Exporting Org Files to DocBook

Baoqiu Cui

Table of Contents

1. Introduction

This is a text file written in Org-mode (http://www.orgmode.org), and it is used as a test case for my Org-to-DocBook exporter. Each section contains some feature in Org-mode...

2. About Sections and Subsections

DocBook supports numbered sections and also *recursive* sections. We choose recursive sections (element `<section>`) in DocBook exporter to make generated DocBook documents much more portable. Sections can be freely moved to different places, either in an article or in a book, at different levels.

As Org files tend to contain a lot of headings, which are transformed to *empty* sections in DocBook. Empty paragraphs, like `<para></para>`, are added for such sections to make generated XML document valid.

2.1. Subsection x.1

2.1.1. Empty Subsection x.1.1

2.1.2. Subsection x.1.2

2.2. Subsection x.2

2.2.1. Subsection x.2.1

2.2.1.1. Subsection x.2.1.1

2.2.1.1.1. Subsection x.2.1.1.1 (Level 5)

2.2.1.1.1.1. Subsection x.2.1.1.1.1 (Level 6)

2.2.1.1.1.1.1. Subsection x.2.1.1.1.1.1 (Level 7)

2.2.1.1.1.1.1.1. Subsection x.2.1.1.1.1.1.1 (Level 8)

2.3. Subsection x.3

3. About Images

Here is an image with caption and label:

Caption of image GNU.png

You can also add attributes to element `<imagedata>` via `#+ATTR_DocBook:` line to control how an image is displayed. You can customize variable `org-export-docbook-default-image-attributes` to set the default attributes that you want to append to all the images in your Org file.

Org-mode Logo (Unicorn) in SVG Format (width="3in")

In Org-mode, an image file can be used as the description of a link, and the image is exported as a click-able image in HTML export. In DocBook format, the image file path is formatted as it is. See the following link: /Users/bcui/Org/docbook-cover.png

3.1. Inline Images in Tables

Image One	Image Two

4. About Timestamps, TODOs, and Tags

This is a timestamp *2009-03-02 Mon*.

For TODOs, we can simply put the TODO keywords in `[...]` to make them stand out in section titles. For tags, they are formatted as superscript. We may need to find other better-looking formats.

4.1. [TODO] My First Item. :work:urgent:

4.2. [DONE] My Second Item

CLOSED: *2009-04-14 Tue 18:51*

4.3. [TODO] My Third Item. :home:_old:

SCHEDULED: *2009-04-14 Tue 18:51* aaa bbb CLOSED: *2009-04-14 Tue 20:30* ccc

5. Inline Tasks (TODOs)

This sections contains some inline task:

⇒⇒⇒ **TODO some small task**

```
                    And here is some extra text
```

⇒⇒⇒ **TODO This is the first inline TODO**

I need to do that following thing for the first TODO:

1. First item for first TODO

2. Second item for first TODO

⇒⇒⇒ **TODO This is the second inline TODO**

This is my second TODO.

⇒⇒⇒ **TODO This is the third inline TODO**

This is my third TODO.

6. About Footnotes

Please see my first footnote. [1] And here is the second one. [2]

Check the first footnote again. [1]

Here is a long footnote. [3]

This is a LaTeX-like anonymous footnote. [4]

---

[1] Body of *first* footnote.
[2] Body of *second* footnote.
[3] This is the body of the *third* footnote, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long, and it is long.... Stop here!
[4] This is the inline definition of this anonymous footnote.

This is another footnote with inline definition *and* a name. [5]

Here I am referring to the previous named footnote. [5]

7. About Links

This is an http link: CNN Web Site.

This is another one: http://www.cnn.com.

An http link with special characters: http://www.cnn.com/xx=5&yy=6&zz=7

Another link with superscript and other special characters in the link description: EMC website ($E = mc^2 > 0$)

FTP link: One Gnu FTP Link.

FTP link: ftp://ftp.gnu.org/path/to/something.

File link: file:~/.emacs::45.

File link: file:~/Org/test0.org.

Check the first footnote for the third time. [1]

8. About Tables

8.1. Regular Tables

If a table has a caption, then it will be formatted as a regular table in DocBook using element `<table>`. Here is an example table that has a caption "My Phone Book":

Table 1. My Phone Book

First Name	Last Name	Tele #
Baoqiu	Cui	123-4567
David	Smith	999-9999
John	Dean	888-8888
Scott	Fisher	111-1111

8.2. Informal Tables

Tables without a caption will be formatted as informal tables in DocBook using element `<informaltable>`. Such tables do not have a caption. See this example:

Item	Owner	Due Date	Status
1	Baoqiu	2009-03-10	Returned
2	Scott	2009-03-20	On Hold

---

[5]This is the definition of second inline footnote, which has a name.

