                 Writing I-Ds and RFCs using Pandoc
                     draft-gieben-pandoc2rfc-00

Abstract

   This document presents a technique for using Pandoc syntax as a
   source format for documents in the Internet-Drafts (I-Ds) and Request
   for Comments (RFC) series.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   This document presents a technique for using Pandoc syntax as a
   source format for documents in the Internet-Drafts (I-Ds) and Request
   for Comments (RFC) series.

   Pandoc [Pandoc] is an "almost plain text" format and therefor
   particularly well suited for editing RFC-like documents. The syntax
   itself is a superset of the syntax championed by Markdown [Markdown].

   Pandoc2rfc generates XML compatible with [RFC2629].

2. Pandoc to RFC

   During the last few years people have been developing markup
   languages that are very easy to remember and type. These languages
   have become known as "almost plain text"-markup languages. One of
   the first was the Markdown ([Markdown]) syntax. One that was
   developed later and incorporates Markdown syntax and a number of
   extensions is Pandoc ([Pandoc]). The power of Pandoc comes from the
   fact that it can be translated to numerous output formats, including,
   but not limited to: HTML, (plain) Markdown and "docbook" XML.

   Pandoc2rfc allows authors to write in (the simple) Pandoc syntax
   which is then transformed to XML and given to xml2rfc. The
   conversions are, in some way amusing, as we start off with (almost)
   plain text, use elaborate XML and end up with plain text again.

```
        +-------------------+   pandoc   +---------+
        | ALMOST PLAIN TEXT |   ------>   | DOCBOOK |
        +-------------------+            +---------+
                  |                           |
        non-existent  |                       | xsltproc
           faster way |                       |
```

```
              v                       v
     +------------+    xml2rfc  +---------+
     | PLAIN TEXT |  <--------  |   XML   |
     +------------+             +---------+

          Attempt to justify Pandoc2rfc.
```

Figure 1

The XML generated (the output after the "xsltproc" step in Figure 1)
is suitable for inclusion in either the "middle" or "back" section of
an RFC.

Even though Pandoc2rfc abstracts away a lot of XML details, there are
still a few places left where some XML needs to be edited. Most
notably in the "front" section of an RFC.

The simplest way to start is to create a template XML file and
include the appropriate XML for this "front" section:

```
<?xml version='1.0' ?>
<!DOCTYPE rfc SYSTEM 'rfc2629.dtd' [
<!ENTITY pandocMiddle PUBLIC '' 'middle.xml'>
<!ENTITY pandocBack   PUBLIC '' 'back.xml'>
<!ENTITY rfc.2629     PUBLIC '' 'reference.RFC.2629.xml'>
]>

<rfc ipr='trust200902' docName='draft-gieben-pandoc2rfc-00'>
 <front>
    <title>Writing I-Ds and RFCs using Pandoc</title>
</front>
<middle>
    &pandocMiddle;
</middle>
<back>
    <references title="Normative References">
            &rfc.2629;
    </references>
    &pandocBack;
</back>

</rfc>
```

A very minimal template.xml.

Figure 2

In this case you need to edit three documents:

1. middle.txt - contains the main body of text;

2. back.txt - holds appendices;

3. this template.xml - probably a fairly static file, but it also
    holds the references.

The up to date source code for Pandoc2rfc can be found at
[Pandoc2rfc], this includes the stylesheet "transform.xsl" used for
the transformation.

2.1. Dependencies

Pandoc2rfc needs "xsltproc" [XSLT] and "pandoc" [Pandoc] to be
installed. When using the template from Figure 2 xml2rfc version 2
needs to be used.

The conversion is done with a stylesheet based on XSLT version 1.0
[W3C.REC-xslt-19991116].

