Using sphinx to build the NeXus documentation

Pete R. Jemian

Advanced Photon Source

2011-10-20

Presented at the 2011 NeXus Code Camp

http://www.nexusformat.org/NIAC2011_CodeCamp

ICMS Content ID: **APS_1425480**

Overview

- Move from wiki to DocBook
  - http://www.nexusformat.org/NIAC2008#Documentation_breakout
  - Printable documentation
  - Assert a certain order to the documentation
  - Use and control a consistent style throughout
  - Can be *versioned* to a specific NeXus release
- Present: NeXus wiki + DocBook (html + PDF)
  - Motivations to consider something different
- Proposed Future: NeXus wiki + sphinx (html + PDF)
  - Examples
  - Advantages
  - Disadvantages
  - Work to be completed before it is usable

Present: NeXus wiki + DocBook

- NeXus wiki is used as main NeXus WWW site
  - entry point to documentation, downloads, and organization
  - Plan to clean up the main page at this Code Camp
- Transition from wiki to DocBook-generated HTML is abrupt (http://trac.nexusformat.org/definitions/ticket/146)
- Hard to keep wiki pages in sync with DocBook source (http://trac.nexusformat.org/definitions/ticket/121)
  - After conversion from wiki to DocBook, some users have updated the wiki but that information has not been placed into the DocBook source
  - Not easy to track this type of contribution
- Users report NeXus manual is too big or hard to follow
  - Introduction chapter reorganized to follow
    - What is NeXus?
    - Motivations for the NeXus standard
- DocBook documentation can be ready for release after this Code Camp!

Why consider changing from DocBook source?

- DocBook is XML
  - Same basic technology as NXDL (not a reason to change)
  - Uses LaTeX to generate PDF (not a reason to change)
  - DocBook capable of handling complex, multivolume documents like the manual

  - XML tags obscure the actual documentation (see next slide, example from NXdata)
  - Few people try to edit the documentation
  - Apparently, the DocBook syntax is a challenge
  - DocBook (html + PDF) requires many minutes to build
  - DocBook toolchain has proven troublesome, at times, to maintain

  - What other tools could be used, yet still build the complexity of our manual?

Example DocBook documentation in NXdata

nxdl2rst | NXnote.nxdl.xml | NXobject.nxdl.xml | nxdl2docbook.xsl | NXdata.nxdl.xml | »3

```
    <doc>
        (required) <db:code>NXdata</db:code> is a template of
        plottable data and their dimension scales.
        <db:indexterm>
            <db:primary>NeXus basic motivation</db:primary>
            <db:secondary>default plot</db:secondary>
        </db:indexterm>
        It is mandatory  that there is at least one <db:code>NXdata</db:code> group
        in each NXentry group.
        Note that the <db:code>variable</db:code> and <db:code>data</db:code>
        can be defined with different names.
        The <db:code>signal</db:code> and <db:code>axes</db:code> attribute of the
        <db:code>data</db:code> item define which items
        are plottable data and which are <db:emphasis>dimension scales</db:emphasis>.
        <db:indexterm>
            <db:primary>NeXus basic motivation</db:primary>
            <db:secondary>default plot</db:secondary>
        </db:indexterm>
        <db:itemizedlist>
            <db:listitem>
                <db:para>Each <db:code>NXdata</db:code> group will consist of only one data set
                    containing plottable data and their standard deviations.</db:para>
            </db:listitem>
            <db:listitem>
                <db:para>This data set may be of arbitrary rank up to a maximum
                    of <db:literal>NX_MAXRANK=32</db:literal>. </db:para>
            </db:listitem>
```

Design | Source

Sphinx: the official documentation tool for Python

- http://sphinx.pocoo.org
- Relies on ReST (restructured text) to provide syntax and control style
- Can build very pretty HTML and PDF including:
  - Lists, tables, sections, Index, Table of Contents (global and local)
  - Equations
  - Graphics
  - Color-styled source code examples for various languages
- Used by many software projects (lots of examples available on the WWW)
- Tool chain easy to install into Python (`easy_install -U sphinx`)
- Source code is simpler, by far, than DocBook, *looks good even as text*
- For the manual parts already converted to sphinx, rebuilds in seconds
- Conversion tool exists on WWW (db2rst.py)
  - but needs more work to handle our manual

- Show short demo using Baow (http://www.baow.com) as Firefox add-on

From the Introduction

db2rst | nxdl2rst | nxxref | introduction.xml | index.rst

DocBook

```
-->
<section xml:id="WhatIsNeXus">
    <title>What is NeXus?</title>
    <para>The NeXus <indexterm significance="preferred">
            <primary>NeXus</primary>
        </indexterm> data format has four components: <variablelist>
            <varlistentry>
                <term>A set of design principles</term>
                <listitem>
                    <para>to help people understand what is in the data file
                </listitem>
            </varlistentry>
            <varlistentry>
                <term>A set of data storage objects</term>
                <listitem>
```

What is NeXus?