Item	Owner	Due Date	Status
3	John	2009-04-20	Overdue

9. About Verses and Quotes

If the source of a verse or quote (a.k.a. *attribution* in DocBook) is found in the verse or quote, an `<attribution>` element will be added to element `<blockquote>`. This helps format verses or quotes more nicely.

9.1. Verses

The following is a verse with attribution ("`-- AlexSchroeder`", which appears in the *last* line in the verse):

> Great clouds overhead
> Tiny black birds rise and fall
> Snow covers Emacs
>
> —AlexSchroeder

The next one is a verse *without* attribution:

> O, for a muse of fire, that would ascend
> The brightest heaven of invention!
> A kingdom for a stage, princes to act,
> And monarchs to behold the swelling scene!

9.2. Block Quotes

This is a block quote with attribution ("`-- Albert Einstein`", which appears in the *last* line of the quote).

> Everything should be made as simple as possible, but not any simpler.
>
> —Albert Einstein

Here is one block quote that has two paragraphs but no attribution:

> God does not care about our mathematical difficulties. He integrates empirically.
>
> Great spirits have always found violent opposition from mediocrities. The latter cannot understand it when a man does not thoughtlessly submit to hereditary prejudices but honestly and courageously uses his intelligence.

10. About Literal Examples

You can include literal examples that should not be subjected to markup. Such examples will be typeset in monospace, so this is well suited for source code and similar examples.

```
Some example from a text file.  This can include characters like '<'
and '>', etc.

The following line should not be regarded as the continuation of a long
footnote:
\par This is NOT part of a footnote.
```

```
The following empty DocBook paragraph has property 'org-protected,
and thus should NOT be removed during the final cleanup time.

<para>
</para>

The following empty section is also protected, so should NOT be
filled with an empty paragraph ("<para></para>") by DocBook
exporter:

<section xml:id="sec-2.1.1">
<title>Empty Subsection x.1.1 </title>
</section>
```

For simplicity when using small examples, you can also start the example lines with a colon followed by a space. There may also be additional whitespace before the colon:

Here is an example:

```
#include <iostream>
#include <iterator>
using namespace std;

int main(int ac, char* av[])
{
    try {
      ...
    }
    catch(exception& e) {
        cerr << "error: " << e.what() << "\n";
        return 1;
    }
    catch(...) {
        cerr << "Exception of unknown type!\n";
    }

    return 0;
}
```

11. About Drawers

Drawers are something that people normally do not want to see. So they should not appear in the DocBook output. This is just to confirm that.

11.1. This is a headline

Still outside the drawer. After the drawer.

12. About Typesetting

The following markups are supported in Org-mode: <u>underline</u>, **emphasize**, `code`, *italic*, `verbatim`, ~~strike-through~~.

More examples:

- `[!#&*+-/:;<=>?]`

- `[\t\n\r ,]`

13. About Lists

13.1. Itemized Lists

This is paragraph one. This is paragraph one. This is paragraph one. This is paragraph one. This is paragraph one. This is paragraph one. This is paragraph one. This is paragraph one.

This is paragraph two. This is paragraph two. This is paragraph two. This is paragraph two. This is paragraph two. This is paragraph two. This is paragraph two. This is paragraph two.

This is an itemized list:

• Item one.

• Item two.

13.2. Ordered Lists

This is an ordered list:

1. Number One.

    Another paragraph for number one item.

    Another another paragraph for number one item.

2. Number Two.

13.3. Definition Lists

This is a variable list (a.k.a. Variable Lists in DocBook):

Term One   Term 1 definition. Term 1 definition. Term 1 definition. Term 1 definition. Term 1 definition. Term 1 definition. Term 1 definition.

Term Two   Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition. Term 2 definition.

13.4. Check Box Lists

This is a check box list:

• `[-]` Check boxes [2/3]

  • `[X]` Check box 1

  • `[ ]` Check box 2

  • `[X]` Check box 3

13.5. Nested Lists with Different List Types

This is a 3-level nested list:

- Item one.

  This item contains a description list (or *variable list* in DocBook):

  Term ONE   This is term one.

  Term TWO   This is term two.

  1. Description one.

     Second paragraph in description one.

