Engineer

Структуровані дані

факти про сторінку, записані для машин

Що це таке

Контактна сторінка компанії одним реченням каже, де розташований офіс. Структуровані дані кажуть це ще раз, як `PostalAddress` з вулицею, поштовим індексом і країною, щоб програмі, яка читає сторінку, не доводилося вгадувати, котре слово означає місто. Так називають практику повторювати те, що сторінка викладає прозою, у вигляді типізованих тверджень спільним словником, про який домовилися видавці й споживачі даних, щоб пошукові системи та інші програми могли читати факти, не розбираючи верстку.

Цей словник здебільшого schema.org: документація Google називає його тим, яким користується більшість структурованих даних для її пошуку. Його заснували Google, Microsoft, Yahoo і Yandex, а з квітня 2015 року головним майданчиком для нього є спільнотна група W3C. Він визначає типи, як-от `Organization`, `Person`, `Service`, `Review` чи `DefinedTerm`, і властивості, які може мати кожен із них, як-от `name`, `address`, `author` чи `price`.

Твердження можна записати трьома синтаксисами. Microdata і RDFa додають атрибути безпосередньо до видимих елементів HTML. JSON-LD тримає твердження окремо, як блок JSON в елементі `script` типу `application/ld+json`; версія 1.1 цього формату стала рекомендацією W3C 16 липня 2020 року. Google приймає всі три на рівних і радить JSON-LD як найпростіший у супроводі на великих обсягах.

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

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

Найважливіший рядок у цьому блоці — `@id` . JSON-LD ідентифікує ним вузол, а вузол без нього анонімний: у термінах специфікації це порожній вузол (blank node) без ідентифікатора, незалежного від документа. Ідентифікатори порожніх вузлів можуть бути спільними для блоків одного документа. Маючи глобальний ідентифікатор, будь-який інший блок на будь-якій іншій сторінці може написати `"publisher": {"@id": "https://example.com/#organization"}` , і специфікація називає це посиланням на вузол: вказівкою на річ, описану деінде, замість нового її опису. Саме ідентифікатори перетворюють розкидані блоки на один граф.

Сусіднє поняття — метадані. Заголовок, опис і теги Open Graph у `head` сторінки описують сторінку як документ, а структуровані дані описують те, про що сторінка: компанію, що за нею стоїть, людину, яка її написала, послугу, яку вона пропонує, чи слово, яке вона пояснює.

Застосуйте поняття до рішення

Застосуйте «Що таке структуровані дані?» до конкретного рішення. Ми можемо пов'язати поняття із системою, яку ви створюєте або купуєте. Сайт чи застосунок, який ШІ-асистенти ваших клієнтів можуть знайти, прочитати й використати, і це перевіряється з кожним випуском, а не обіцяється один раз.

Переглянути наші послуги Зв'язатися з нами →

Перевірте припущення за терміном

Використайте пояснення та пов'язані роботи, щоб сформулювати запитання для проєкту. Ми можемо перевірити припущення разом до вибору інструмента. Побудувала сайт на Hugo 3 мовами — 108 файлів шаблонів обслуговують три мови й чотири формати

Побудувала сайт на Hugo 3 мовами Зв'язатися з нами →

Чому це важливо

Рішення тут одне: скільки сказати. Чесна відповідь зазвичай така: менше, ніж дозволяє словник, але повністю. Синтаксис нічим не заважає видавцеві заповнити кожну властивість, яку пропонує schema.org. Власні настанови Google радять протилежне: краще менше рекомендованих властивостей, зате повних і точних, ніж усі можливі з неповними чи хибними даними, і не розмічати того, чого читач не бачить на сторінці, навіть коли це правда.

Причина в тому, що кожен розмічений факт — друга копія того, що сторінка вже каже, а друга копія може розійтися з першою. Ціна, виправлена у видимій таблиці й забута в JSON, оцінка, яку бачать лише машини, адреса, щоразу набрана трохи інакше: через кожну з них сторінка каже людям одне, а програмам інше. Основна праця полягає не в тому, щоб написати структуровані дані, а в тому, щоб вони лишалися правдивими. Тож корисно питати, за які твердження сторінка може поручитися і чи походять вони з того самого джерела, що й видимий текст.

Виграш — розбірливість. Пошукова система чи агент, що можуть звести сотню згадок до однієї компанії, знають більше, ніж ті, хто читає сотню не пов'язаних між собою назв. Сторінки, які дотримуються настанов, можуть отримати право на розширені результати (rich results), тобто показ із додатковими подробицями, але саме право їх не гарантує. Змінюється те, наскільки добре машина розуміє, про що сторінка.

Як це виглядає на практиці

Сторінка подяк цього сайту обіцяє розмічати тільки правдиве й видиме, а історія його структурованих даних показує, чому це речення довелося записати.

До переробки в липні 2026 року компанію було описано 1 671 раз у 963 блоках, і в жодного з них не було сталого ідентифікатора, тож споживач даних не мав підстав вважати організацію на одній сторінці тією самою, що й на іншій. Тепер граф спирається на три ідентифікатори, утворені від origin сайту: для організації, для вебсайту й для засновниці. Усі інші блоки посилаються на них через `@id`, а не описують їх знову. Ідентифікатори однакові в усіх мовах: коментар у шаблоні пояснює, що окремий данський ідентифікатор оголосив би другу організацію з тим самим реєстраційним номером. На тому етапі було зафіксовано шістнадцять типів, а списки називали свої елементи лише через URL, бо повний запис 107 позицій, які тоді були на головній сторінці портфоліо, зробив би стиснену сторінку на 52% важчою.

