Engineer

Datos estructurados

Datos de una página escritos para las máquinas

Qué es

La página de contacto de una empresa explica en una frase dónde está la oficina; los datos estructurados lo repiten como una `PostalAddress` con calle, código postal y país, para que un programa que lea la página no tenga que adivinar qué palabra es la localidad. Consisten en repetir lo que una página expresa en prosa mediante enunciados tipados en un vocabulario acordado entre quienes publican y quienes consumen los datos, para que los buscadores y otros programas puedan leer los hechos sin interpretar la disposición visual de la página.

El vocabulario suele ser schema.org, que la documentación de Google identifica como el que utiliza la mayoría de los datos estructurados de su buscador. Lo fundaron Google, Microsoft, Yahoo y Yandex, y desde abril de 2015 su principal foro es un grupo comunitario del W3C. Define tipos, como `Organization`, `Person`, `Service`, `Review` o `DefinedTerm`, y las propiedades que puede tener cada uno, como `name`, `address`, `author` o `price`.

Los enunciados pueden escribirse con tres sintaxis. Microdata y RDFa añaden atributos a los propios elementos HTML visibles. JSON-LD mantiene los enunciados aparte, en un bloque JSON dentro de un elemento `script` de tipo `application/ld+json`, y su versión 1.1 es una Recomendación del W3C publicada el 16 de julio de 2020. Google acepta las tres por igual y recomienda JSON-LD por ser la más fácil de mantener a gran escala.

```
<script type="application/ld+json">
    {
        "@context": "https://schema.org",
        "@type": "Organization",
        "@id": "https://example.com/#organization",
        "name": "Example Ltd",
```

```
        "url": "https://example.com/"
    }
</script>
```

La línea más importante de ese bloque es `@id`. JSON-LD utiliza ese campo para identificar un nodo; un nodo sin él es un nodo en blanco, sin un identificador independiente del documento. Los identificadores de nodos en blanco pueden compartirse entre bloques de un mismo documento. Con un identificador utilizable globalmente, cualquier otro bloque de cualquier otra página puede escribir `"publisher": {"@id": "https://example.com/#organization"}`, lo que la especificación denomina referencia a un nodo: un puntero a algo descrito en otro lugar, en vez de una nueva descripción de ello. Los identificadores convierten un conjunto disperso de bloques en un único grafo.

Un concepto cercano es el de los metadatos. El título, la descripción y las etiquetas Open Graph de la cabecera describen la página como documento, mientras que los datos estructurados describen aquello de lo que trata: la empresa que está detrás, la persona que la escribió, el servicio que ofrece o la palabra que define.

Utilice el concepto en una decisión

Aplique «¿Qué son los datos estructurados?» a una decisión real. Podemos ayudarle a relacionar el concepto con el sistema que está construyendo o comprando. Un sitio o una aplicación que los asistentes de IA de sus clientes puedan encontrar, leer y utilizar, con comprobaciones en cada versión en lugar de una promesa inicial.

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. Sitio Hugo multilingüe en 3 idiomas — 108 archivos de plantillas para tres idiomas y cuatro formatos

Sitio Hugo multilingüe en 3 idiomas Póngase en contacto →

Por qué importa

La decisión es cuánto decir, y la respuesta honesta suele ser menos de lo que permite el vocabulario, pero de forma completa. Nada en la sintaxis impide rellenar todas las propiedades que ofrece schema.org. Las propias directrices de Google apuntan en sentido contrario: proporcionar menos propiedades recomendadas, completas y exactas, en lugar de todas las posibles con datos parciales o erróneos, y no marcar nada que el lector no pueda ver en la página, aunque sea cierto.

El motivo es que cada hecho marcado es una segunda copia de algo que ya dice la página, y una segunda copia puede divergir. Un precio modificado en la tabla visible y olvidado en el JSON, una valoración que solo ven las máquinas, una dirección escrita de forma ligeramente distinta en cada página: todo ello hace que la página diga una cosa a las personas y otra a los programas. El coste de los datos estructurados está menos en escribirlos que en mantenerlos fieles a la realidad, así que conviene preguntar qué enunciados puede respaldar la página y si proceden de la misma fuente que el texto visible.

Lo que se obtiene es legibilidad. Un buscador o agente capaz de vincular cien menciones a una sola empresa sabe más que otro que lea cien nombres sin relación. Las páginas que siguen las directrices pueden optar a resultados enriquecidos, es decir, resultados con información adicional, aunque cumplir los requisitos no garantiza ninguno. Lo que cambia es cuánto entiende una máquina de aquello de lo que trata la página.

Cómo se aplica en la práctica

En sus créditos, el sitio se compromete a marcar únicamente lo que sea cierto y visible en la página, y la historia de sus datos estructurados muestra por qué hacía falta dejar esa frase por escrito.

