

docs.ruby-lang.org/ja/ の生成方法を変えた

Kazuhiro NISHIYAMA

【大阪オンライン開催】 *RubyKaigi 2024*
*KaigiEffect*発表会
2024-06-13

self.introduction

- 西山 和広
- Ruby のコミッター
- github など: @znz
- 株式会社Ruby開発 www.ruby-dev.jp

前の方法

- docs.ruby-lang.org で生成
- bc-setup-all で bitclust の db-* を生成
- bc-static-all で static html を生成 (/ja/バージョン/ の内容)
- update-rurema-index で rurema-search のインデックスを更新 (/ja/search/ の内容)

問題発生

- パターンマッチのドキュメントを ruby/ruby の rdoc を翻訳する形で追加 <https://github.com/rurema/doctree/pull/2773> のマージで問題発生
- CI はサポートバージョンでしか動いていなかった

```
% curl -s 'https://cache.ruby-lang.org/pub/misc/ci_versions/cruby.json' | jq -c '. + []'
["3.1", "3.2", "3.3", "head"]
```

docs の ruby

- Debian GNU/Linux 11 (bullseye) の `/usr/bin/ruby` だと
ruby 2.7.4p191 (2021-07-07 revision a21a3b7d23) [x86_64-linux-gnu]
- <https://snapcraft.io/ruby> の最新安定版を使っていたこともあったが
rurema-search との兼ね合いで `/usr/bin/ruby` に戻していた

HTML 生成の Docker 化

- 生成された db と html を入れても .git は 140M ぐらい
- [GitHub のリポジトリサイズ制限](#)
> リポジトリは小さく保ち、理想としては 1GB 未満、および 5GB 未満にすることを強くお勧めします。
- 生成したファイルもリポジトリ管理に

rurema-search の Docker 化

- rurema-search のインデックス作成も Docker 化
- インデックスはバイナリで git 管理には向かなさそう
- かつ毎回再生成すれば良さそう
- リポジトリ管理にはせず

古いドキュメントの保存

- HTML ファイルは ja/1.8.7 から残っていた
- <https://github.com/rurema/generated-documents> に保存
- db-* は EC2 を今の docs-2020 に移行したときに残していなかった
- db-2.4.0 以降のみ現存
- rurema-search には db-* が必要だったが 2.3 以前はないまま
- <https://docs.ruby-lang.org/ja/search/> は現状維持

GitHub Actions で生成

- github.com/ruby ではなく github.com/rurema に作成
- 権限がなくて S3 へ置けない
- リポジトリに置くことにした
- 更新は 生成して pull request を作成して自動マージする workflow

docs 側の更新方法変更

- pull してきて反映
- bc-setup-all で rurema/generated-documents をとってきて db-* の symlink 作成
- bc-static-all は static html を rsync で反映
- update-rurema-index は今まで通り

今後の予定

- rurema/generated-documents の生成済ファイルは埋め込まれているタグなどの関係で docs.ruby-lang.org 専用 → うまく分離したい
- docs.ruby-lang.org の環境軽量化
 - HTML 生成部分は完了
 - rurema-search は生成されるインデックスだけで 600M 越え (heroku の slug の 500M 制限超過) で静的ファイルのホスティング + Heroku への移行は無理そう
- <https://github.com/ruby/docs.ruby-lang.org>にある ansible の playbook も現状と合わない
ので EC2 インスタンス作り直し?

残作業

- bitclust への型付けをしつつコードリーディングの続き
- kramdown への型付け (まだなければ)
- 開発環境の devcontainer 化 (bitclust 開発者向けと doctree 執筆者向け)
- bitclust の markdown 対応

rurema の markdown 対応

- bitclust に markdown 対応機能追加
- markdown 移行前に doctree の pull request 一掃
- doctree で markdown に一部書き換え
- doctree の書き換えでわかった bitclust で markdown 対応の問題点修正
- rurema-search の markdown 対応

rurema の markdown 対応

- doctree で全面的に markdown 対応
- doctree の RDベース記法のドキュメント削除
- bitclust から RD 対応を削除

その他のやりたいこと

- irb でのドキュメント表示対応
- ドキュメント内部での ruby.wasm での実行対応
- ドキュメント執筆補助ツール (bitclust の tools) の再整備

docs.ruby-lang.org 関連

- (済) rdoc 生成のコンテナ化
- 脆弱性のある古い js の対処(?) (古い jquery などが残っているかどうかなどの確認から)
- (済) ja html 生成のコンテナ化
- GA の削除? (共通 js ファイルにして docs.ruby-lang.org 以外だと空ファイルとかできると良さそう?)
- (済) 更新しない古いバージョンをアーカイブファイルでも保存・配布

docs.ruby-lang.org 関連

- (済) 新しいバージョンも rurema-search で必要ならアーカイブでも配布
- 古いバージョンの db-* の再生成
- HTML 配信元を EC2 から S3 バックエンドか何かに移行(?)
- (途中まで済) rurema-search のコンテナ化かサーバーレス化か何か

直近

- (済) rurema-search で master が 3.4 のようにバージョンで出てきてリンク切れになる問題の対策
- <https://docs.ruby-lang.org/ja/> と <https://docs.ruby-lang.org/en/> のサポート終了バージョンの更新