start
- index-range-end

For example, an index entry on cheese can start with:

```
<indexterm>cheese<index-range-start/></indexterm>
```

The range can close with:

```
<indexterm>cheese<index-range-end/></indexterm>
```

Due to the potential for orphaned range markers during map assembly, page ranges cannot span topics at the topic level. Index ranges that start within a topic must end in the same topic, excluding nested topics. Topic spanning can only occur at the map level by inserting `indexterm` elements into map metadata. ~~Processors can handle unpaired ranges by generating individual references. That is, if there is an `indexterm` with a range start marker but does not have a corresponding `indexterm` that ends the range, it should just generate a single page number reference as if there was no range start marker. On the other hand, an `indexterm` that terminates a range but has no corresponding `indexterm` that starts the range should be dropped from output.~~

Scope

Small: introduce two new elements to the `indexterm` element's content model.

Use Case

An author adds a page spanning index entry: `<indexterm>DITA<index-range-start/></indexterm>` . Later in the same topic, she adds a range terminating marker: `<indexterm>DITA<index-range-end/></indexterm>` . This spans 4 pages on paper, so the generated PDF looks like:

• DITA, 46-49

Costs

Minimal: only two new elements are added to one element's content model. Processors will have to be updated to interpret these elements, but they can degrade gracefully by ignoring them.

Benefits

This will add an important feature to DITA's support for book indexes, one that is fairly common.

Implementation

For outputs such as HTML that are not book-like, page numbers may make no sense. If a processor sees a valid page range, the recommended action is that it should generate a hyperlink from that index entry to the start of the range.

When page range markers are not properly paired, the following recommended processing should result in few surprises:

• If there is an `indexterm` with a range start marker but does not have a corresponding `indexterm` that ends the range, it should just generate a single page number reference in a book as if there was no range start marker.
• On the other hand, an `indexterm` that terminates a range but has no corresponding `indexterm` that starts the range should be dropped entirely from output.

Time Required

Less than a day.