Engineer

メタデータ

ほかのデータを説明し、見つけられるようにするデータ

どのようなものか

メタデータは、別のデータを対象とするデータです。文書に伴う題名、日付、言語、著者、形式、サイズなどで、人やプログラムは、中身を読む前に、並べ、信用できるかを判断し、処理できます。古い標準ではmeta-dataとハイフンを入れ、日常では、HTTPのヘッダーフィールド、オフィス文書のプロパティー、写真のタグ、Markdownページ先頭のフロントマターなど、より狭い名前でも呼びます。

よくある分類は三つです。記述メタデータは、題名、作成者、主題、要約など、何についてのものかを示します。構造メタデータは、章、ページ、読む順など、部分の組み合わせを示します。管理メタデータは、いつ作成、変更し、誰が行い、どのライセンス、形式かなど、出所と許される用途を示します。2003年にISO 15836となった十五要素のDublin Coreは、最もよく知られた共通の記述項目一覧で、その題名、作成者、日付、言語、権利は、その後のほぼすべての形式にも、別の名前で現れます。

Webページは、層ごとにメタデータを持ちます。最初はHTTP応答です。RFC 9110は、ヘッダーフィールドが「メッセージの意味を変更または拡張し、送信者を説明し、内容を定義し、追加の文脈を与える場合がある」と述べます。`Content-Type`、`Content-Language`、`Last-Modified`は、本文が一バイト届く前から、ページを説明します。HTML内には、`<title>`、`lang`属性、`<meta>`要素があります。後者の名前として、HTML標準は`description`、`author`、`generator`などを定めます。その上には、他者が読む語彙があります。リンクのプレビュー用のOpen Graphプロパティーと、検索エンジンやアシスタント用の構造化データです。

では、データとメタデータを分けるのは何でしょうか。バイトより、読み手です。目録の一件は、本のメタデータであり、目録の側からは、それ自体がデータ一式です。写真の撮影時刻は、眺める人には細部でも、撮影場所を地図にする人

には素材です。メタデータと構造化データの区別は、より狭い問いです。構造化データは、意味を推測せずにプログラムが項目を読めるよう、共通の語彙で書くメタデータです。

概念を判断に生かす

「メタデータとは？」を実際の判断に生かしましょう。その概念を、開発または購入するシステムに結び付けて考えるお手伝いをします。 取り込み時に不適切なデータを検出する規則と検査を設け、各数値の責任者と出所を明らかにします。

サービスを見る お問い合わせ →

用語の背後にある前提を検証する

解説とリンク先の実績を参考に、プロジェクトについての質問を整理してください。ツールを選ぶ前に、前提を一緒に検討できます。 14部門を支える共通インフラ計画 — 14 共通インフラ計画の対象となった研究部門

14部門を支える共通インフラ計画 お問い合わせ →

なぜ重要なのか

各項目は、小さな約束です。重要なのは、人の手入力、版の履歴、データベースの列、そのときファイルシステムが返す値のどこから、値を得るかです。プログラムは、約束を確認せずに動きます。メタデータの目的でもあり、危険でもあります。

キャッシュとクローラーは、それを基に仕事を割り振ります。ページのコピーを持つブラウザーは、受け取った `Last-Modified` の日付を送り返します。変わっていなければ、ページ全体ではなく、短い `304 Not Modified` を受け取ります。RFC 9110は、最終変更を「合理的かつ一貫して決められる」表現には、その日付を送るよう求めます。条件は逆にも働き、一貫した日付がなければ、標準は要求しません。サイトマップの変更日は、取り直す対象を選ぶクローラーにも、同じ役割を果たします。

人と支援技術も頼ります。 `lang` 属性は、スクリーンリーダーに発音規則を伝え、題名は、タブ、ブックマーク、検索結果に表示されます。権利も同じです。パッケージマネージャーが依存先のライセンスを知るのは、依存先自身が宣言するからで、その一覧に基づくライセンス審査の正確さは、元の宣言の正確さそのものです。

すべての項目は、正しさを保つ必要があります。四十項目なら、間違う機会も四十あります。誰も読まないものもあります。HTML標準は、 `keywords` が検索結果のスパムに使われたため、多くの検索エンジンは考慮しないと記します。また、写真の位置や文書の改訂履歴は、本文以上のことを伝え、メタデータから情報が漏れる場合もあります。ずれない出所から値を入れる少数の項目の方が、多数あるより、通常は優れます。

実際の活用例

このサイトの三つのプロジェクトに、メタデータが現れます。それぞれ、サイトマップ、PDF文書群、依存関係のグラフという、違う種類のファイルです。

あるサイトマップは、違うメタデータから変更日を取得し、176ページ中172ページが同じ瞬間を示す結果になりました。それらは内容ではなく、バイトに関するファイル更新時刻で、内容を書き出すたびに、すべて書き直していました。修正後は、明示したフィールド、コミット履歴、ファイルの順で代替元を調べます。落とし穴も記します。フィールド名そのものを一覧に入れなければ、生成器は一度も読みません。今は、ページの元の記録で、各項目を書き、改訂した日と、仕事を実際に行った年を分けます。誰も知らない日付は、「精密だが誤った数値は、ないより悪い」という規則で、空欄にします。記録された実装段階では、一か月の編集後に同期し直すと、変更が生じたのは186ファイル中8ファイルでした。