3. Building an Internet-Draft

Assuming the setup from Section 2, we can build an I-D as follows (in
a Unix-like environment):

```
        pandoc -t docbook -s middle.txt | xsltproc \
            --nonet transform.xsl - > middle.xml
        pandoc -t docbook -s back.txt | xsltproc \
            --nonet transform.xsl - > back.xml

        # Create text output
        xml2rfc template.xml -f draft.txt --text

        # ... or create HTML output
        xml2rfc template.xml -f draft.html --html

        # ... or create (self-contained) XML output
        xml2rfc template.xml -f draft.xml --exp
```

                Building an I-D with Pandoc2rfc.

                        Figure 3

The "draft.txt" and "draft.xml" are then ready to be uploaded to the
I-D tracker.

4. Supported Features

Almost everything xml2rfc can do is supported, see Table 1 in
Appendix A for a "cheat sheet".

o Sections with an anchor and title attributes

o Lists

   * style=symbols, use "* " for each item;

   * style=numbers, use a digit "1. " for each item;

   * style=empty, use "#. " for each item;

* style=format %i, use roman lowercase numerals: "ii. ";

* style=format (%d), use roman uppercase numerals "II. ";

       * style=letters, use lower- or uppercase letters: "a. " and
          "A. " (note: two spaces);

       * style=hanging, use the Pandoc definition list syntax;

   o Spanx style=verb, style=emph and style=strong, respectively use:
      "`text`", "_text_" or "**text**";

   o Block quote - not supported by xml2rfc, so this is converted to
      "<list style="hanging">" paragraph;

   o Figure/artwork with an anchor and postamble (Section 7.1);

   o Tables with an anchor and postamble (Section 7.2);

   o References (Section 7.3)

       * external (eref);

       * internal (xref):

          + section (handled by Pandoc);

          + figures (handled by XSLT);

          + tables (handled by XSLT).

   o Citations, by using internal references;

   o Indexes, by (ab)using footnotes (Section 7.4).

5. Unsupported Features

   o Pandoc markup in the caption for figures/artwork. Pandoc markup
      for table captions -is- supported;

   o crefs: for comments, but you can use HTML comments: "<!-- ...
      -->";

   o preamble in figures and tables.

6. Acknowledgements

   The following people have helped shape Pandoc2rfc: Benno Overeinder,
   Erlend Hamnaberg, Matthijs Mekking and Trygve Laugstoel.

   This document was prepared using Pandoc2rfc.

7. Pandoc Style

In some cases the Pandoc syntax is slightly misused to get the
desired output, in the following paragraphs we will detail these.
Also the meta data features of Pandoc are not used in Pandoc2rfc.

7.1. Figure/Artwork

   Indent the paragraph with 4 spaces as mandated by Pandoc. Note that
   xml2rfc supports a caption with "<artwork>". If you add a "Figure:
   some text" as the last line, the artwork gets a "<postamble>" with
   the text after "Figure: ". It will also be possible to reference the
   artwork. If a caption is supplied the artwork will be centered on
   the page.

7.2. Tables

   A table can be entered by using Pandoc's table syntax. All tables
   are translated to a "<texttable>" element in xml2rfc. You can choose
   multiple styles as input, but they all are converted to the same
   style (plain "<texttable>") table in xml2rfc.

   The table caption is -always- translated to a "<postamble>". The
   "<preamble>" tag isn't supported. If a table has a caption, it will
   also be possible to reference the table.

7.3. References

   Any reference like: "[Click here](URI)", is an external reference and
   it converted to an "<ulink>". An internal (i.e. see Section X)
   reference is typeset with: "[Click here](#localid)" or
   "[](#localid)". This converts to a "<link>".

   For referencing RFCs, you will need add the reference source in the
   template, as an external XML entity. After that you can use:
   "[](#RFC2119)"

   Note that referencing figures/artworks and tables is done slightly
   different. The reason for this is that Pandoc does not have native
   support/syntax for this. We work around it, by "faking" the
   reference in the xslt. Thus for referencing figures/artworks and
   tables, you need:

   o To take the first 10 characters of the caption (i.e. this is the
      text -after- the string "Table: " or "Figure: ");

   o Translate spaces and single quotes ' to a minus "-";

   o Translate uppercase letters to lowercase;

   o For tables prefix the anchor with "tab:" and for figures use
      "fig:".