The NeXus data format has four components:

A set of design principles

to help people understand what is in the data files.

A set of data storage objects

(base classes and application definitions) to allow the development of more portable analysis software.

db2rst | nxdl2rst | nxxref | introduction.xml | index.rst

sphinx

```
.. $Id: index.rst 889 2011-09-13 06:28:14Z Pete Jemian $

.. _WhatIsNeXus:

==========
What is NeXus?
==========

The NeXus data format has four components:

.. index:: NeXus

#. :ref:`Introduction-DesignPrinciples`
   to help people understand what is in the data files.
#. :ref:`Introduction-DataStorageObjects`
   (base classes and application definitions) to allow
   the development of more portable analysis software.
#. :ref:`Introduction-SetOfSubroutines`
   (utilities) to make it easy to read and write NeXus data files.
#. :ref:`Scientific_Community`
   provides the scientific data, advice, and continued involvement
   with the NeXus standard. NeXus provides a forum for the scientific
   community to exchange ideas in data storage.
```

db2rst | nxdl2rst | nxxref | introduction.xml | index.rst | 1.1. What is NeXus? — NeXus: User Guide and Refer

file://C:/Users/Pete/Documents/eclipse/NeXus/definitions/trunk/sphinx/build/html/volume1/introduction/whatis/index.html

NeXus: User Guide and Reference Documentation 2011-documentation

PREVIOUS | NEXT | INDEX

1.1. What is NeXus?

The NeXus data format has four components:

1. *A Set of Design Principles* to help people understand what is in the data files.
2. *A Set of Data Storage Objects* (base classes and application definitions) to allow the development of more portable analysis software.
3. *A Set of Subroutines* (utilities) to make it easy to read and write NeXus data files.
4. *NeXus Scientific Community* provides the scientific data, advice, and continued involvement with the NeXus standard. NeXus provides a forum for the scientific community to exchange ideas in data storage.

Sphinx disadvantages

- Requires indentation to indicate substructure (like Python)
  - Could be a problem for documentation in NXDL (XML) files since an XML editor may reflow the text content of a <doc> element
- No automatic numbering of equations, figures, or tables
- Multi-volume books (like our manual) can be a challenge
  - Better to make separate books
  - Sphinx can link these together but I have not tried this yet
- Chapter / Section / Appendix numbering can be hard to control
- Cannot have lists or tables within table cells
  - see this example from NXdata
- Harder to generate from XSLT
  - Easy to generate from Python, though

HTML from DocBook source

ttributes	Type	Units	Description (and Occurrences)
	NXgeometry		Position of crystal
	Bragg Laue		How this crystal is used. Choices are in the list.
	NX_CHAR		Type or material of monochromating substance. Chemical formula can be specified separate indicate the (hkl) orientation. Use the "d_spacing" field to record the lattice plane spacing. T from an enumeration to a string since common usage showed a wider variety of use than a si list at the time of the change: • PG • Ge • Si • Cu • Fe3Si • CoFe • Cu2MnAl • Multilayer • Diamond Note that "PG" is Highly Oriented Pyrolytic Graphite. Also, "Cu2MnAl" had the comment "H

