

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY



**Technológie: Mikroformáty - popis technológie, jej
možností a využitia (vrátane napr. možnosti a príkladu
využitia v oblasti digitálnych knižníc)**

ESEJ

Meno: Ivan Stana

Ročník: 3.

UČO: 256650

URL projektu:

<http://microformats.org/>

1 Úvod

Mikroformáty sú zbierka štandardov, ktoré umožňujú identifikovať špecifické druhy informácií, ako napríklad ľudí, vizitku alebo pozície na mape na webových stránkach s využitím (X)HTML značiek. Sú to metainformácie a takto identifikované informácie sú potom navyše strojovo čitateľné. To umožňuje napríklad aplikáciám nakladať s tými informáciami alebo webovému vyhľadávaču nakladať s obsahom na stránke omnoho efektívnejšie.

Vývoj ako taký nemá na starosti žiadna organizácia, ale prebieha v komunite okolo *microformats.org* hlavne vo wiki, mailing listu a IRC kanálu. Niektoré mikroformáty, ako rel-nofollow vznikli inde.

2 Princípy:

Základné princípy a ciele:

- riešiť špecifický problém
- začať najjednoduchšie ako sa dá
 - riešiť najskôr jednoduchšie problémy (malý problém, jednoduché riešenie; rozdeliť si veľký problém na jednoduché podproblémy)
 - postupne pridávať evolučné vylepšenia
- návrh je určený predovšetkým pre ľudí a až potom pre stroje
 - má byť prezentovateľný (teda tiež viditeľný) a parsovateľný
 - viditeľné dáta sú pre ľudí oveľa lepšie ako neviditeľné metadáta
 - prispôbiť ho súčasným trendom, spôsobom používania a správania sa, napr. (X)HTML, blogovanie
 - ručné začleňovanie mikroformátu do kódu má byť jednoduché
- recyklácia základných prvkov zo všeobecne prijatých štandardov
 - využitie sémantického, zmysluplného (X)HTML – vytvoreného napríklad podľa pravidiel POSH (plain old semantic HTML)¹
 - recyklácia existujúcich mikroformátov
 - * ako celku – napríklad použitie *hCard* pre reprezentovanie ľudí (v inom mikroformáte)
 - * čiastočná – využiť isté názvy sémantických tried (z mikroformátov), ale v súlade s princípmi pomenovania ²
 - široko prijatých a používaných schém z interoperabilných RFC

¹<http://microformats.org/wiki/POSH>

²<http://microformats.org/wiki/naming-principles>

- modularita/vložiteľnosť
 - návrh, ktorý sa dá znovupoužiť a začleniť do existujúcich formátov a mikroformátov
- umožniť a podporovať decentralizovaný a distribuovaný výskum, obsah, služby
 - výslovne podporovať pôvodného „ducha Webu“

Ďalšie prebraté princípy:

- **DRY** (Don't repeat yourself) – neopakovať sa
- **Least Surprise** – princíp, ktorý hovorí, že užívateľ by nemal byť prekvapený správaním programu, technológie, atď.
- **Paretov princíp** (pravidlo 80/20)

Dôsledky využívania týchto princípov:

