Top Ten Seven Reasons we use SMW at PNNL

Keith Fligg
Pacific Northwest National Laboratory
keith.fligg@pnl.gov
#7: Knowledge Encapsulation Framework.

- KEF – Knowledge Encapsulation Framework
  - Lab Directed Research and Development

- Features
  - SMW + Blog (Wordpress) + Forum (PHPBB)
  - Single Sign On via Kerberos

- Growth
  - About 15-20 sites using KEF
  - Growing pains moving from research task to client-ready
  - Kerberos (for SSO) was slow
  - Needed a build system
#6: A build system.

- Jenkins
- Ant
- Subversion
  - core – Core MediaWiki, Wordpress and PHPBB
  - core_extensions – SMW, Halo, SemanticForms, etc.
  - pnnl – PNNL global code and changes to above
  - <site> – Site specific code and changes to above
- Setup scripts
- KEF ‘Modules’
Jenkins

- Automatic builds when subversion changes to development/test server
- Manual builds to production or external server
- Subversion is not cutting it
  - We need separate development and production branches
  - Working on transitioning to git
KEF Subversion Setup and Hudson Layering

- Site Specific
- PNNL
- Extensions
- MediaWiki Core
KEF Setup

- KEF Modules that SMW instances can utilize
  - Automatic property and category creation
  - Automatic image and page creation (templates, Semantic Forms, etc.)
  - XML Specification
## PNNL Extension

<table>
<thead>
<tr>
<th>Folder</th>
<th>Last Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNNL</td>
<td>Today, 12:41 PM</td>
</tr>
<tr>
<td>includes</td>
<td>Apr 21, 2011 4:37 PM</td>
</tr>
<tr>
<td>PNNL_BibTeX.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>PNNL_Common.php</td>
<td>Apr 21, 2011 4:37 PM</td>
</tr>
<tr>
<td>PNNL_Globals.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>PNNL_phpBB_Bridge.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>PNNL_Settings.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>maintenance</td>
<td>Apr 23, 2011 10:31 AM</td>
</tr>
<tr>
<td>categories.xml</td>
<td>Apr 15, 2011 7:08 AM</td>
</tr>
<tr>
<td>cleanCategory.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>groups.xml</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>KEFCategories.xsd</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>KEFHaloACLGroups.xsd</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>KEFProperties.xsd</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>KEFSemanticData.xsd</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>moveToCategory.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>obliteratePages.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>PNNL_Setup.inc</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>PNNL_Setup.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>properties.xml</td>
<td>Apr 23, 2011 10:31 AM</td>
</tr>
<tr>
<td>renameProperty.php</td>
<td>Apr 14, 2011 8:27 PM</td>
</tr>
<tr>
<td>modules</td>
<td>Apr 14, 2011 9:36 PM</td>
</tr>
<tr>
<td>ADM</td>
<td>Apr 14, 2011 8:29 PM</td>
</tr>
<tr>
<td>AnalyticalGaming</td>
<td>Apr 14, 2011 8:29 PM</td>
</tr>
</tbody>
</table>
### Site Specific Overlay

