

Webové technológie 11

Aplikácie Webových technológií

Peter Bednár

Sémantické údaje – príklad (1)

- Chceme na stránke reprezentovať informácie o osobe ako napr. meno, adresu, kontaktné údaje a pod.:

```
<h1>Peter Bednár</h1>
```

```
<p>
```

```
Pracujem ako odborný asistent na Katedre kybernetiky  
a umelej inteligencie.
```

```
<dl>
```

```
<dt>Adresa:</dt><dd>Vysokoškolská 4, Košice</dd>
```

```
<dt>Tel. číslo:</dt><dd>(+421 55) 602 4219</dd>
```

```
<dl>
```

```
</p>
```

Sémantické údaje – príklad (2)

Peter Bednár

Pracujem ako odborný asistent na Katedre kybernetiky a umelej inteligencie.

Adresa:

Vysokoškolská 4, Košice

Tel. číslo:

(+421 55) 602 4219

- Informácie v HTML sú členené pre zobrazenie čitateľné pre človeka
- Nie sú štruktúrované pre strojové spracovanie informácií

Sémantické údaje

- Strojové spracovanie informácií zo stránok potrebujeme napr. pre zlepšenie vyhľadávania informácií na webe
 - Štruktúrovanie dát je užitočné pre vyhľadávače, ktoré automaticky indexujú HTML stránky pre vyhľadávanie
- Pre strojové spracovanie potrebujeme:
 1. Jednoznačne definovať označenia typov informácií (napr. označenie pre priezvisko: `surname`, `family-name`, `familyName`, ...?)
 2. Formát ako tieto označenia vložiť do HTML kódu

Pomenovanie typov informácií

- **Schéma štruktúrovaných dát** - slovník jednoznačne definovaných označení, ktoré pomenúvajú typy entít a ich vlastnosti
- Typy entít sú usporiadané do hierarchie od všeobecnejších typov po špecifickejšie
- Každá entita má definovaný zoznam vlastností, ktoré ju môžu popisovať
- Každá vlastnosť má definovaný typ hodnôt, ktoré môže vlastnosť nadobúdať
 - Dátové hodnoty (reťazce, čísla, dátumy, a pod.)
 - Odkazy na iné entity

Schema.org

- Všeobecná schéma definovaná pre štruktúrovanie obsahu na webe
- Spoločná iniciatíva Google, Microsoft, Pinterest, Yandex
- V súčasnosti využívaná pre indexovanie 10 mil. stránok na webe
- Špecifikácia: www.schema.org
- Na typy sa jednoznačne odkazuje pomocou URL

Schema.org – príklad

- Typ pre popis osôb:
- Názov typu: [Person](#)
- URL: <http://schema.org/Person>
- Umiestnenie v hierarchii: [Thing](#) (najvšeobecnejší typ) > Person
- Vlastnosti:
 - name: Text
 - address: Text alebo [PostalAddress](#)
 - telephone: Text
 - colleague: [Person](#) alebo URL
 - ...

Microdata

- Rozširuje HTML o nové atribúty elementov pre štruktúrovaný sémantický popis obsahu stránok
- Umožňuje anotovať HTML stránky a označiť v nich štruktúrované dáta
- W3C štandard, špecifikácia: <https://www.w3.org/TR/microdata/>
- Atribúty:
 - **itemscope** – binárny atribút, označuje element pre popis novej entity
 - **itemtype** – definuje typ entity pomocou URL
 - **itemprop** – označuje vlastnosť entity, anotovaný element obsahuje hodnotu vlastnosti entity

Microdata – príklad (1)

- Štruktúrovanie informácií si môžete skontrolovať pomocou nástroja na <https://search.google.com/structured-data/testing-tool>

```
<p itemscope itemtype="http://schema.org/Person">  
  Volám sa <span itemprop="name">Peter Bednár</span> a  
  pracujem ako <span itemprop="jobTitle">odborný  
  asistent</span>.  
</p>
```

Microdata – príklad (2)

- Je možné štruktúrovať aj popis vnorených entít/záznamov, napr.:

```
<div itemscope itemtype="http://schema.org/Organization">
  <div itemprop="name">
    Katedra kybernetiky a umelej inteligencie
  </div>
  <div>Adresa:</div>
  <div itemprop="address" itemscope
    itemtype="http://schema.org/PostalAddress">
    <span itemprop="streetAddress">Vysokoškolská 4</span>
    <span itemprop="addressLocality">Košice</span>
  </div>
</div>
```

