

Y36WMM

3.přednáška (pokračování)

# HTML metadata - pokračování

## SEO

Miroslav Bureš

# HTML

## metadata

# Hlavička (X)HTML

```
<?xml version='1.0' encoding='utf-8' ?>
```

```
<!DOCTYPE html
```

```
PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
```

```
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

```
<html xmlns="http://www.w3.org/1999/xhtml" lang="cs"
  xml:lang="cs">
```

```
<head>
```

```
<meta http-equiv="Content-Language" content="cs" />
```

```
<meta http-equiv="Content-Type" content="text/html;
  charset=utf-8" />
```

```
<meta name="author" content="Tomáš Vichta" />
```

```
<meta name="description" content="Volební kandidátka Homera" />
```

```
<meta name="keywords"
  content="volby,homer,simpson,kandidátka,volební program" />
```

```
<meta name="robots" content="index,nofollow" />
```

```
<meta name="copyright" content="ČVUT-FEL 2009" />
```

```
<title>Homer - kandidátka</title>
```

```
</head>
```

```
...
```

# Tabulka – atribut scope

```
<table border="1">
  <tr>
    <th></th>
    <th scope="col">Month</th>
    <th scope="col">Savings</th>
  </tr>
  <tr>
    <td scope="row">1</td>
    <td>January</td>
    <td>$100</td>
  </tr>
  <tr>
    <td scope="row">2</td>
    <td>February</td>
    <td>$80</td>
  </tr>
</table>
```

# Tabulka – atribut axis

```
<table border="1" width="100%">
  <tr>
    <th axis="name">Name</th>
    <th axis="contact">Email</th>
    <th axis="contact">Phone</th>
    <th axis="contact">Address</th>
  </tr>
  <tr>
    <td>John Doe</td>
    <td>someone@example.com</td>
    <td>+45342323</td>
    <td>Rosevn 56,4300 Sandnes,Norway</td>
  </tr>
</table>
```

# Tabulka – atribut headers

```
<table border="1" width="100%">
  <tr>
    <th id="name">Name</th>
    <th id="email">Email</th>
    <th id="phone">Phone</th>
    <th id="addr">Address</th>
  </tr>
  <tr>
    <td headers="name">John Doe</td>
    <td headers="email">someone@example.com</td>
    <td headers="phone">+45342323</td>
    <td headers="addr">Rosevn 56,4300 Sandnes,Norway</td>
  </tr>
</table>
```

# Jazyk obsahu

```
<html xmlns="http://www.w3.org/1999/xhtml" lang="cs"
      xml:lang="cs">
```

...

```
<p>
```

Také mluvím několika jazyky:

```
<span lang="en" xml:lang="en">I speak English</span>,
```

```
<span lang="de" xml:lang="de">Ich spreche deutsch</span>,
```

```
<span lang="fr" xml:lang="fr">Je parle français</span>.
```

```
</p>
```

SEO



# SEO

- Search Engine Optimization
- Zvyšování návštěvnosti stránek pomocí lepších výsledků ve vyhledávacích
- Použití „přirozených“, neplacených možností, jak stránku pro vyhledávače zviditelnit (placené zvýraznění stránky ve vyhledávači do SEO nepatří)
- Vychází ze znalosti algoritmu indexování používaném vyhledávací službou
- Legální a nelegální techniky – white hat a black hat

# Indexování stránek

- Stránky jsou procházeny roboty (crawler, spider) vyhledávací služby
- Podle definovaného algoritmu je určený rank stránky, čím lepší rank, tím lepší umístění stránky ve výsledcích, které vrací vyhledávač
- Vstupy do ranku typicky:
  - Počet odkazů směřujících na stránky, rozlišuje se váha
  - Shoda klíčového slova pro které se indexuje s jeho výskyty v metadatech a obsahu stránky
- Některé vyhledávací služby nabízejí zaručení crawlingu za poplatek

# Zahrnutí stránky do indexování

- Většinou se robot na stránku dostane přes link z již indexované stránky
- Google Sitemaps  
<http://www.google.com/webmasters/>
- Může být součástí hostingové služby

# White hat

- „Legální“ technika
- Zajištění viditelnosti v prohlížeči pomocí:
  - Správných metadat
  - Úpravy obsahu webu o klíčová slova, která často uživatelé vyhledávají (*je toto opravdu korektní?*)
  - Přidání více odkazů, které směřují na hlavní stránky webu (*ze stránek samotných i zvenku, ale do rozumné míry, jinak začne být kontraproduktivní, viz black hat*)
  - Zpřístupnění relevantních a schování nerelevantních stránek robotům
    - robots.txt v root adresáři domény
    - `<meta name="robots" content="index,follow" />` v `<head>`

# White hat – čitelnost stránky

- Když přistoupí robot, musí mít možnost stránku automaticky přečíst:
  - Validita HTML/XHTML
  - Použití sémantických značek
  - Návaznost stránek klasickými odkazy (redirect např. javascriptem, frames, nebo izolované stránky → robot se na stránku nedostane)

# Black hat (spamindexing)

- „Nelegální“ technika
- Umělé zvyšování ranku stránky bez ohledu na její reálný obsah:
  - Link farms
  - Keyword stuffing
  - Article spinning
  - Cloaking
- Pokud vyhledávací služba rozpozná nelegální techniku (už to efektivně dokáží):
  - Sníží page rank
  - Vyřadí stránku z indexu natrvalo
  - Pokud při strojovém prohledávání vznikne podezření na spamindexing, může následovat „ruční“ revize stránky

# Link farms

- Generování linků směřujících na stránky, u kterých chceme zvýšit rank
- Prováděno ve většině případů automaticky
- Reciprocal link management services

# Keyword stuffing

- Přidání častých klíčových slov:
  - Do metadat dokumentu
    - Některé vyhledávací služby je kvůli tomuto začaly obecně ignorovat <meta> tagy
  - Přidání častých klíčových slov na stránky tak, aby nebyly vidět:
    - Barva textu stejná jako barva pozadí / velmi podobná
    - Zobrazení mimo viditelnou část obrazovky prohlížeče (CSS absolutní pozice)
    - Zobrazení text před obrázek, aby nebyl vidět (pomocí CSS)
    - Neviditelný <div>
    - Neviditelný <iframe>
- U většiny vyhledávacích služeb už nemá význam



# Article spinning

- Cílem je vyhnout se penalizaci v ranku za duplicitní obsah stránek
- Přepisování stávajícího obsahu stránek:
  - Ručně
  - Automaticky nahrazováním slov za synonyma podle slovníku

# Cloaking

- Pokud na server přistupuje robot, server mu poskytne jinou stránku než v případě člověka (přístup z browseru)

Děkuji za pozornost

[buresm3@fel.cvut.cz](mailto:buresm3@fel.cvut.cz)