記録された検証段階では、アクセシビリティ標準に照らして検証した九つの生成PDFが、設定したアクセシビリティプロファイルの106規則すべてに合格しました。作業には、タグ付きの構造、代替テキスト、明示した読み順に加え、メタデータでの文書言語と実際の題名の宣言が含まれます。その記録された検証では、同じ文書が長期保存用プロファイルの146規則のうち、ちょうど一つに失敗しました。ページは、その規則を名指しし、惜しくも届かない結果と呼びます。145規則の成功は、適合ではないからです。

Rustの644クレートのライセンスを、ビルドごとに検査した例では、監査が明示する限界は、メタデータの限界でした。各クレートの宣言したライセンスを読みますが、宣言は間違ったり不足したりするので、誤った宣言をするクレートについて、何も証明しません。意図的に二度失敗させ、検査自体が動くことは示しました。しかし、どんな検査も、読む宣言の正直さまでは示せません。

問題が起きるところ

次の不具合には、共通点があります。それを探さない検査は、すべて成功します。

配備するたび、すべてのページが変更を報告します。 静的ファイルの更新時刻を `Last-Modified` にするWebサーバーや、同じ方法でサイトマップの日付を決める生成器は、コピーした瞬間を伝えます。チェックアウト、書き出し、コンテナービルドは、すべてをコピーします。キャッシュはすべて再検証し、クローラーは

何も学べません。日付は、版の履歴か人が書くフィールドから取り、不明なら、ないままにします。Hugoは、一覧の順に出所を読み、フィールド自身の名前も必要です。

```
enableGitInfo = true

[frontmatter]
date = ["date", ":git", ":fileModTime"]
lastmod = ["lastmod", ":git", ":fileModTime"]
```

題名やリンクのプレビューが文字化けします。 文字コードは、ほかのすべてのバイトに関するメタデータです。宣言より先に文章を見た解析器は、推測する必要があります。HTML標準は、宣言する要素全体が、先頭の1,024バイト内に入ることを求めます。`<head>`の最初に置けば、推測は不要です。このサイトでは、BingのURL Inspectionが、題名の長いダッシュが文字コード宣言より前にあったため、題名とOpen Graphの二つの文字列を、Latin-1として読んだことがあります。HTTPヘッダーもUTF-8を示していましたが、ヘッダーなしでバイトを読む道具には、ページしかありません。今は、文字コード宣言が最初の要素で、ビルドしたすべてのページで、その順を検査します。

同じ事実が、場所によって違います。 題名は、`<title>`、`og:title`、構造化データ、PDFなら文書のプロパティーにあります。言語は、`lang`、`Content-Language`、`og:locale`にあります。別々に入力すればずれ、読む側がそれぞれ別のものを信じます。一つのテンプレートか関数で一度だけ計算し、出力するすべての場所が、その値を読むようにします。

作者が公開するつもりのなかった情報まで、ファイルが伝えます。 写真にはカメラの情報、時刻、多くは位置が入り、オフィス文書には、著者名、編集履歴、テンプレートのパスが入ります。`generator`タグは、ページを作ったソフトウェアを示します。誰も公開すると決めていないものは、書き出すときに除去し、すべての項目を、知らない人が読む文章として扱います。

参考資料

- RFC 9110、第6.3節：ヘッダーフィールド — HTTPが本文より先にメタデータを運ぶ仕組みと、フィールドのできることです。
- RFC 9110、第8.8.2節：Last-Modified — 更新日を送る条件と、未来の日付を置き換える理由です。
- HTML標準：meta要素 — 標準のメタデータ名と、文字コードの宣言規則です。

- DCMIメタデータ語彙 — Dublin Coreの十五要素と、その周りの語彙です。
- Open Graphプロトコル — 各ページに必須の四つのプロパティーと、そのほかの任意項目です。

用語集

実際の制約に合わせて選ぶ

代表的なデータセット、そこから答えを得たい問い、現在のボトルネックをお知らせください。品質、処理量、復旧の測り方を合意できます。

料金を見る お問い合わせ →

用語を実務に結び付ける

「メタデータとは？」がお客様の組織にどう当てはまるかをご相談ください。現在のシステムの具体例が一つあれば、話を始められます。

サービスを見る

welcome@engineer.company +45 50 10 61 84 デンマーク、コペンハーゲン

このページを持ち歩く

連絡先カード

心を込めて丁寧に制作： D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

著作権 © 2025 – 現在 · Engineer ApS · すべての権利を留保します。

印刷元： https://engineer.company/ja/glossary/metadata/