The New NeIC Wikis

Petter Urkedal

February 11, 2013

About...

... creation of two NeIC wikis to replace the current NDGF wiki.

... moving from the Confluence software to MediaWiki.

Outline

About MediaWiki

Authentication and Access Control

Tools for Organizing Content

Moving Existing Content

Discussion

About MediaWiki

MediaWiki…

… was written for Wikipedia and first released in 2002.

… has gained popularity outside Wikipedia and its sister projects.

… is managed by the MediaWiki foundation who also coordinate the many extensions.

… is designed to be open both for editors and readers.

… is Free Software.

Authentication and Authorization

- You authenticate with your X509 certificate.
- The DN is not a usable MediaWiki user name.
- The CN part of it is neither secure nor usable.

Authentication and Authorization

- You authenticate with your X509 certificate.
- DNs are mapped to a full names and email addresses using a plain-text file. ∴ Ask us to authorize your DN.

Authentication and Authorization

- You authenticate with your X509 certificate.
- DNs are mapped to a full names and email addresses using a plain-text file. ∴ Ask us to authorize your DN.
- Public wiki: The log-in page is protected by SSL authentication. You click the log-in link to log in.
- Private wiki: The whole wiki is protected by SSL authentication. You are automatically logged in.

Access Control

- Designed for open content: Good at moderation, bad at access control.
- Extensions available, but warnings about information leaks.
- Protecting a whole wiki is no problem.

Access Control

- Designed for open content: Good at moderation, bad at access control.
- Extensions available, but warnings about information leaks.
- Protecting a whole wiki is no problem.

Therefore,

- We created two wikis, a public and an internal, and use the PrivatePageProtection extension for fine-grained access.
- Thus, for the internal wiki, unauthorized users don't have access to the features which potentially leak information.

Access Control, cont.

- The PrivatePageProtection extension use MediaWiki groups.
  - New groups are added in the configuration
  - Membership is edited within the wiki.
- A page is read and write protected by adding an instruction `{{#allow-groups` $G_1 | \cdots | G_n$`}}` to the page.
- We don't have full control of write protection which is independent of read protection, but we can...
  - ... create a user group with only read access.
  - ... make some pages editable only by staff.

Namespaces

… are not for organizing the main content.

… are used for meta-content and wiki user interface.

… are accessed as `Namespace:Page Title`.

Some notable namespaces include

- `Special` — Wiki UI like search, user management, quality control, etc.
- `Talk` — One for each page, used to discuss issues with the content.
- `User` — Each user has one for personal content.
- `Help` — Our documentation of the wiki itself.

Subpages in MediaWiki

- … are pages containing at least one slash and which *reside in a namespace where subpages are enabled*. I.e. page titled "$T_1/\cdots/T_n$" is a subpage of "$T_1/\cdots/T_{n-1}$"
- … are enabled in `Talk` and `User` namespaces by deafult.
- Extensively used in current wiki.
- Enabled for the main namespace in the new wiki.

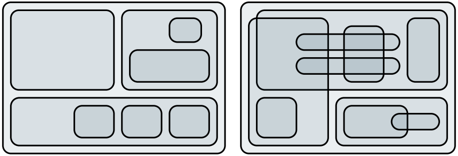
Categories

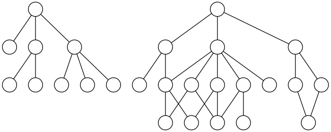
… are the default classification tool in MediaWiki.

… are sets of pages and other categories.

- A page or a category can be members of several categories.
- A category can contain content, though this is meant for defining it rather than elaborate content.
- A page or category is added to a category by placing a tag [[Category:Topic]] in the source.

Subpages vs Categories



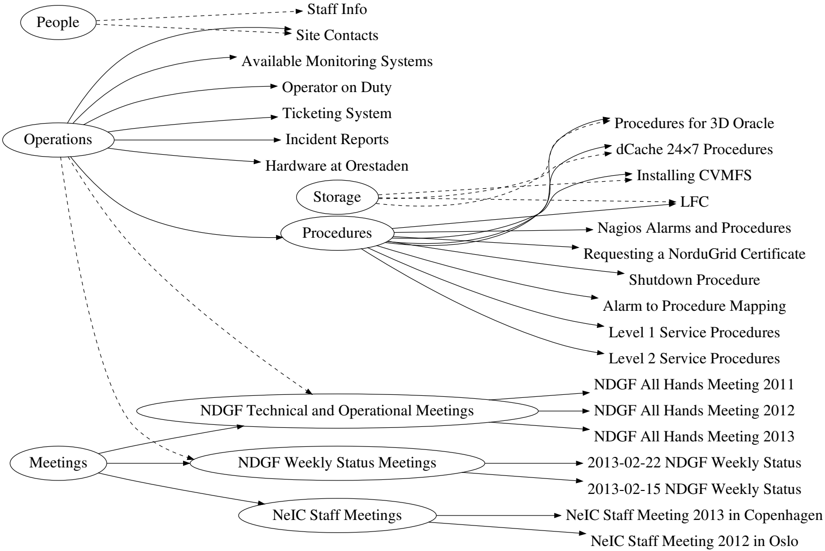


Subpages…

- … organize the content into a tree.
- … works as an intrinsic feature of pages.
- … are encoded in the title; need planning.

