Engineer

Metadatos

Datos que describen otros datos para poder encontrarlos

Qué es

Los metadatos son datos que describen otros datos: el título, fecha, idioma, autor, formato y tamaño que acompañan a un documento para que personas y programas puedan ordenarlo, confiar en él y actuar antes de leerlo. En estándares antiguos aparece la grafía inglesa meta-data, con guion; en el uso diario reciben nombres más concretos: campos de cabecera en HTTP, propiedades de un archivo ofimático, etiquetas de una foto o front matter al principio de una página Markdown.

Una división habitual distingue tres tipos. Los metadatos descriptivos indican de qué trata algo: título, creador, tema, resumen. Los estructurales explican cómo encajan sus partes: capítulos, páginas, orden de lectura. Los administrativos indican de dónde procede y qué puede hacerse con ello: cuándo se creó y modificó, quién lo hizo, con qué licencia y en qué formato. Dublin Core, quince elementos que se convirtieron en el estándar ISO 15836 en 2003, es la lista compartida de campos descriptivos más conocida; su título, creador, fecha, idioma y derechos reaparecen con otros nombres en casi todos los formatos posteriores.

Una página web lleva metadatos en varias capas. Primero llega la respuesta HTTP. RFC 9110 dice que un campo de cabecera «podría modificar o ampliar la semántica del mensaje, describir al remitente, definir el contenido o proporcionar contexto adicional»; `Content-Type`, `Content-Language` y `Last-Modified` describen la página antes de que llegue un solo byte de ella. Dentro del HTML están `<title>`, el atributo `lang` y elementos `<meta>` cuyos nombres define el estándar HTML, como `description`, `author` y `generator`. Sobre ellos hay vocabularios que leen terceros: propiedades Open Graph para las vistas previas de enlaces y datos estructurados para buscadores y asistentes.

¿Qué separa entonces los metadatos de los datos? Más el lector que los bytes. Un registro de catálogo aporta metadatos sobre un libro y constituye el conjunto de datos completo para el catálogo; la hora de captura de una foto es un detalle para quien la mira y materia prima para quien cartografía dónde se toman las fotos. La distinción entre metadatos y datos estructurados es más concreta: los datos estructurados son metadatos escritos con un vocabulario compartido, para que un programa lea un campo sin adivinar qué significa.

Utilice el concepto en una decisión

Aplique «¿Qué son los metadatos?» a una decisión real. Podemos ayudarle a relacionar el concepto con el sistema que está construyendo o comprando. Reglas y comprobaciones que detectan datos incorrectos al entrar y aclaran quién responde de cada cifra y de dónde procede.

Explore nuestros servicios Póngase en contacto →

Compruebe los supuestos detrás del término

Utilice la explicación y los trabajos enlazados para formular preguntas sobre su proyecto. Podemos examinar los supuestos con usted antes de elegir una herramienta. Un plan de infraestructura para 14 departamentos — 14 departamentos de investigación con un único plan de infraestructura

Un plan de infraestructura para 14 departamentos Póngase en contacto →

Por qué importa

Cada campo es una pequeña promesa; lo decisivo es de dónde obtiene su valor: de una persona que lo escribe, del historial de versiones, de una columna de base de datos o de lo que diga el sistema de archivos. Los programas actúan según esas promesas sin comprobarlas. Esa es la utilidad de los metadatos y también su riesgo.

Las cachés y los rastreadores distribuyen su trabajo en función de ellos. Un navegador que conserva una copia devuelve la fecha `Last-Modified` que recibió; si nada ha cambiado, obtiene un breve `304 Not Modified` en lugar de otra copia de la página. RFC 9110 pide al servidor enviar esa fecha para toda representación cuya última modificación «pueda determinarse de forma razonable y coherente». La condición funciona en ambos sentidos: si no existe una fecha coherente, el estándar no la pide. La fecha de modificación de un mapa del sitio cumple la misma función para un rastreador que decide qué volver a descargar.

También dependen de ellos las personas y las tecnologías de asistencia. El atributo `lang` indica al lector de pantalla qué reglas de pronunciación aplicar; el título aparece en pestañas, marcadores y resultados de búsqueda. Los derechos viajan del mismo modo: un gestor de paquetes solo conoce la licencia de una dependencia porque esta la declara; una revisión de licencias basada en esa lista tiene exactamente la calidad de las declaraciones que la sustentan.

Cada campo debe seguir siendo cierto: una página con cuarenta campos tiene cuarenta ocasiones de equivocarse. Algunos no los lee nadie; el estándar HTML señala que muchos buscadores no consideran `keywords` porque se usó para manipular los resultados. Los metadatos también revelan información: la posición de una foto o el historial de revisiones de un documento pueden contar más que su contenido. Suele ser mejor tener menos campos, cada uno alimentado por un origen que no pueda desajustarse, que muchos.

Cómo se aplica en la práctica

Los metadatos aparecen en tres proyectos de este sitio, cada vez en un tipo de archivo distinto: un mapa del sitio, un conjunto de PDF y un grafo de dependencias.