Microdata – príklad (3)

- Element `<meta/>` je možné použiť na zadefinovanie hodnoty vlastnosti, ktorá nie je priamo zobrazená resp. je zobrazená v neštandardnom formáte, napr.:

```
<div itemscope itemtype="http://schema.org/Person">  
  Pracujem na  
  <span itemprop="worksFor" itemscope  
    itemtype="http://schema.org/Organization">KKUI  
    <meta itemprop="name" content=  
      "Katedra kybernetiky a umelej inteligencie"/>  
  </span>.  
</div>
```

Microdata – príklad (4)

- Konkrétne entity môžu mať na webe kanonickú reprezentáciu na ktorú je možné sa odkazovať cez URL pomocou elementu `<link/>`, napr.:

```
<p itemscope itemtype="http://schema.org/Movie">  
  Film <span itemprop="name">Star Wars</span> bol  
  natočený v roku <time itemprop="datePublished"  
    datetime="1977-05-25">1977</time>.  
  <link itemprop="url" href=  
    "https://www.imdb.com/title/tt0076759"/>  
  <link itemprop="sameAs" href=  
    "https://en.wikipedia.org/wiki/Star_Wars_(film)"/>  
</p>
```

Sémantický Web a RDF

- Technológie sémantického webu definujú spôsob ako štruktúrovane organizovať a zdieľať informácie a znalosti publikované na webe
- Resource Description Framework – RDF
 - Základný model pre reprezentáciu znalostí definovaný W3C
- Entity sa označujú ako tzv. zdroje (*resources*), ktoré sú jednoznačne identifikované pomocou URI
- Informácie o entitách (subjektoch) sa reprezentujú ako trojica `<subjekt> <predikát> <objekt>`
 - `<predikát>` - označuje vlastnosť entity pomocou URI
 - `<objekt>` - označuje hodnotu vlastnosti
 - Dátová hodnota, alebo odkaz na relevantnú entitu (zdroj)

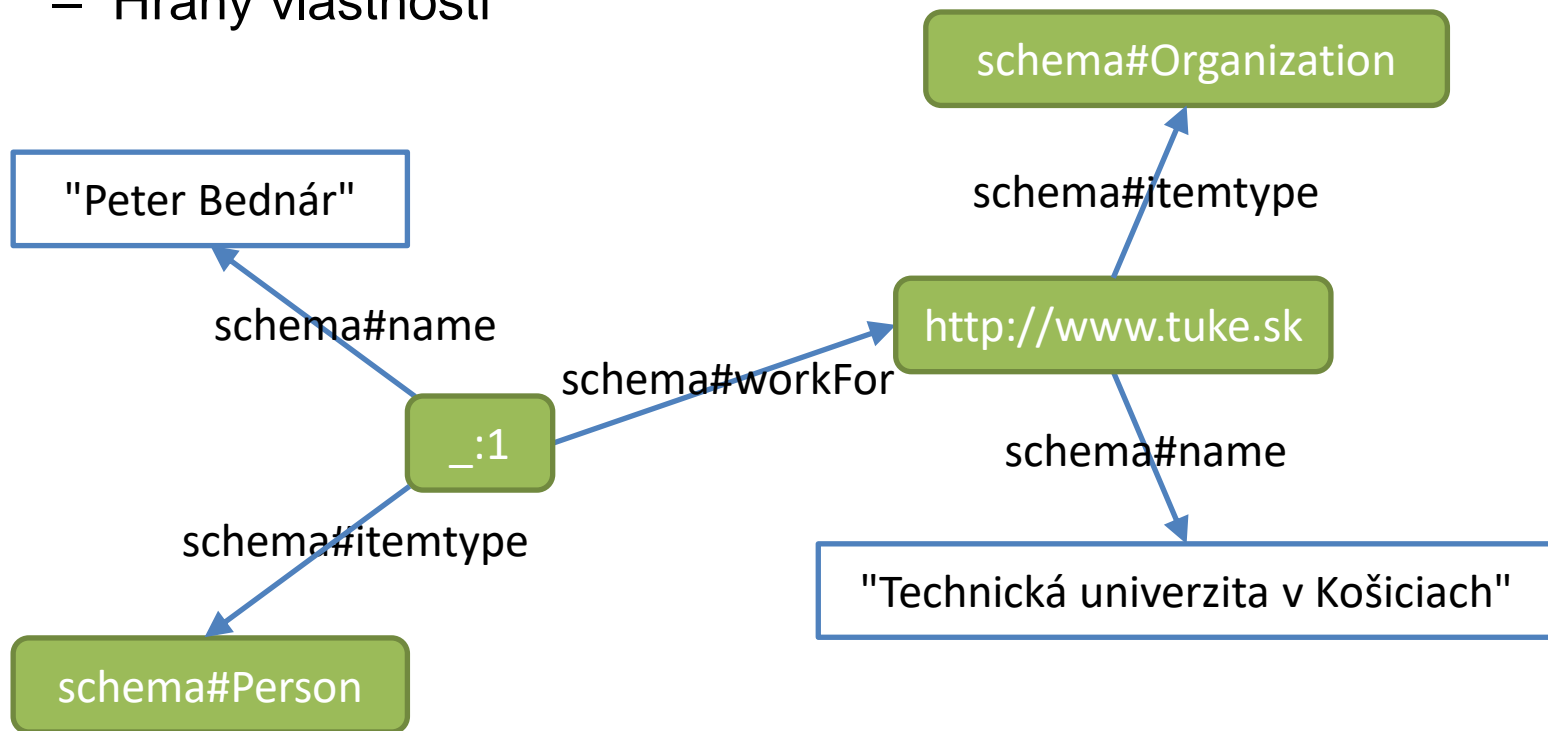
RDF – príklad (1)