Categories…

- … organize the content into acyclic graphs.
- … are not regular pages, but can contain free prose.
- … have minor impact on the articles; can be added at any time.



People

Staff Info

Site Contacts

Operations

Available Monitoring Systems

Operator on Duty

Ticketing System

Incident Reports

Hardware at Orestaden

Storage

Procedures

Procedures for 3D Oracle

dCache 24×7 Procedures

Installing CVMFS

LFC

Nagios Alarms and Procedures

Requesting a NorduGrid Certificate

Shutdown Procedure

Alarm to Procedure Mapping

Level 1 Service Procedures

Level 2 Service Procedures

Meetings

NDGF Technical and Operational Meetings

NDGF All Hands Meeting 2011

NDGF All Hands Meeting 2012

NDGF All Hands Meeting 2013

NDGF Weekly Status Meetings

2013-02-22 NDGF Weekly Status

2013-02-15 NDGF Weekly Status

NeIC Staff Meetings

NeIC Staff Meeting 2013 in Copenhagen

NeIC Staff Meeting 2012 in Oslo

The Current Content

- Confluence wiki contains about 11 600 pages. Many meeting pages.
- Most of it we may want to preserve, but not use time reorganizing.
- Off-line dump good enough? Or can we convert it?

Confluence Database Dumps

- Zip archive containing a 139 MiB XML as a single file and 851 attachment files.

```
<?xml version="1.0" encoding="UTF-8"?>
<hibernate-generic datetime="2013-02-06 13:23:31">
  [...]
  <object class="Page" package="com.atlassian.confluence.pages">
    <id name="id">33918872</id>
    <property name="position"/>
    <property name="parent" class="Page" package="com.atlassian.confluence.pages">
      <id name="id">7832011</id>
    </property>
    <collection name="ancestors" class="java.util.List">[...]</collection>
    <property name="space" class="Space" package="com.atlassian.confluence.spaces">
      <id name="id">7864321</id>
    </property>
    <property name="title"><![CDATA[Meeting-2013-02-11 NDGF All Hands Meeting]]></property>
    <collection name="bodyContents" class="java.util.Collection">[...]</collection>
    <collection name="outgoingLinks" class="java.util.Collection">[...]</collection>
    <collection name="referralLinks" class="java.util.Collection">[...]</collection>
    <property name="version">28</property>
    <property name="creatorName"><![CDATA[maswan]]></property>
    <property name="creationDate">2012-12-11 16:48:34.000</property>
    <property name="lastModifierName"><![CDATA[270784@ku.dk]]></property>
    <property name="lastModificationDate">2013-01-21 15:56:34.000</property>
    <property name="versionComment"><![CDATA[]]></property>
    <collection name="historicalVersions" class="java.util.Collection">
      <element class="Page" package="com.atlassian.confluence.pages">
        <id name="id">33919294</id>
      </element>
      <element class="Page" package="com.atlassian.confluence.pages">
        <id name="id">33920589</id>
      </element>
      <element class="Page" package="com.atlassian.confluence.pages">
        <id name="id">33920851</id>
      </element>
    </collection>
    <property name="contentStatus"><![CDATA[current]]></property>
  </object>
  [...]
</hibernate-generic>
```

Confluence Database Dumps

- Zip archive containing a 139 MiB XML as a single file and 851 attachment files.
- The XML looks like a generic dump of an OODBMS.
- Each of the top-level elements represent an object, which
  - has an ID
  - has on of several classes
  - contains properties (name, type, value or reference)
  - contains collections (name, type, list of references)

Confluence Database Dumps

- Zip archive containing a 139 MiB XML as a single file and 851 attachment files.
- The XML looks like a generic dump of an OODBMS.
- Each of the top-level elements represent an object, which
  - has an ID
  - has on of several classes
  - contains properties (name, type, value or reference)
  - contains collections (name, type, list of references)

67270	`ReferralLink`
28125	`OutgoingLink`
11575	`BodyContent`
11569	`Page`
2340	`BucketPropertySetItem`
851	`Attachment`
265	`ContentPermission`
136	`SpacePermission`

82	`ContentPermissionSet`
33	`Labelling`
25	`Label`
5	`TrackbackLink`
4	`Comment`
4	`ConfluenceBandanaRecord`
2	`PageTemplate`
2	`SpaceDescription`
1	`Space`

Conversion?

- Seems feasible to extract sets of pages based on title and subpage structure.
- There are some utilities for supervised conversion, but they are probably not good enough for full automation.
- MediaWiki has a nice API which allows bots to query and edit.

Discussion

- What authorization groups do we need?
- How much of the existing wiki should we copy?
- Are there areas which need reorganization?
- Where and how can we put subpages to good use?
- In particular, how do we organize the meeting pages?