

MICROFORMATS & MICRODATA

Štruktúrované dáta na webe

Bc. Adam Kundracik

Prečo štruktúrovať webový obsah?

Web sa od svojho vzniku vyvíja

Pôvodne bol len kolekciou dokumentov, dnes je to dynamická sieť prepojených informácií.

Potreba zabezpečenia aby počítače rozumeli obsahu stránok

Výzva - Strojové porozumenie webu

Chceme aby akákoľvek aplikácia , web presne pochopili čo na našej stránke máme

Prichádza prvá idea štruktúrovaných dát

Microformats

Vznik okolo roku 2004

Prvá verzia

Cieľ: Rozšíriť výrazové možnosti HTML

Využívanie existujúcich atribútov “class” a “rel”

Štruktúrované publikovanie dát

Publikovanie informácií o: Ľuďoch, Miestach, Udalostiach, Recenziách, Produktoch

Triedy v microformats

hCard (Kontaktné informácie)

vcard – Hlavná trieda pre kontakt
fn - Formatted Name (celé meno)
n – Štrukturované meno

- given-name - Krsné meno
- family-name - Priezvisko
- additional-name – Prostredné meno
- honorific-prefix - Titul pred menom (Dr., Mgr.)
- honorific-suffix - Titul za menom(PhD., CSc.)

nickname
email
tel – Telefónne číslo
adr - Adresa

- street-address
- locality - mesto
- region
- postal-code
- country-name

org - Organizácia
title – Pracovná pozícia
bday - Dátum narodenia
url - Webová stránka

hCalendar (Udalosti)

vevent – Hlavná trieda pre udalosť
summary - Názov udalosti
dtstart - Dátum a čas začiatku
dtend - Dátum a čas konca
location - Miesto konania
description
url - Odkaz na udalosť
duration - Dĺžka trvania

hReview (Recenzie)

hreview - Hlavná trieda pre recenziu
summary - Nadpis recenzie
reviewer - Autor recenzie
dtreviewed - Dátum recenzie
rating – Hodnotenie
item - Recenzovaný predmet
type - Typ recenzie

hProduct (Produkt)

hproduct - Hlavná trieda pro produkt
fn - Název produktu
photo - Fotografie
description
price
brand
url

- hAtom -  [Readability](#) (hNews), numerous  [Spinn3r](#) customers
- hCard -     [Readability](#)  [H2VX](#)  [Firefox Operator](#)  [Microformats for Chrome](#)
- hCalendar -    [H2VX](#)  [Firefox Operator](#)  [Microformats for Chrome](#)
- hProduct -   
- hRecipe -     [Microformats for Chrome](#)
- hResume -  [Guardian Jobs](#),  [Madgex Labs](#) library clients
- hReview -     [Microformats for Chrome](#)
- hReview-aggregate -    [Microformats for Chrome](#)
- rel-me - , [RelMeAuth/IndieAuth](#) **Web Sign-in**
- rel-author - 

```
<span class="vcard">  
  <a class="fn n url"      href="http://waterpigs.co.uk">  
    Barnaby Walters  
  </a>  
</span>
```

MICROFORMATS2

Vylepšená verzia
pôvodných
microformátov

Zjednodušenie
publikovania a
spracovania dát

Otvorený
komunitný
projekt

Viacero open-
source
implementácií

Príklady Microformats2 tried

hCard v Microformats2

- `h-card` - Hlavná trieda pre kontakt
- `p-name` - Meno
- `p-given-name` - Krstné meno
- `p-family-name` - Priezvisko
- `p-org` - Organizácia
- `p-job-title` - Pracovná pozícia
- `p-adr` - Adresa
- `p-street-address` - Ulica
- `p-locality` - Mesto
- `p-region` - Región
- `p-postal-code` - PSČ
- `p-tel` - Telefón
- `u-email` - Email
- `u-url` - Webová stránka

hEvent v Microformats2

- `h-event` - Hlavná trieda pre udalosť
- `p-name` - Názov udalosti
- `dt-start` - Začiatok udalosti
- `dt-end` - Koniec udalosti
- `p-location` - Miesto konania
- `p-description` - Popis

hReview v Microformats2

- `h-review` - Hlavná trieda pre recenziu
- `p-name` - Nadpis recenzie
- `p-reviewer` - Autor recenzie
- `dt-reviewed` - Dátum recenzie
- `p-rating` - Hodnotenie
- `p-item` - Recenzovaný predmet

h-feed

- h-feed: Kontajner preznam položiek
- p-name: Názov feedu
- h-entry: Jednotlivá položka
- p-name: Názov položky
- dt-published: Dátum publikácie
- p-summary: Krátky popis
- u-url: Odkaz na celú položku
- p-author: Autor položky

MICROFORMATS VS MICROFORMATS2:

Rozdiely

Microformats:

- Používali pevne definované triedy
- Menej konzistentná štruktúra
- Komplikovanejšie parsovanie
- Príklad: `class="vcard"`, `class="fn"`