- `schema#` je prefix definovaný ako <http://schema.org/>
- `_:1`, `_:2` označujú tzv. anonymné zdroje, ktorým sa priradí vygenerované URI odlišné od ostatných

Subjekt	Predikát	Objekt
<code>_:1</code>	<code>schema#itemtype</code>	<code>schema#Person</code>
<code>_:1</code>	<code>schema#name</code>	"Peter Bednár"
<code>_:1</code>	<code>schema#worksFor</code>	http://www.tuke.sk
http://www.tuke.sk	<code>schema#itemtype</code>	<code>schema#Organization</code>
http://www.tuke.sk	<code>schema#name</code>	"Technická Univerzita v Košiciach"

RDF – príklad (2)

- Triplety je možné zobrazit' ako sémantický graf:
 - Uzly reprezentujú zdroje a dátové hodnoty
 - Hrany vlastnosti



RDFa

- Štandardný formát pre anotovanie HTML podľa znalostného modelu RDF
- Definovaný W3C
 - Dokumentácia: <https://rdfa.info/>
- Podobne ako Microdata definuje sadu atribútov pre HTML elementy
 - **vocab** – špecifikuje URL schémy, ktorá definuje typy a vlastnosti použité pri popise entít (napr. <http://schema.org>), názvy typov a vlastností je potom možné zapisovať skrátene
 - **typeof** – definuje typ entity (zdroja), hodnota je URL
 - **property** – definuje vlastnosť, hodnota je URL RDF predikátu, anotovaný element obsahuje hodnotu vlastnosti entity
 - **content** – definuje hodnotu ak nie je vyjadrená priamo ako obsah anotovaného elementu

RDFa – príklad

```
<div vocab="http://schema.org/" typeof="Person">  
  Volám sa <span property="name">Peter Bednár</span> a  
  pracujem ako  
  <span property="jobTitle">odborný asistent</span> na  
  <span property="worksFor" typeof="Organization">  
    <a property="url" href="http://kkui.fei.tuke.sk">  
      KKUI</a>.  
    <meta property="name" content="Katedra kybernetiky a  
      umelej inteligencie"/>  
  </span>  
</div>
```

Anotovanie dát služieb – JSON-LD

- Okrem HTML stránok je možné anotovať aj dáta webových služieb
- JSON-LD (*JavaScript Object Notation for Linked Data*)
 - Štandard odporúčaný W3C pre anotovanie dát v JSON formáte
 - Špecifikácia: <https://www.w3.org/TR/json-ld/>
- Základné JSON vlastnosti:
 - **@context** – definuje kontext platný pre celý JSON objekt, zvyčajne obsahuje odkaz na schému
 - **@type** – definuje typ JSON objektu, hodnota je názov typu ktorý je definovaný v schéme

JSON-LD – príklad

- Názvy vlastností JSON objektov zodpovedajú vlastnostiam definovaných v type schémy

```
{  
  "@context": "http://schema.org",  
  "@type": "Person",  
  "name": "Peter Bednár",  
  "address": {  
    "@type": "PostalAddress",  
    "streetAddress": "Vysokoškolská 4",  
    "addressLocality": "Košice"  
  }  
}
```