

# 10. SEO – Obsah meta, konkrétní elementy v html kódu

Web pro kodéry  
(Petr Kosnar, ČVUT, FJFI, KFE, PINF 2008)

# Obsah

---

- Obsah stránek
- Meta data
- Meta Title
- Meta Description
- Meta Keywords
- Zdrojový kód
- Odkazy
- Vyhledávací roboty
- Přesměrování
- Rss

# Meta tagy

---

- Informace o dokumentu
  
- Název dokumentu – TITLE
- Popis dokumentu – DESCRIPTION
- Klíčová slova – KEYWORDS
- Kódování (znaková sada)
  
- Přesměrování – REFRESH
  - Kromě Seznamu tento redirect neumí vyhledávače následovat
    - => Nevhodné (používat raději HTTP hlavičky 301,302)
- A další...

# Meta Title

---

- **Nejdůležitější!**
- Pro každou stránku vlastní, unikátní, popisující výstižně obsah stránky
  - Klíčová slova
- Na konci je název webu
  - Na začátku jen v případě velmi známé značky
    - Když uživatele naláká právě tato značka víc než konkrétní titulek
- Přiměřeně dlouhý
  - Okolo 70 znaků
    - Zbytek může být oříznut (zejména Seznamem)
- Nemusí se shodovat s URL, hlavním nadpisem stránky ani ničím jiným
- Může se měnit
- Zobrazuje se na SERP jako text odkazu na web

# Meta Description

---

- Nemá vliv na hodnocení vyhledávači
- Nemusí obsahovat klíčová slova
  - Může, ale nemá to žádný význam
- Pro každou stránku unikátní a vystihující obsah stránky
  
- Zobrazuje se v Googlu na SERP pod textem odkazu
  - Pokud chybí, zobrazuje se zde začátek stránky
  - Seznam vždy generuje úryvek ze stránky
  
- Měl by výstižně a lákavě popisovat obsah stránky, zaujmout a přimět k návštěvě tohoto odkazu.

# Meta Keywords

---

- Nemají dnes žádný vliv na SEO

# HTML kód

---

- Validita (X)HTML
- Strukturování a sémantika
- Důležitý obsah na začátku kódu
  - Ne kvůli tomu, že by to vyhledávače lépe hodnotily, ale kvůli tvoření úryvků ze stránky
    - Zvyšujeme pravděpodobnost vygenerování správného a relevantního úryvku
- Atribut **title**
  - Nemá u odkazů žádný význam. Neindexuje se
- Atribut **alt**
  - U obrázků má význam i pro indexaci obrázků (Google Images)
- Sémantické značky typu **strong**, **h1-h6**, **em**, ... Mají velmi malý význam
  - Protože by byly znevýhodněny weby s kvalitním obsahem, ale špatným kódem
    - To by uživatelům nepřineslo žádný užitek

# Odkazy

---

- **Textový odkaz a**
  - Obsahuje vhodný text popisující účel/cíl odkazu
    - Srovnejte:  
Výsledky hlasování naleznete [zde](#)  
Prohlédněte si [výsledky hlasování](#)
- **Obrázkový odkaz**
  - **Atribut `alt`**
    - Má u obrázkových odkazů stejný význam jako text uvnitř textového odkazu
    - Lepší než „číst dále...“ za perexem je umístit obrázkovou ikonu s vhodným `alt`em



# Robots.txt

---

- Soubor s kterým můžete povolit nebo zakázat robotovi přístup do určité složky na Vašich stránkách.
- Je umístěn v rootu webu
  
- Příklad:
  - Zákaz veškeré indexace:  
User-agent: \*  
Disallow: /
  
  - Zákaz indexace všeho ve složce /privat pro google:  
User-Agent: Googlebot  
Disallow: /privat/
  
- Pomocí Google Webmaster Tools lze analyzovat a generovat robots.txt
- Validátor: <http://tool.motoricerca.info/robots-checker.phtml>

# Indexace obsahu

---

- Boti obvykle neumí interpretovat JavaScript
  - Skrytí obsahu před boty pomocí JS
  - Podstrčení obsahu botům pomocí noscript
- Vyhledávače neakceptují cookies
  - Skrytí obsahu před roboty
    - Administrační rozhraní
    - Stránky dostupné po přihlášení
- Pokud máme tiskovou verzi, skryjeme jí pomocí JS
  - Vyhledávače jí obvykle indexují raději, protože je pro ně čitelnější (bez obrázků a prezentačních prvků, jednodušší, ...)
  - Pokud jí schovám pomocí robots.txt, přijdu o indexaci celé stránky
    - To je nejspíš nežádoucí
- V rámci viditelného textu se indexují i obsahy formulářových prvků
  - Pokud nejsou žádoucí, skryjeme pomocí JavaScriptu
- Příliš velká hustota klíčových slov není postihována, ale ani nepřináší výhody
  - Nezbyvá místo pro ostatní slova
  - Raději pestrost blízkých slov
- V zahraničí jsou časté RSS vyhledávače => pozor na správnou hlavičku

# Pár drbů o botech...

---

- Seznam má malý timeout
  - Pokud je stránka moc velká nebo pomalý server, nezaindexuje jí a zkusí to často až za dlouho (třeba i několik měsíců)
- Seznam pracuje hodně s proximitou klíčových slov
  - Google moc ne
- Seznam pracuje hodně s hustotou klíčových slov
  - Google moc ne
- Crawler neuvádí referera, nebo uvádí sebe
- Seznam zpracovává HTTP hlavičku Accept-Language
  - Google ne
- Seznam je velmi přísný při hodnocení duplicitního obsahu
  - Někdy obsah!
  - Google je mírnější
- Google preferuje větší množství kratších stránek
  - Seznam preferuje méně delších stránek

# Přesměrování

---

- Nejvhodnější je přímo na serveru
  - Nastavení Apache nebo pomocí .htaccess
- Vždy s příslušnou hlavičkou HTTP Status!
  - HTTP 301 - Moved Permanently
  - HTTP 302 - Found (Temporarily under different URI)
- Nepřesměrovávat 404 na homepage!
  - Tvoří se tím nekonečný duplicitní obsah
- Vyřazené stránky by měly mít stále nějaký relevantní obsah
  - Produkty, které již nejsou v prodeji, neaktuální články, atp.
  - Nejlépe zachovat jejich původní obsah, nebo aktuální verzi
    - HTTP 200 – OK
  - Pokud je to nutné, a daná stránka nebude už nikdy potřeba, přesměrovat pomocí HTTP 300 - Multiple Choices nebo obdobně

## Další informace

---

- <http://www.seo-expert.cz>
- <http://www.google.com/webmasters>
- <http://www.seo-profesional.cz>
- <http://www.seochat.com>
- <http://www.w3.org/Protocols/rfc2616/rfc2616-sec10.html>

# Kontrolní úkol

---

- Optimalizujte HTML kód svých stránek pro lepší SEO
  - Včetně robots.txt
  - Včetně užitečných meta tagů