Antes de la revisión de julio de 2026, 963 bloques repetían los datos de la empresa 1.671 veces, sin ningún identificador estable, por lo que ningún consumidor tenía fundamento para tratar la organización de una página como la misma de otra. El grafo se apoya ahora en tres identificadores creados a partir del origen del sitio, para la organización, el sitio web y la fundadora; todos los demás bloques los referencian mediante `@id` en lugar de volver a describirlos. Los identificadores son iguales en todos los idiomas: un comentario de la plantilla explica que un identificador danés distinto declararía una segunda organización con el mismo número de registro empresarial. En aquella etapa se registraron dieciséis tipos, y las listas identificaban sus elementos solo por URL, porque detallar las 107 entradas que tenía entonces el índice del portafolio habría aumentado el peso de la página comprimida un 52 %.

Varios hechos se leen de la misma fuente que la página visible. Los precios de los servicios proceden del mismo archivo de datos que los precios en pantalla, por lo que a la máquina y al lector se les ofrece la misma tarifa. El periodo que abarca un trabajo procede de la plantilla parcial que escribe la frase sobre él. Las rutas de navegación siguen el recorrido visible, y las posiciones de las listas siguen el

orden en pantalla, nunca una clasificación publicada solo para las máquinas. El nodo de la propia página utiliza la misma dirección que su enlace canónico, de modo que el marcado y la cabecera hacen una misma afirmación.

Lo que se omite es tan deliberado como lo que se declara. Los textos que la empresa escribe sobre su propio trabajo no se marcan como reseñas, porque eso inventaría un autor para sus propias palabras. La plantilla de testimonios solo utiliza nodos Review para las entradas publicadas con consentimiento concedido, sin añadir una puntuación. Cada página del glosario, incluida esta, lleva un `DefinedTerm` cuya descripción es la definición de una línea que aparece en la lista, de modo que la máquina lee la misma definición que el lector.

Una comprobación del sitio generado verifica que todos los bloques se puedan analizar, que todos los identificadores se resuelvan y que se haya generado cada URL que menciona el grafo; por tanto, un grafo que apunte a una página inexistente hace fallar el control de calidad obligatorio. La misma página de créditos enumera Schema Markup Validator para leer los datos y Rich Results Test de Google para ver qué hace un buscador concreto con ellos.

Dónde surgen los problemas

Los resultados enriquecidos siguen sin llegar aunque todos los validadores den el visto bueno. El marcado es JSON-LD válido y utiliza tipos reales de schema.org, pero la página sigue apareciendo como un resultado ordinario. Los datos describen cosas que el lector no puede ver, como una puntuación que nadie dio, preguntas ocultas o un precio distinto del que aparece en pantalla; Google advierte de que incumplir sus directrices puede dejar una página fuera de los resultados enriquecidos, por muy correcta que sea la sintaxis. Marque solo lo que muestre la página y genere ambas versiones desde una misma fuente para que no diverjan.

Cientos de bloques describen una empresa y nada los conecta. Cada página repite el nombre, el logotipo y la dirección de la organización, a veces con pequeñas diferencias, y corregir un detalle obliga a encontrarlos todos. Los nodos son anónimos, o cada página y cada idioma crean un identificador propio para la misma entidad. Asigne a cada entidad real un único `@id` construido a partir de una URL estable con un fragmento, descríbala una sola vez y referénciela en los demás lugares.

Un cambio de plantilla rompe el marcado de todas las páginas a la vez. Una comilla suelta, una coma final o un valor vacío invalidan todos los bloques y los consumidores no pueden utilizar ninguno. La guía de Google advierte de que las plantillas pueden romper el marcado después del despliegue. El JSON se compuso como texto, pegando valores entre llaves, de modo que un campo ausente deja una cadena vacía o una estructura rota. Construya el enunciado como una estructura de datos y serialícela, como hacen las plantillas de este sitio con `jsonify` de Hugo; omita las propiedades sin valor en lugar de emitirlas vacías y analice todos los bloques de la salida generada.

El grafo menciona páginas que faltan. Las URL del marcado llevan a un 404, a un servidor de preproducción o a una réplica del sitio. Las direcciones se escribieron a mano o se construyeron con una regla distinta de la utilizada para los enlaces y la dirección canónica. Obtenga cada URL del marcado con la misma función que produce la dirección canónica de la página y compruebe que se haya generado cada una.

Lecturas adicionales

- schema.org — el vocabulario compartido de tipos y propiedades, con una página para cada tipo.
- JSON-LD 1.1 — la Recomendación del W3C para la sintaxis, incluidos los identificadores y las referencias a nodos.
- schema.org: DefinedTerm — el tipo que utiliza cada página del glosario para describir la palabra que explica.
- Google Search Central: introducción a los datos estructurados — las tres sintaxis, la regla de visibilidad y cómo se prueban los resultados enriquecidos.
- Schema Markup Validator — contrasta los datos estructurados de cualquier página con el vocabulario de schema.org.

Glosario

Elija según sus restricciones reales

Describa el flujo de trabajo, las interfaces existentes y los usuarios que necesitan el cambio. Podemos definir una entrega comprobable y un traspaso que permita a su equipo mantener lo entregado.

Vea cuánto cuesta Póngase en contacto →

Haga práctica la terminología

Pregunte cómo se aplica «¿Qué son los datos estructurados?» 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/structured-data/