Microformats2:

- Zavedenie konzistentných prefix pre triedy
 - `h-*` pre hlavné typy (hlavné objekty)
 - `p-*` pre vlastnosti
 - `u-*` pre URL
 - `dt-*` pre dátumy a časy

```
<a class="h-card u-url"  
href="http://waterpigs.co.uk">  
    Barnaby Walters  
</a>
```

MICROFORMATS VS MICROFORMATS2

Microformats formát:

```
<span class="vcard">
```

```
  <a class="fn n url" href="http://waterpigs.co.uk">Barnaby  
  Walters</a>
```

```
</span>
```

Microformats2 formát:

```
<a class="h-card u-url" href="http://waterpigs.co.uk">Barnaby Walters</a>
```

Prepíš na Microformats v1 Kontrola na <https://pin13.net/mf2/>

```
<!DOCTYPE html>

<html lang="sk">

<head>

  <title>Seminár o webových technológiách</title>

</head>

<body>

  <div>

    <h1>Seminár o webových technológiách</h1>

    <p>Datum konaní: 15. mája 2024 od 9:00 do 16:00</p>

    <p>Miesto: Stredná škola informatiky, Učebňa A1</p>

  </div>

</body>

</html>
```

Úloha: Rovnaké HTML do Microformats2

Kontrola v parseri
<https://pin13.net/mf2/>

Výstup z parseru

```
{
  "items": [
    {
      "type": ["h-event"],
      "properties": {
        "name": ["Seminár o webových technológiách"],
        "location": ["Stredná škola informatiky, Učebňa A1"],
        "start": ["2024-05-15T09:00"],
        "end": ["2024-05-15T16:00"]
      }
    },
    {
      "rels": {},
      "rel-urls": {},
      "debug": {
        "package": "https://packagist.org/packages/mf2/mf2",
        "source": "https://github.com/indieweb/php-mf2",
        "version": "v0.5.0",
        "note": [
          "This output was generated from the php-mf2 library available at",
          "https://github.com/indieweb/php-mf2",
          "Please file any issues with the parser at https://github.com/indieweb/php-",
          "mf2/issues",
          "Using the Masterminds HTML5 parser"
        ]
      }
    }
  ]
}
```

Úloha pre vás

Osobná vizitka

- Vytvorte osobnú kontaktnú kartu pomocou:
 - Microformats v1 (vcard)
 - Microformats v2 (h-card)
- Povinné prvky:
 - Meno a priezvisko
 - Kontaktné informácie
 - Organizácia/škola

MICRODATA

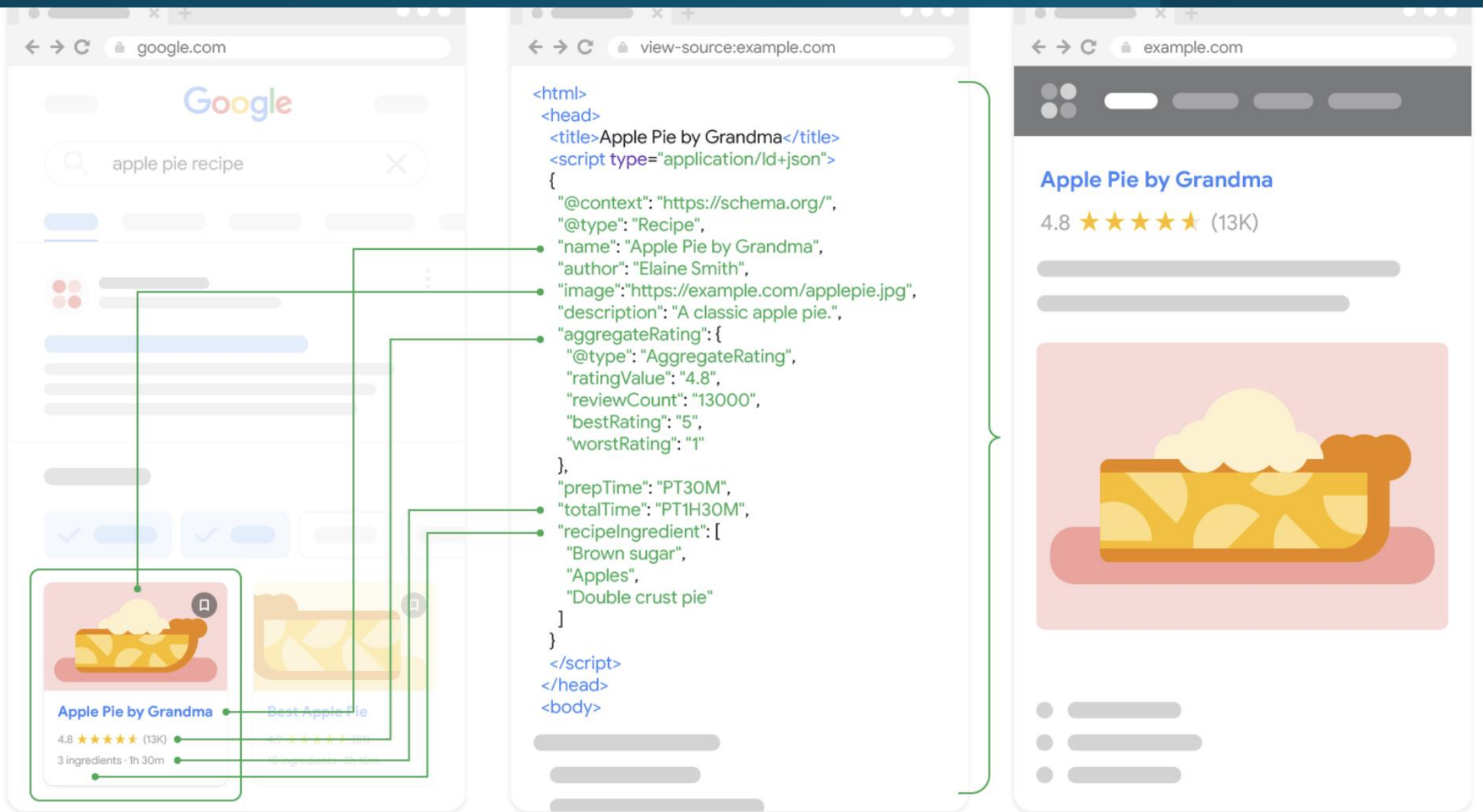
- HTML štandard
- spôsob štruktúrovania webového obsahu, ktorý umožňuje strojom lepšie rozumieť významu informácií na webových stránkach.
- **Sémantické značkovanie**
 - Pridávanie významových atribútov priamo do HTML
 - Používanie špeciálnych atribútov:
 - itemscope
 - itemtype
 - Itemprop
- Jednoduchá čitateľnosť pre stroje aj ľudí
- **Typy informácií**
 - Osoby
 - Organizácie
 - Udalosti
 - Produkty
 - Recenzie
 - Miesta

