BlueSpice MediaWiki

EMWCon 2021: Confluence Migration Tool

www.bluespice.com

BlueSpice MediaWiki

Agenda

- History
- Considerations
- Technical background
- Migration process
- Quick Demo
- Questions

BlueSpice MediaWiki

History

- Two customer projects
  - Customer specific implementation
- Decision to make build a generic tool

BlueSpice MediaWiki

Considerations

- MediaWiki distribution agnostic
  - Create MediaWiki Import/Export-XML (consumeable by "importDump.php" or "Special:Import")
  - Create "upload files" with unique (collision free) names ("Massupload"/ "importImages.php")
- Support custom namespaces as target
- Support multiple confluence spaces in on migration
- Allow for re-structuration
- Migration can not be 100% accurate
  - Assist manual "post-migration" maintenance work

BlueSpice MediaWiki

Technical Background

- Datasource: Confluence XML export (per space)
  - Comes as ZIP archive
  - Contains "entities.xml" ("Java-Hibernate-file") and "attachments/"-subfolder

- XML holds
  - Pages
  - "BodyContents" (= Revisions)
  - Attachments
  - Comments (unsupported)
  - Blog entries (unsupported)

BlueSpice MediaWiki

Technical Background

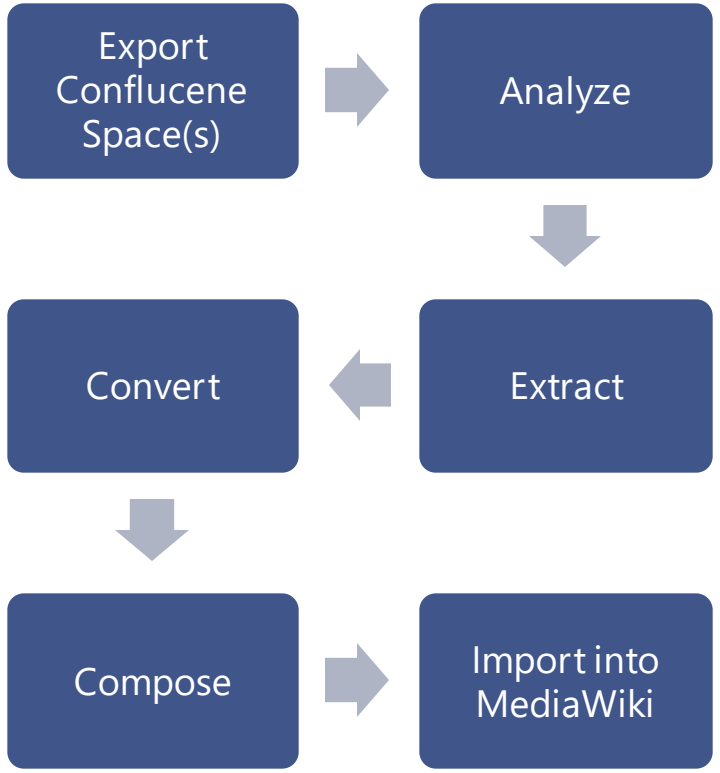
- "attachments/"-subfolder holds
  - Binary data of files
  - "obfuscated" filenames (Page-ID/Attachment-ID/Revision-ID)

- "BodyContents"
  - Confluence Storage Format
  - XHTML + Custom-XML-Elements (internal links, attachment-links, macros, layouts, ...)
  - Timestamp
  - User-Ids (can not be migrated)

BlueSpice MediaWiki

Migration process

- Command line tool with four Steps
  - Analyze
  - Extract
  - Convert
  - Compose
- Data shared between steps
- Data can be manipulated in between



BlueSpice MediaWiki

Migration process / Analyze

- Namespace-prefix-mapping: Confluence-Space-ID to MediaWiki-namespace-prefix
- Title-mapping: Confluence-Page-ID to MediaWiki-page-name
  - Invalid title detection (e.g. max. title length issues)
- Title-Attachment-mapping:
- Title-Metadata-Mapping: Categories, …
- Title-Revision-Mapping
- …

BlueSpice MediaWiki

Migration process / Extract

- Create one XML file per "BodyContent"
- Potential pre-transformations

BlueSpice MediaWiki

Migration process / Convert

- Conversion of "Confluence Storage Format"-XML
  - WikiText links (e.g. "Media:" for "attachments")
  - Image embeddings ("File:"-links)
  - Macros (WikiText-template calls, "Maintenance categories")
- Use `pandoc` to convert XHTML to WikiText
- Post-processing / cleanups

BlueSpice MediaWiki

Migration process / Compose

- Create MediaWiki-Import/Export-XML
  - Add metadata (Attachment-Links, Categories, ...)
- Copy all "attachment" files into a directory

BlueSpice MediaWiki

DEMO

BlueSpice MediaWiki

Questions

Get it from https://github.com/hallowelt/migrate-confluence

Feedback & contributions are welcome!

Contact

Robert Vogel

Hallo Welt! GmbH • Postfach 11 02 19 • 93015 Regensburg

E-Mail: vogel@hallowelt.com

Telefon: +49 (0)941 660 80 185

Telefax : +49 (0)941 660 80 189

www.bluespice.com

www.hallowelt.com