   The figure from Section 2 will get "fig:a-very-min" as an anchor.

Note that duplicate anchors are an XML validation error which will
make xml2rfc fail. These are not detected during the XSL
transformation.

7.4. Indexes

The footnote syntax of Pandoc is slightly abused to support an index.
Sub items are also supported. Use an exclamation mark ("!") to
separate them:

    [^1]: item!sub item

8. Security Considerations

This document raises no security issues.

9. IANA Considerations

This document has no actions for IANA.

10. Normative References

[Markdown]
          Gruber, J., "Markdown", 2004, <http://daringfireball.net/
          projects/markdown/>.

[Pandoc2rfc]
          Gieben, R., "Pandoc2rfc git repository", October 2012,
          <http://github.com/miekg/pandoc2rfc>.

[Pandoc]   MacFarlane, J., "Pandoc, a universal document converter",
          2006, <http://johnmacfarlane.net/pandoc/>.

[RFC2629] Rose, M.T., "Writing I-Ds and RFCs using XML", RFC 2629,
          June 1999.

[W3C.REC-xslt-19991116]
          Clark, J., "XSL Transformations (XSLT) Version 1.0", World
          Wide Web Consortium Recommendation REC-xslt-19991116,
          November 1999, <http://www.w3.org/TR/1999/REC-
          xslt-19991116>.

[XSLT]   Veillard, D., "The XSLT C library for GNOME", 2006, <http:
          //xmlsoft.org/XSLT/xsltproc2.html>.

Appendix A. Cheat Sheet

```
+--------------------+-----------------+---------------------------+
| Textual Construct  | Pandoc Syntax   | Xml2rfc Syntax            |
+--------------------+-----------------+---------------------------+
| Section Header     | "# Section"     | "<section title= ...>"    |
| Unordered List     | "* item"        | "<list style="symbols">"  |
| Unordered List     | "#. item"       | "<list style="empty">"    |
| Ordered List       | "1. item"       | "<list style="numbers">"  |
| Ordered List       | "a.  item"      | "<list style="letters">"  |
| Ordered List       | "ii. item"      | "<list style="format      |
|                    |                 | %i">"                     |
| Ordered List       | "II. item"      | "<list style="format      |
|                    |                 | (%d)">"                   |
| Ordered List       | "A.  item"      | "<list style="format      |
|                    |                 | (%C)">"                   |
| Definition List    | Definition      | "<list style="hanging">"  |
| Emphasis           | "_text_"        | "<spanx style="emph">"    |
| Strong Emphasis    | "**text**"      | "<spanx style="strong">"  |
| Verbatim           | "`text`"        | "<spanx style="verb">"    |
| Block Quote        | "> quote"       | "<list style="hanging">"  |
| Index              | Footnotes       | "<iref item="index"/>"    |
| Table              | Tables          | "<texttable>"             |
| Figure/Artwork     | Code Blocks     | "<figure><artwork>"       |
| External Reference | "[Click](URI)"  | "<eref target="..."/>"    |
| Internal Reference | "[](#id)"       | "<xref target="id">"      |
| Figure Reference   | "[](#fig:...)"  | "<xref target="..."/>"    |
| Table Reference    | "[](#tab:...)"  | "<xref target="..."/>"    |
| Citations          | "[](#RFC2119)"  | "<xref target="RFC2119">" |
+--------------------+-----------------+---------------------------+
```

The most important textual constructs that can be used in Pandoc2rfc.

Table 1

Author's Address

   R. (Miek) Gieben
   SIDN Labs

   Email: miek@miek.nl
   URI: http://miek.nl/