Sphinx: work to be completed

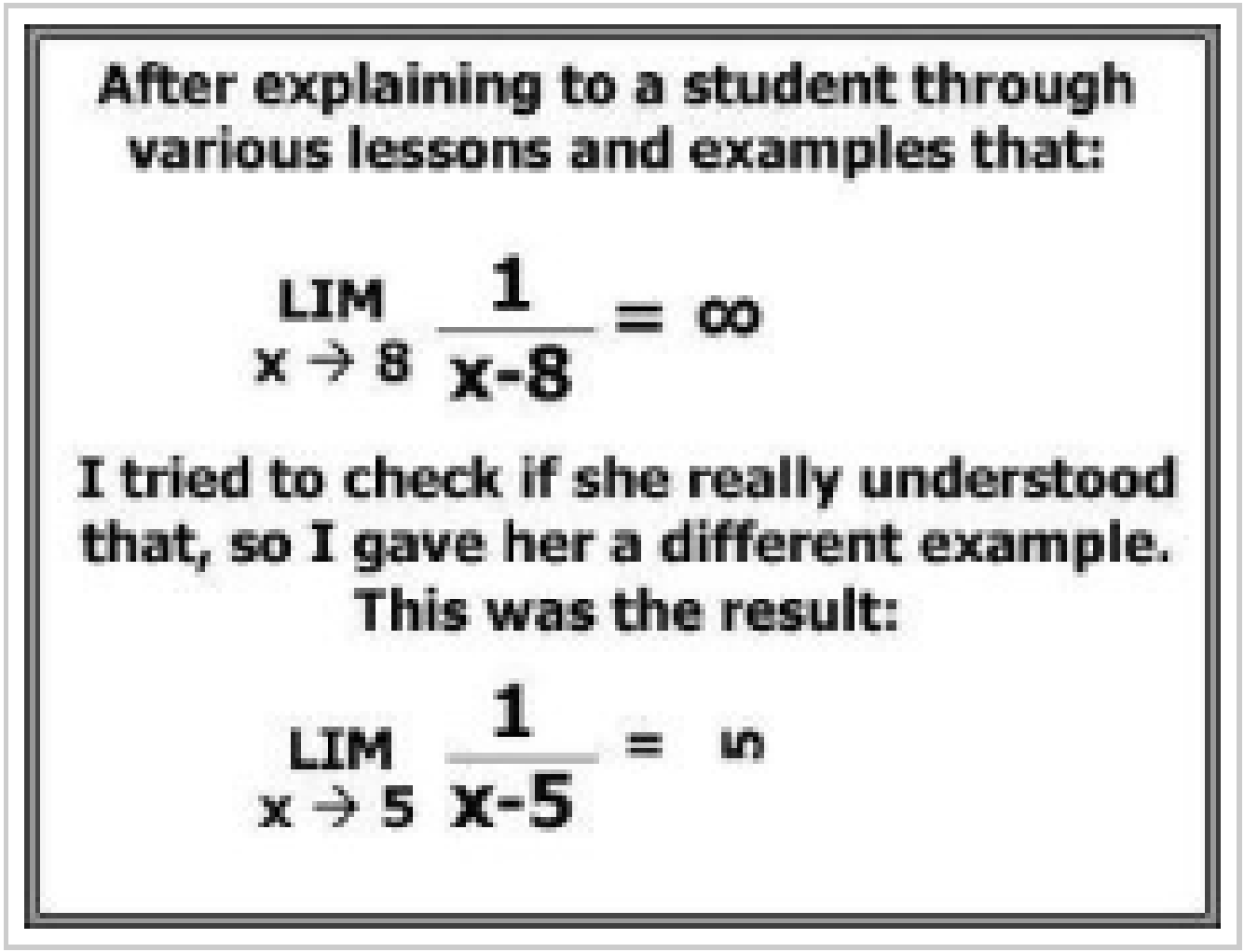
```
TODO items
----------

* Rendering of tables in NXDL classes is not done yet
* fix math source formatting between html and pdf
  * see examples at http://theoretical-physics.net/dev/src/math/integration.html
* tables, examples, and figures: treat them consistently with titles, captions, and cross-references
* stop the section numbering for very deep subsections (2.1.4.1.2.1.3.1.4.5.1.4.1... is just ridiculous)
* Should we produce two or more separate books?
* Convert NXDL doc strings into ReST

* note there is a figure number extension: https://bitbucket.org/arjones6/sphinx-numfig/wiki/Home
```