Кілька фактів беруться з того самого джерела, що й видима сторінка. Ціни на послуги походять із того самого файлу даних, що й ціни на екрані, тож машина й читач бачать однаковий тариф. Період, який охоплює робота, береться з того самого partial-шаблону, що пише речення про цей період. Хлібні крихти повторюють шлях, який бачить читач, а позиції в списках

відповідають порядку на екрані, а не рейтингові, опублікованому лише для машин. Власний вузол сторінки має ту саму адресу, що й її канонічне посилання, тож розмітка й `head` стверджують одне й те саме.

Відмови так само свідомі, як і твердження. Власні описи робіт компанії не позначено як `Review`, бо так з'явився б вигаданий автор її власних слів. Шаблон відгуків використовує вузли `Review` лише для записів, опублікованих за наданою згодою, не додаючи оцінки. Кожна стаття Словника, і ця теж, несе `DefinedTerm`, чий опис — однорядкове пояснення зі списку, тож визначення, яке читає машина, те саме, що бачить читач.

Перевірка зібраного сайту вимагає, щоб кожен блок розбирався, кожен ідентифікатор вказував на наявний вузол, а кожна URL, яку називає граф, справді була зібрана, тож ворота якості не пропускають граф, що вказує на сторінку, якої немає. Та сама сторінка подяк називає Schema Markup Validator, щоб прочитати дані, і Rich Results Test від Google, щоб побачити, що з ними робить одна пошукова система.

Де виникають проблеми

Розширені результати лишаються недосяжними, хоча всі валідатори приймають розмітку. Розмітка — коректний JSON-LD зі справжніми типами schema.org, а сторінка й далі показується як звичайний результат. Дані описують те, чого читач не бачить: оцінку, якої ніхто не ставив, запитання, приховані на сторінці, чи ціну, що відрізняється від тієї, що на екрані, а Google попереджає, що розмітка, яка порушує його настанови, може позбавити сторінку права на розширені результати, хоч який чистий синтаксис. Розмічайте лише те, що сторінка показує, і генеруйте розмітку й видимий текст з одного джерела, щоб вони не могли розійтися.

Сотні блоків описують одну компанію, і кожен описує її наново. Кожна сторінка повторює назву, логотип і адресу організації, часом із дрібними відмінностями, і щоб виправити одну деталь, треба знайти всі її копії. Вузли анонімні, або кожна сторінка й кожна мова видає власний ідентифікатор для тієї самої сутності. Дайте кожній реальній речі один `@id`, побудований зі сталої URL з фрагментом, опишіть її один раз і посилайтеся на неї в усіх інших місцях.

Одна зміна шаблону ламає розмітку на всіх сторінках водночас. Зайві лапки, кома після останнього елемента чи порожнє значення роблять кожен блок недійсним, і споживачі не можуть використати з нього нічого. Посібник Google попереджає, що шаблонізація може зламати розмітку після розгортання. JSON складали як текст, вставляючи значення між дужками, тож відсутнє поле лишає порожній рядок або зламану структуру. Будуйте твердження як структуру даних і серіалізуйте її, як це роблять шаблони цього сайту через `jsonify` у Hugo, пропускайте властивості без значення, а не виводьте їх порожніми, і розбирайте кожен блок у зібраному результаті.

Граф називає сторінки, яких бракує. URL у розмітці ведуть на 404, на тестовий хост чи на дзеркальну копію. Адреси набрано вручну або побудовано за іншим правилом, ніж посилання й канонічну URL. Виводьте кожну URL у розмітці тією самою функцією, що дає канонічну адресу сторінки, і перевіряйте, що кожну з них зібрано.

Що почитати далі

- schema.org — спільний словник типів і властивостей з окремою сторінкою для кожного типу.
- JSON-LD 1.1 — рекомендація W3C для синтаксису, зокрема ідентифікатори вузлів і посилання на вузли.
- schema.org: DefinedTerm — тип, яким кожна стаття Словника описує слово, що його пояснює.
- Google Search Central: introduction to structured data — три синтаксиси, правило видимості й те, як перевіряти розширені результати.
- Schema Markup Validator — читає структуровані дані будь-якої сторінки й звіряє їх зі словником schema.org.

Словник

Обирайте за фактичними умовами

Опишіть робочий процес, наявні інтерфейси та користувачів, яким потрібна зміна. Ми можемо визначити перевірюваний результат і спосіб передачі, який дасть вашій команді змогу підтримувати його.

Дивитися, скільки це коштує Зв'язатися з нами →

Зробіть термінологію практичною

Запитайте, як «Що таке структуровані дані?» стосується вашої організації. Конкретного прикладу з поточної системи достатньо для початку розмови.

Переглянути наші послуги

welcome@engineer.company +45 50 10 61 84 Копенгаген, Данія

Візьміть цю сторінку із собою

Візитівка

Дбайливо зроблено з любов'ю — D.

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

Авторське право © 2025 – дотепер · Engineer ApS · Усі права захищено.

Надруковано з https://engineer.company/uk/glossary/structured-data/