- **Dátová integrita.**
Viditeľné dáta = presnejšie dáta. Pri návrhu predovšetkým pre ľudí a tým, že sú dáta prezentovateľné (a teda viditeľné a skontrolovateľné ľuďmi), sú tým pádom presnejšie nielen vtedy, keď sa dokument vytvorí (keďže ľudia pozerajúci stránky si chyby veľmi rýchlo všimnú), ale tiež v dlhšom časovom horizonte sa rýchlejšie zistí, keď sa dáta stanú zastaralými alebo neaktuálnymi. Je to priamy kontrast k „samostatným súborom“ a neviditeľným dátam, ako napríklad tým, ktoré obsahujú *<meta>* značky.
 - **Neopakovať sa** (podľa princípu DRY) – vďaka tomuto princípu je menšia šanca na nekonzistentnosť dát.
 - **Viacjazyčná integrita.** Možno nie je princípom, ale mnohí z tých, ktorí majú do činenia s mikroformátmi zistili, že konzistentnosť, ktorá vyplýva z používania kódovania UTF-8 pomáha zaručiť, že textový obsah sa nepokazí pri manipulácii, hlavne keď obsahuje aj iné znaky, ako z ASCII7.
- **Znižovanie prekážok pre vydavateľov.**
 Jedným z cieľov mikroformátov je, na rozdiel od snáh iných formátov, byť v ich návrhu viac zameraný na vydavateľov. To však neznamená, že návrhári mikroformátov chcú docieľiť, aby návrhy boli úplne bez snahy vydavateľov. Jasne sa od nich niečo málo od nich žiada. Ale žiada od nich menej, než väčšina iných štandardov, ktoré chcú, aby sa učili nové jazyky, vytvárali nové súbory, menné priestory, atď.
 Napríklad mikroformát *hCalendar*, v porovnaní s formátom *iCalendar* (RFC 2445), znižuje prekážky pre vydavateľov, lebo nevyžaduje učenie sa nového jazyka (.isc) a vytváranie samostatných .ics súborov, atď. Aj keď sa v návrhu *hCalendaru* vyskytlo niekoľko chýb v návrhu, ale aj napriek tomu, že treba rozpoznať tých niekoľko chýb, *hCalendar* okamžite znižuje prekážky pre vydavateľov, aby mohli publikovať a zdieľať udalosti alebo svoj program v otvorenom formáte, na rozdiel od iných formátov kalendárov.
 Princípom, ktorý znižuje prekážky pre vydavateľov, je hlavne pravidlo „návrhu predovšetkým pre ľudí a až potom pre stroje“. To, že návrh je zameraný na ľudí znamená, že

uprednostní, aby ľudia vo všeobecnosti mohli radšej ľahko publikovať informácie v mikroformátoch, než aby stroje mohli ľahšie rozparsovať mikroformáty. Keďže oveľa menej ľudí vyvíja parsery, než publikuje obsah, taký návrh má výhodu v tom, že z ľahšieho publikovania má prospech viac ľudí.

Ale treba si všimnúť, že cieľom je skôr znižovanie prekážok pre vydavateľov, než úplné odstránenie prekážok za každú cenu. Niektorí by mohli mať extrémny názor, že všetko by mali spraviť parsery, ktoré by mali byť čo najinteligentnejšie ako sa dá. Niektoré metódy, ktoré to riešia, sú založené na pravdepodobnosti, dokážu dávať celkom dobre výsledky. Každopádne, detekcia dát na základe pravdepodobnosti a podobne nie je dostatočne dobrá, ak chceme dosiahnuť dátovú integritu. Z toho vyplýva, že mikroformáty nezávisia a nesmú závisieť na pravdepodobnostných metódach, ako napríklad detekcia entít z obsahu (*content entity detection*).

- **Užívateľské znovupoužívanie dát.**

Používaním princípu POSH (plain old semantic HTML) a dodatočné sémantické označovanie cez mikroformáty, umožňuje oveľa väčšie znovupoužívanie dát používateľmi. Napríklad portabilita – niektoré príklady užívateľského znovupoužívania dát: *dátová portabilita* (možnosť užívateľa presúvať a používať jeho dáta v priestore a čase) a *portabilita týkajúca sa sociálnych sietí* (ide o možnosť prenesenia informácií o sebe a svojich nastavení z jednej sociálnej siete do druhej).

Istá časť z tohto dôsledku je dosiahnutá implicitne niečím, čo tvorcovia mikroformátov nechcú aby vydavatelia robili. Konkrétne od vydavateľov chcú, aby označovali svoje dáta sémanticky, čo umožní všeobecné znovupoužitie. Ale nechú od nich, aby sémanticky značkovali dáta so značkami kvôli špecifickým znovupoužitiam.