Spôsoby zavedenia štrukturovaných dát na web

JSON-LD

Microdata

RDFa



Štruktúra Microdata

1. **itemscope**

- Označuje, že element obsahuje sadu mikrodát
- Definuje začiatok samostatného informačného kontextu
- Signalizuje začiatok popisu konkrétneho objektu

2. **itemtype**

- Určuje presný typ objektu podľa schema.org
- Definuje schému/kategóriu pre daný objekt
- Poskytuje kontextové informácie o type entity

3. **itemprop**

- Definuje konkrétne vlastnosti objektu
- Popisuje jednotlivé atribúty entity
- Umožňuje pridať špecifické informácie

Povolené je aj vnáranie objektov

```
<div itemscope itemtype="http://schema.org/Restaurant">
  <!-- Vlastnosti reštaurácie -->
  <span itemprop="name">Gurmán</span>
  <span itemprop="servesCuisine">Slovenská kuchyňa</span>

  <!-- Vnorená adresa -->
  <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
    <span itemprop="streetAddress">Hlavná 15</span>
    <span itemprop="addressLocality">Bratislava</span>
    <span itemprop="postalCode">811 01</span>
  </div>

  <!-- Hodnotenie -->
  <div itemprop="aggregateRating" itemscope itemtype="http://schema.org/AggregateRating">
    <span itemprop="ratingValue">4.5</span>/5
    <span itemprop="reviewCount">120</span> recenzií
  </div>
</div>
```

Slovník (Vocabulary) pre Microdata

- <https://schema.org/>
- kolaboratívny projekt (Google, Microsoft, Yahoo a Yandex)
- 2011
- Vznikol s cieľom vytvoriť jednotný štandard pre štruktúrované dáta na webe

Úloha

- Vytvorme vizitku pomocou microdata, ktorá bude zahŕňať meno, telefónne číslo a email

.....
.....
.....
.....

.....
.....
.....
.....

```
<div itemscope itemtype="http://schema.org/Person">  
  <span itemprop="name">Martin Novák</span>  
  <span itemprop="telephone">+421 902 123 456</span>  
  <span itemprop="email">martin.novak@firma.sk</span>  
</div>
```

Validácia HTML a microdata

<https://validator.schema.org/>

<https://search.google.com/test/rich-results>

Úloha

- Vytvorte pomocou microdata udalosť
- Postačujúce properties: meno, dátum začatia, poloha, popis
- Hint: <https://schema.org/Event>

Úloha 2

- Pridajte vnorený objekt ktorý popisuje, kto udalosť usporiadal a napíšte meno, adresu a telefónne číslo.
- Hint: <https://schema.org/Organization>

Extrakcia JSON-LD pomocou TypeScriptu

- Použitie knižnice cucumber/microdata
<https://github.com/cucumber/microdata>