<table>
<thead>
<tr>
<th>Folder</th>
<th>Date/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>wiki</td>
<td>Today, 12:39 PM</td>
</tr>
<tr>
<td>extensions</td>
<td>Today, 12:39 PM</td>
</tr>
<tr>
<td>ExhibitExtension</td>
<td>Apr 21, 2011 4:11 PM</td>
</tr>
<tr>
<td>PNNL</td>
<td>Today, 12:39 PM</td>
</tr>
<tr>
<td>includes</td>
<td>Apr 14, 2011 6:44 PM</td>
</tr>
<tr>
<td>PNNL_Config.inc</td>
<td>Apr 14, 2011 6:44 PM</td>
</tr>
<tr>
<td>maintenance</td>
<td>Apr 14, 2011 6:44 PM</td>
</tr>
<tr>
<td>PNNL_Setup.inc</td>
<td>Apr 14, 2011 6:44 PM</td>
</tr>
<tr>
<td>modules</td>
<td>Today, 12:43 PM</td>
</tr>
<tr>
<td>KEF</td>
<td>Apr 14, 2011 6:44 PM</td>
</tr>
<tr>
<td>NSF</td>
<td>Apr 21, 2011 9:35 PM</td>
</tr>
<tr>
<td>maintenance</td>
<td>Apr 23, 2011 10:49 AM</td>
</tr>
<tr>
<td>images</td>
<td>Apr 14, 2011 8:16 PM</td>
</tr>
<tr>
<td>ldrd-data</td>
<td>Apr 23, 2011 10:31 AM</td>
</tr>
<tr>
<td>addpubtemplate.php</td>
<td>Apr 14, 2011 8:13 PM</td>
</tr>
<tr>
<td>importldr.php</td>
<td>Apr 23, 2011 10:49 AM</td>
</tr>
<tr>
<td>metadata.xml</td>
<td>Apr 24, 2011 1:27 PM</td>
</tr>
<tr>
<td>nsfload.php</td>
<td>Apr 14, 2011 8:16 PM</td>
</tr>
</tbody>
</table>
Example site specific PNNL_Config.inc

```php
# the PNNL_Setup.php script uses this file to determine what modules are being
# included so that they can be initialized.
#
require_once("./modules/Global/maintenance/Global_Setup.inc");
require_once("./modules/ITD/maintenance/ITD_Setup.inc");
require_once("./modules/BachModel/maintenance/BachModel_Setup.inc");
require_once("./modules/KEF/maintenance/KEF_Setup.inc");
require_once("./modules/RSS/maintenance/RSS_Setup.inc");
require_once("./modules/AnalyticalGaming/maintenance/AG_Setup.inc");

setup->complete();
```
Example site specific PNNL_Settings.inc

```
# This file includes those PNNL modules that are required for the IBD site.

#
global $pnnlDir;

require_once($pnnlDir . "/modules/BachModel/includes/BachModel_Settings.inc");
require_once($pnnlDir . "/modules/KEF/includes/KEF_Settings.inc");
require_once($pnnlDir . "/modules/PDF2MediaWiki/includes/pdf2mw_Settings.inc");
require_once($pnnlDir . "/modules/AnalyticalGaming/includes/AG_Settings.inc");
require_once($pnnlDir . "/modules/Annotator/includes/A7R_Settings.inc");
require_once($pnnlDir . "/modules/REST/includes/REST_Settings.inc");
require_once($pnnlDir . "/modules/NetLogo/includes/NetLogoHooks.php");
```
#5: We needed a means of supporting internal communication and collaboration

- 15 to 20 sites with several hundred users thus far
- Internal ‘initiative’ sites
  - Carbon Sequestration Initiative
  - Microbial Communities Initiative
  - Signature Sciences Initiative
  - And others
- Project Sites
  - Marine Hydro Kinetic CMS
  - International Building Exchange
  - Naval Postgraduate School LiDAR Research
And even more sites

- Division and Communities of Interest internal sites
  - Computational and Statistical Analytics
  - Visual Analytics, etc.

- Research
  - National Science Foundation – Science and Innovation Research
  - Automated Content Analysis
#4: We wanted to play with Simile Exhibit.

► SRF Exhibit didn’t do all we needed
► Discovered a nascent MW Exhibit plugin
  ▪ Added support to do ask queries from the browser
  ▪ Added support for Exhibit collections
  ▪ Added support for Exhibit converters
► Found disabled extensions and little known, or used facets.
More Exhibit goodness…

▶ Added new facets
  - ‘Saved Filter Facet’ saves facet selections as collections
More Exhibit goodness…

- Date selection facet and view
More Exhibit goodness...