TODO items

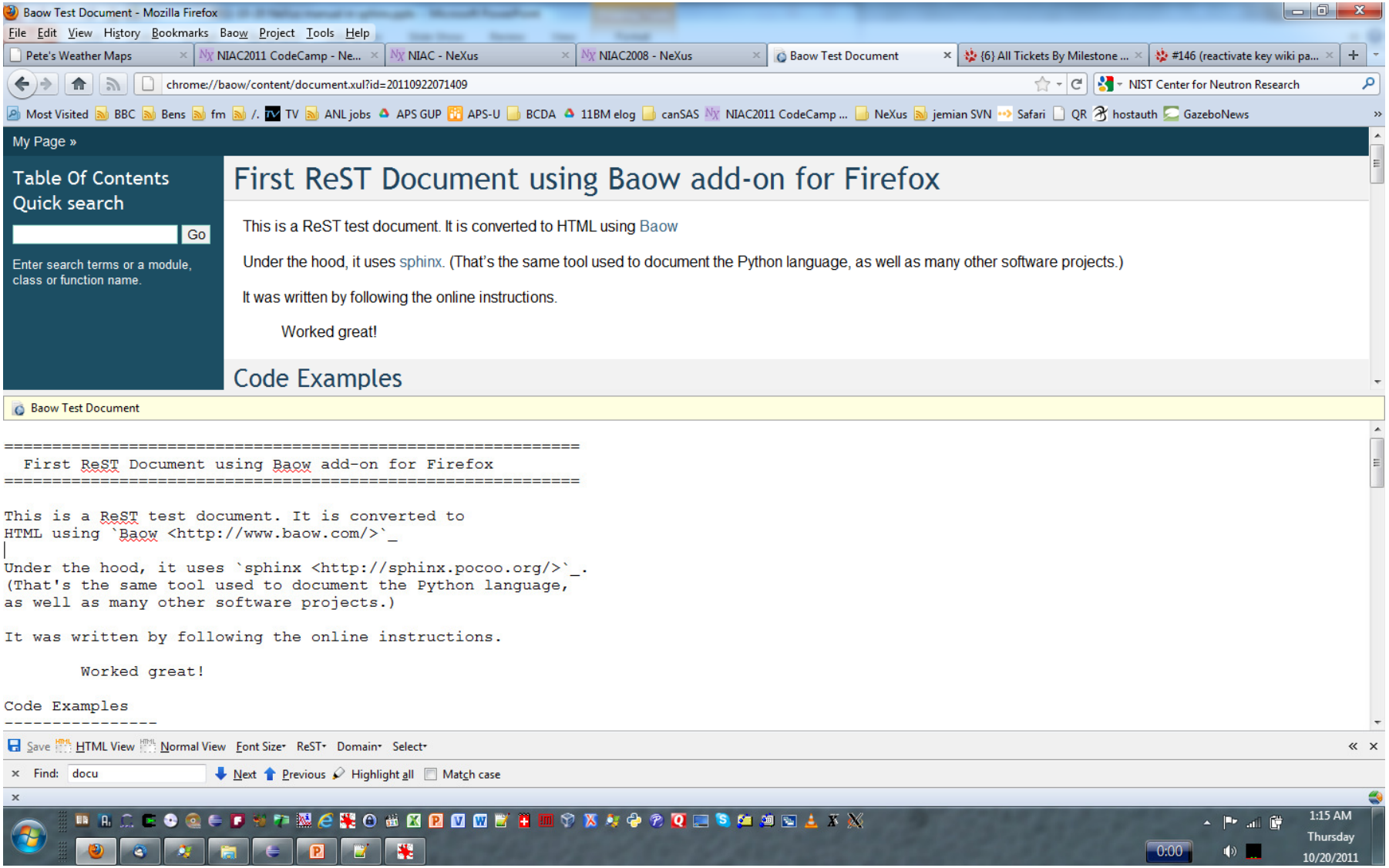
- Rendering of tables in NXDL classes is not done yet
- fix math source formatting between html and pdf * see examples at http://theoretical-physics.net/dev/src/math/integration.html
- tables, examples, and figures: treat them consistently with titles, captions, and cross-references
- stop the section numbering for very deep subsections (2.1.4.1.2.1.3.1.4.5.1.4.1... is just ridiculous)
- Should we produce two or more separate books?
- Convert NXDL doc strings into ReST
- note there is a figure number extension: https://bitbucket.org/arjones6/sphinx-numfig/wiki/Home

NeXus home page

Thank you for your attention!



Demo of Sphinx using Baow add-on for Firefox




COOKIE
DOUGH
Fondue
555CKT


$5.00
Deep Fried
Bacon Wrapped
Hot Dog
On A Stick