     Third paragraph in description one. After this paragraph, we have two consecutive lists with different list types. These two lists should have been exported as *two* different lists, but Org mode currently does not support this:

     a. Ordered list item one

     b. Ordered list item two

     c. Itemized list item one

     d. Itemized list item two

     e. Itemized list item three

  2. Description two.

     ```
     Code in description two (line 1)
     Code in description two (line 2)
     ```

  3. Description three.

     Another paragraph for "Term TWO".

  Another paragraph for "Item one".

- Item two.

14. Random Pieces

14.1. Lord of the Rings

My favorite scenes are (in this order)

1. The attack of the Rohirrim

2. Eowyn's fight with the witch king

   - this was already my favorite scene in the book

   - I really like Miranda Otto.

3. Peter Jackson being shot by Legolas

   - on DVD only

He makes a really funny face when it happens.

But in the end, no individual scenes matter but the film as a whole. Important actors in this film are:

Elijah Wood   He plays Frodo

Sean Austin   He plays Sam, Frodo's friend. I still remember him very well from his role as Mikey Walsh in the Goonies.

15. About Code Reference Labels

The following is a two-line program listing. Since options "-n -r" are specified, we display the line numbers but remove the code reference labels.

```
1:  (save-excursion
2:     (goto-char (point-min))
```

In line 1 we remember the current positon. Line 2 jumps to point-min.

The following listing continues from the previous one (option "+n" is specified). Since option "-r" does not exist, code reference labels are *not* removed:

```
3:  (defgroup org-export-docbook nil
4:     "Options for exporting Org-mode files to DocBook."  (doc)
5:     :tag "Org Export DocBook"                           (tag)
6:     :group 'org-export)
```

Line (doc) is the group documentation, and line (tag) is the group tag.

16. About Subscripts and Superscripts

"`E = mc^{2}`" is exported as $E = mc^2$.

$x_1^2 + x_2^2 = y^2$

$x_1^2 + x_2^2 = y^2$

17. About Quoting DocBook Code

An Org file can quote DocBook XML code using one of the following constructs:

```
#+DOCBOOK: Literal DocBook code for export
```

or

```
#+BEGIN_DOCBOOK
All lines between these markers are exported literally.
#+END_DOCBOOK
```

Here are three examples (notice that the last one is a **warning** admonition in DocBook, something that Org-mode does not provide):

One line written in DocBook

Another paragraph written in DocBook XML code: $E = mc^2$.

⊗ Warning

You should know what you are doing when quoting DocBook XML code in your Org file. Invalid DocBook XML file may be generated by DocBook exporter if you are not careful!

This is a paragraph right after the DocBook code block, and should be wrapped into a paragraph.

18. About Text Centering

`#+BEGIN_CENTER` ... `#+END_CENTER` is a newly added feature in Org-mode 6.24trans to support text centering. This feature is currently supported via `<informaltable>` (with one row and one column) in DocBook exporter. Element `<para role="centered">` does not seem to work well with FOP.

<div align="center">Everything should be made as simple as possible, but not any simpler.</div>

<div align="center">Repeat the above paragraph:</div>

<div align="center">Everything should be made as simple as possible, but not any simpler.</div>

This is a paragraph right after the centered block, and should be wrapped into a paragraph.

19. Special Characters

Special characters that are written in TeX-like syntax, such as `\alpha` (α), `\Gamma` (Γ), `\Zeta` (Z), and `\sigma` (σ), are supported by DocBook exporter. These characters are rewritten to HTML entities like `α`, `Γ`, `Ζ`, and `σ`, so as long as the DocBook XML file includes the right entities, these special characters are recognized correctly.

You can customize variable `org-export-docbook-doctype` to include the entities you need. By default, this variable contains the following value:

```
"<!DOCTYPE article [
<!ENTITY % xhtml1-symbol PUBLIC
\"-//W3C//ENTITIES Symbol for HTML//EN//XML\"
\"http://www.w3.org/2003/entities/2007/xhtml1-symbol.ent\"
>
%xhtml1-symbol;
]>
"
```

For example, the following line in an Org file will generate this: $\alpha \le \sum 2^{\beta^\gamma} + \delta$

```
\alpha \le \sum 2^{\beta^{\gamma}} + \delta
```

Characters like <, >, and & should be exported as HTML entities.

20. About Macros

We can use macros to include current date and time in Org file, like this: 2009-05-20 00:28:23. We can also use macros to expand some commonly used abbreviations. For example: "Time To Live" (expanded from macro `TTL`) for cache can be configured in the following lines in the configuration file:

```
Time To Live of cache: 1000 minutes.
```