El mapa de un sitio tomaba las fechas de última modificación de los metadatos equivocados: 172 de sus 176 páginas acabaron compartiendo un instante. Eran las marcas de tiempo de modificación de los archivos, metadatos sobre los bytes en lugar del contenido, y cada exportación de contenido volvía a escribirlas todas. La corrección recorre una cadena de alternativas: primero un campo explícito, después el historial de commits y, por último, el archivo. También documenta una trampa: el nombre literal del campo debe aparecer en la lista o el generador nunca lo lee. Los registros que sustentan las páginas conservan ahora el día de redacción y revisión de cada entrada separado de los años en que se realizó el trabajo; una fecha desconocida queda vacía, según la regla de que «una cifra precisa pero incorrecta es peor que su ausencia». En la etapa de implementación documentada, una nueva sincronización tras un mes de cambios modificaba 8 archivos de 186.

En la etapa de validación documentada, nueve documentos PDF generados y evaluados contra el estándar de accesibilidad superaron las 106 reglas configuradas del perfil de accesibilidad. El trabajo incluyó un idioma declarado y un título real en los metadatos, además de estructura etiquetada, texto alternativo y un orden de lectura explícito. En esa ejecución documentada, los mismos documentos incumplieron exactamente una regla de las 146 del perfil de archivo. La página lo describe como un resultado cercano, con la regla fallida identificada, porque 145 reglas superadas no equivalen a conformidad.

Al comprobar las licencias de 644 crates de Rust en cada compilación, el límite declarado de la auditoría era un límite de los metadatos. Lee la licencia que declara cada crate; las declaraciones pueden ser incorrectas o incompletas y la

comprobación no demuestra nada sobre un crate que se describa mal. Se provocó su fallo dos veces, lo que demuestra que la comprobación funciona; nada que haga una comprobación puede demostrar que las declaraciones que lee sean honestas.

Dónde surgen los problemas

Los fallos siguientes comparten una característica: se superan todas las comprobaciones que no los busquen.

Todas las páginas informan de un cambio después de cada despliegue. Un servidor que envía la marca de tiempo de modificación de un archivo estático como `Last-Modified`, o un generador que fecha así el mapa del sitio, anuncia cuándo se copió el archivo. Obtener una copia del repositorio, exportar o construir un contenedor los copia todos. Las cachés vuelven a validar todo y los rastreadores no aprenden nada. Obtenga las fechas del historial de versiones o de un campo escrito por una persona y deje ausente una fecha desconocida. En Hugo se recorren los orígenes de una lista en orden; el nombre del propio campo debe figurar en ella:

```
enableGitInfo = true

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

El título o la vista previa de un enlace muestra caracteres dañados. La codificación de caracteres es un metadato que describe todos los demás bytes; si el analizador encuentra texto antes de la declaración, debe adivinar. El estándar HTML exige que el elemento que la declara quepa por completo en los primeros 1.024 bytes; colocarlo primero en `<head>` elimina la necesidad de adivinar. En este sitio, URL Inspection de Bing leyó una vez el título y las dos cadenas Open Graph como Latin-1 porque una raya del título aparecía antes de la declaración de codificación. La cabecera HTTP también indicaba UTF-8, pero una herramienta que lea los bytes sin cabecera solo dispone de la página. La declaración es ahora el primer elemento y una comprobación confirma ese orden en cada página compilada.

El mismo dato dice cosas distintas en lugares distintos. El título de una página está en `<title>`, en `og:title`, en sus datos estructurados y, para un PDF, en las propiedades del documento; el idioma aparece en `lang`, `Content-Language` y `og:locale`. Si se escriben por separado, divergen y cada consumidor cree una versión distinta. Calcule cada valor una vez, en una plantilla o función, y haga que todos los lugares donde se imprime lean ese valor.

El archivo dice más de lo que su autor quería publicar. Las fotos incluyen información sobre la cámara, la hora y, a menudo, la posición; los documentos ofimáticos llevan nombres de autores, historial de cambios y rutas de plantillas; una etiqueta `generator` identifica el software que construyó la página. Elimine al exportar lo que nadie decidió publicar y trate cada campo como texto que leerá una persona desconocida.

Lecturas adicionales

- RFC 9110, sección 6.3: campos de cabecera — cómo transporta HTTP los metadatos antes del contenido y qué puede hacer un campo de cabecera.
- RFC 9110, sección 8.8.2: Last-Modified — cuándo debería enviar el servidor una fecha de modificación y por qué se sustituye una fecha futura.
- Estándar HTML: el elemento meta — los nombres de metadatos estándar y las reglas para declarar una codificación de caracteres.
- Términos de metadatos DCMI — los quince elementos Dublin Core y el vocabulario construido alrededor de ellos.
- Protocolo Open Graph — sus cuatro propiedades obligatorias en cada página y las opcionales que las acompañan.

Glosario

Elija según sus restricciones reales

Traiga un conjunto de datos representativo, las preguntas que debe responder y el cuello de botella actual. Podemos acordar cómo medir la calidad, la capacidad de procesamiento y la recuperación.

Vea cuánto cuesta Póngase en contacto →

Haga práctica la terminología

Pregunte cómo se aplica «¿Qué son los metadatos?» a su organización. Basta un ejemplo concreto de su sistema actual para iniciar la conversación.

Explore nuestros servicios

welcome@engineer.company +45 50 10 61 84 Copenhague, Dinamarca

Llévese esta página

Tarjeta de contacto

Creado con esmero y cariño por D.

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

Derechos de autor © 2025 – actualidad · Engineer ApS · Todos los derechos reservados.

Impreso desde https://engineer.company/es/glossary/metadata/