3 Zoznam mikroformátov

Stabilné:

- *hCalendar* (1:1 reprezentácia iCalendaru (RFC2445)) – kalendárovací mikroformát, zápis udalostí.
- *hCard* – reprezentácia ľudí, spoločností, organizácií a miest.
- *rel-license* – signalizovanie licencií, pod ktorými je obsah uvoľnený.
- *rel-nofollow* – pridaním `rel="nofollow"` do odkazu naznačujeme, že odkaz by nemal mať žiadnu váhu pri analýze odkazov na danej stránke (týka sa napr. botov z vyhľadávačov). Typické použitie je aplikovanie na odkazy v komentároch od iných ľudí alebo odkazy na stránky, ku ktorým sa autor nechce hlásiť.
- *rel-tag* – pridaním `rel="tag"` do odkazu, stránka naznačuje, že s ňou tá cieľová stránka nejako súvisí.
- *VoteLinks* – mikroformát na vyjadrenie podpory, ľahostajnosti alebo vyjadrenie proti podpore nejakej veci. Pridáva sa do odkazov.

- *XFN* (The XHTML Friends Network) – reprezentácia ľudských vzťahov cez odkazy.
- *XMDP* (XHTML Meta Data Profiles) – definovanie metadátových profilov cez (X)HTML, ktoré sú ľahko čitateľné ľuďmi aj strojmi. Značkovanie je profilom XHTML.
- *XOXO* (Extensible Open XHTML Outlines)

Koncepty (špecifikácia sa môže zmeniť, ale už sú v pokročilom vývojovom štádiu):

- *adr* – označenie adresy.
- *geo* – označenie geografických koordinátov (šírka, dĺžka).
- *hAtom* – označuje epizodický obsah (napr. blogové príspevky).
- *hAudio* – vloženie informácií o zvukovej nahrávke.
- *hListing* – definovanie malých reklám, textových inzerátov.
- *hMedia* – formát týkajúci sa publikovania obrázkov, videa a audia.
- *hNews* – na novinové novinky, správy.
- *hProduct* – publikovanie a vloženie informácií o produktoch.
- *hRecipe* – začlenenie informácií o receptoch na varenie.
- *hResume* – na publikovanie životopisov.
- *hReview* – na vloženie recenzií.
- *rel-directory* – pridaním *rel="directory"* do odkazu sa naznačuje, že cieľ odkazu je materská sekcia/stránka súčasnej stránky.
- *rel-enclosure* – pridaním *rel="enclosure"* do odkazu sa označuje, že cieľ odkazu je určený na stiahnutie a cachovanie.
- *rel-home* – pridaním *rel="home"* do odkazu, stránka naznačuje, že cieľ odkazu je úvodná stránka webu, ku ktorému súčasná stránka patrí.
- *rel-payment* – pridaním *rel="payment"* do odkazu sa naznačuje, že cieľ odkazu poskytuje spôsob ako ukázať alebo poskytnúť podporu pre súčasnú stránku. Napríklad poskytnúť finančnú podporu vlastníčkovi súčasnej stránky.
- *robots exclusion* – mikroformát, ktorý definuje pokyny pre webových robotov. Dá sa aplikovať nielen na celú stránku, ale aj na ľubovoľné prvky alebo množinu prvkov na stránke.
- *xFolk* (z „xFolksomony“) – formát na publikovanie zbierok bookmarkov (obľúbených záložiek).

4 Príklad - mikroformát hCard

hCard je jednoduchý, otvorený, distribuovaný formát pre reprezentáciu ľudí, spoločností, organizácií a miest. Používa 1:1 reprezentáciu vlastností a hodnôt *vCardu* (RFC2426) v sémantickom HTML alebo XHTML³. *hCard* je jedným z niekoľkých otvorených mikroformátov, ktoré sa dajú použiť pre vloženie do HTML, XHTML, Atom, RSS a ľubovoľného XML.

Odkaz: <http://microformats.org/wiki/hcard>

Príklad:

```
<div class="vcard">
  <a class="fn org url" href="http://example.org/">Inštitút mágie</a>
  <div class="adr">
    <span class="type">Hlavná budova</span>:
    <div class="street-address">Modrých svetiel, 50</div>
    <span class="locality">Mesto dávneho mora</span>,
    <abbr class="region" title="Lómen">L0</abbr>
    <span class="postal-code">94301</span>
    <div class="country-name">Krajina temného lesa</div>
  </div>
  <div class="tel">
    <span class="type">Práca</span> +2513215451-234-567-890
  </div>
  <div class="tel">
    <span class="type">Fax</span> +2513215451-234-567-8901
  </div>
  <div>Email:
    <span class="email">info@example.org</span>
  </div>
</div>
```