Graph View
More Exhibit goodness…

Property Export Selection View

- General Features
  - Initial Approval Date
  - Duration
  - Fourth Year SW
  - Research Type
  - Number of Awards
  - Sector Code Counts
  - Proposal Ratings

- Publications Features
- Collaboration Features
- IP Features
- Staff Features
- Finance Features
More Exhibit goodness…

Tag Cloud

Cloud

.NET  3D Graphics  Agile Development  AJAX  Algorithms  Apache  Blog  C/C++  C#  C++
Cognitive Science  Computational Linguistics  CSS  Database  Eclipse  Ethnography  Flash
Flex  Graphic Design  HTML  Human Computer Interaction  Human Factors  Human-
Centered Design  IIS  Information Visualization  Java  JavaScript  Javascript  Knowledge
Representation  Linux  Literature Review  Machine Learning  MediaWiki  MySQL  Natural Language
Processing  OpenGL  Perl  Photography  Photoshop  PHP  Program Management  Project
Management  project management  Python  RDF  Semantic Forms  Semantic MediaWiki
Semantic Mediawiki  Semantic Web  Social Media Analytics  Software Architectures  SQL
SQL Server  Statistics  Usability  User Experience Design  User Interface Design  User-Centered
Evaluation  Video  video and animation  Visual Analytics  Visual Studio  Web Development  Web
Services  Wiki  XML  XSLT

267  Staff Members

sorted by: has_name; then by...  •  grouped as sorted

1.

Aaron Alva

Phone Number  (509) 375-5630
Email Address  aaron.alva@pnl.gov
Job Title  Intern
Technical Group  Secure Cyber Systems

2.

Aaron Baumann

Phone Number  (509) 372-4549
Future Exhibit goodness…

Radar Graphs

Legend
- Machine Learning String Tools for Operational and Network Security - 730
- Molten Hydroxide CO2 Absorption with Steam Regeneration – 761
- Validation of Environmental Biomarkers in Periphyton Communities Exposed to Uranium – 812
- A Multi-Modal Integration Framework for Chemical Imaging – 966
- Correlative High Resolution Imaging and Spectroscopy to Characterize the Structure and Biogeochemical Function of Microbial Biofilms – 993
- Instrumentation, Control, Diagnostics, and Prognostics for Solid Oxide Fuel Cell – 40
- Detection of the Sources of Atmospheric and Nuclear Explosion Particulate Matter – 87
- Nondestructive Characterization of Life-Limiting Degradation of Nuclear Power Plants – 116
Future Exhibit goodness…

- Relationship Graphs
We have developed a Natural Language Processing pipeline that takes unstructured text and adds semantics.

- People
- Places
- Dates
- Rhetoric
- Statistically Improbable Phrases
- Sentiment (target and of groups)

Automated Discovery Mechanism finds documents that are similar to ‘seed’ documents in the wiki.

We developed a proof of concept PDF to Wiki markup converter.
Automated Annotation Pipeline (AAP)

1. Pages to be annotated are selected from KEF

2. Pages are processed and annotated by a pipeline of NLP tools

3. The annotated pages are put into KEF as a new revision
Automated Discovery Mechanism (ADM)

1. Seed docs are selected from KEF
2. Queries are extracted from the seed docs
3. Search terms are sent to data sources and results are harvested
4. Relevancy of each result is calculated
5. The most relevant result docs are put in KEF
#2: We wanted to write a Wiki markup parser.

- One of the KEF requirements was an GUI annotation tool that could:
  - Nest annotations from the NLP pipeline
  - Create annotations across paragraph boundaries.
- Started by hacking on the Halo Advanced Annotation Tool. But I just couldn’t get something working that was clean.
- So I build a real wiki text parser that generated AST’s that I then used to render the wiki text in a Sproutcore application.
Parser details

- PEG parser generator to deal with the vagaries of wiki text like the quote mess
- Not fully implemented, e.g., tables and images don’t work
- Generates JavaScript, so it runs in the browser.
- It is SLOW. Firefox and MSIE are nearly unusable, but Safari works well.
- A new version will likely create the AST’s on the server and cache them.
Annotation Details

- Store annotations as Semantic Internal Objects, with properties for
  - Property name(s)
  - Property value(s)
  - Text
  - Text offsets
Obligatory Screen Shot
#1: We wanted to work on an open source project.

- We are working on being able to contribute to MediaWiki and Semantic MediaWiki and other extensions
  - Bug fixes
  - Enhancements

- We are in the planning stages of open sourcing KEF.
  - Currently specifying what will be included
    - NLP pipeline (parts of it)
    - Annotator
    - Build system
    - What else?
The end…

Thanks!