Pri vložení na stránku sa zobrazí toto, ale tieto informácie sa dajú vyexportovať vhodným programom:

Inštitút mágie (odkaz)
Hlavná budova:
Modrých svetiel, 50
Mesto dávneho mora, L0 94301
Krajina temného lesa
Práca +2513215451-234-567-890
Fax +2513215451-234-567-8901
Email: info@example.org

³<http://microformats.org/wiki/semantic-xhtml>

5 Príklad použitia v digitálnych knižniciach:

V digitálnych knižniciach sa dajú mikroformáty použiť na identifikovanie informácií o autoroch a kontaktov s využitím *hCard*, o geografickej pozícii cez *Geo* mikroformát alebo životopisov cez *hResume*, ktoré môžu byť spracované strojmi a napríklad vyextrahovať z textu čisto životopisy a spraviť z nich prehľad, či štatistiku.

Sémanticky označovaný text sa dá využiť na rôzne mixovanie častí z rôznych článkov či kníh. Ďalej sa dá takto označovaný text využiť na vyhľadávanie podobne tematicky zameraných článkov, podobne písaných článkov.

6 Vlastné zhodnotenie projektu a jeho prínosu:

Mikroformáty považujem za celkom zaujímavú technológiu, ktorej úplný potenciál sa podľa môjho názoru ešte nevyužíva. Na rozpoznávanie niekoľkých mikroformátov na webe som použil plugin do Firefoxu – *Operator* (alebo ešte tiež *Tails Export*). Páči sa mi napríklad, ak stránka využíva *hCard*, jednoduché pridanie kontaktu do adresáru.

Filozofia a ciele tvorcov mikroformátov sa mi čiastočne páčia, ale myslím si, že niekedy nie je cieľom tvorcu obsahu ho označovať, aby si ho mohli cudzí roboti pohodlne spracovať a ktovie, čo so získanými výsledkami robiť.

A tiež sa mi nepáči ich cieľ umožnenia a podporovania decentralizovaného a distribuovaného výskumu, obsahu, služieb. Myšlienka sama o sebe nie je zlá, ale pri službách je dôležité, aby všetci boli dôveryhodní a nemali nejaké bočné úmysly ako napríklad zbieranie informácií o užívateľoch a ich následné zneužívanie a podobne.

Ale na druhú stranu, niektoré mikroformáty umožnia ľahšiu navigáciu po webe, hlavne pre veľmi neskúsených, či postihnutých užívateľov.

7 Zdroje:

- <http://microformats.org/>
- <http://zdrojak.root.cz/serialy/kodujme-semanticke-s-mikroformaty/>
- <http://microformats.cz/>

8 Metadáta v Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Technológia: Mikroformáty - popis technológie,
jej možností a využitia (vrátane napr. možnosti a príkladu využitia
v oblasti digitálnych knižníc)" />
<meta name="DC.Creator" content="Ivan Stana" />
<meta name="DC.Subject" content="Technológia: Mikroformáty - popis technológie,
jej možností a využitia (vrátane napr. možnosti a príkladu využitia
v oblasti digitálnych knižníc)" />
```

```
<meta name="DC.Description" content="Mikroformáty sú zbierka štandardov,  
ktoré umožňujú identifikovať špecifické druhy informácií, ako napríklad ľudí,  
vizitku alebo pozície na mape na webových stránkach s využitím (X)HTML značiek." />  
<meta name="DC.Date" content="4.12.2009" />  
<meta name="DC.Type" content="Text" />  
<meta name="DC.Type" content="Esej" />  
<meta name="DC.Format" content="application/pdf" />  
<meta name="DC.Format" content="computerFile" />  
<meta name="DC.Identifier"  
content="http://www.fi.muni.cz/~xstana/dl-mikroformaty.pdf" />  
<meta name="DC.Source" content="http://microformats.org/" />  
<meta name="DC.Source"  
content="http://zdrojak.root.cz/serialy/kodujme-semanticky-s-mikroformaty/" />  
<meta name="DC.Source" content="http://microformats.cz/" />  
<meta name="DC.Language" content="sk" />
```