

# SiteMap V.16.02

V prípade že máte na Internete množstvo **.html** stránok asi by ste chceli aby boli Vaše stránky nájdené cez Google. Vyhľadávač Google nevie o Vašej aktivite. Jednoduché je registrovať si Vašu stránku napr. <http://jankovic.azet.com> .

Google potom používa vyhľadávací program tzv. Spider, ktorý si pozrie Váš súbor **index.html** a tam hľadá odkazy na ďalšie Vaše .html stránky.

Je dobrým zvykom mať v hlavnom tj. koreňovom adresári ešte súbor **robots.txt** , kde necháte pre Spider nejaký odkaz. Tento súbor môže vyzeráť takto :

```
User-agent: *  
Allow: /
```

V tomto prípade mu pán Jankovič oznamuje, aby si pozrel všetky stránky ktore nájde prikazom **Allow** . Sú aj iné možnosti, tie si dohľadajte na Internete.

Môže sa ale stať, že používate v menu JavaScript a Spider vo Vašom súbore index.html nenajde použiteľný odkaz. Ja som narazil na free hosting, kde sa serveru "spiderovanie" jednoducho neľúbi a tak nenajde a nenajde. Keďže je služba free, niet sa kde sťažovať, dokonca ani na lampárni.

Pre tieto účely zaviedol Google pomôcku pre Spidera akýsi zoznam stránok kam sa má Spider pozrieť. Keďže Spider si načíta stránku ako Internet-Explorer, nájde spoľahlivo všetko čo mu ponúknete, teda ak taký súbor na Internete je. Upravený súbor robots.txt potom vyzerá takto:

```
User-agent: *  
Allow: /
```

```
Sitemap: http://jankovic.azet.com/sitemap.xml
```

Existujú aj programy ktoré online alebo z Vášho PC "spiderujú" Vaše stránky a potrebný súbor s menom **sitemap.xml** vytvoria na Vašom PC. Lenže aj tie môžu naraziť a potom nastane problém. No dal by sa súbor sitemap.xml vytvoriť aj ručne, akorát taký ako som ja potreboval by mal viac ako 154 riadkov.

## 1. Obmedzenia programu

Program Sitemap.exe ktorý som vytvoril má určité obmedzenia ale mne celkom vyhovuje. Aby sme predišli sklamaniu tak ich tu uvádzam.

## Bod 1.

Na mojom PC mám rovnakú štruktúru adresárov ako mám na Internete. Je potrebné dodržiavať veľké a malé písmená. Teda aj keď to doteraz nebolo nutné, pre použitie programu musia byť názvy podadresárov a súborov s dodržaním malých a veľkých písmen a v každom adresári je súbor index.html

## Bod 2.

Program neprehľadáva adresáre na Internete, ale na Vašom PC . V každom adresári aj podadresári hľadá súbor index.html . Ak sa v podadresári súbor index.html nenajde predpokladá sa, že taký podadresár nie je na Internete.

## Bod 3.

Iné .html súbory ako index.html si program nevšíma. Ak máte v nejakom podadresári viac .html súborov predpokladá sa, že sa jedná o pokračovanie. Nie je v záujme veci aby Google indexoval napr. druhý diel príbehu, lebo tým by záujemca nečítal príbeh od začiatku.

Takže teraz názorný príklad :

1	<b>index.html</b>	jankovic.azet.com
2	novinky.html	
3	<b>Kvety</b> <b>index.html</b>	
4	<b>Orchidea</b> <b>index.html</b>	
5	<b>Tulipan</b> <b>index.html</b>	
6	Sklad	popis.html
7		
8	<b>Zvierata</b> <b>index.html</b>	
9	<b>Pes</b> <b>index.html</b>	
10		casopisy.html
11		knihy.html
12	<b>Medved</b> <b>index.html</b>	
13		hnedy_medved.htm

Riadok 1. bude mať adresu <http://jankovic.azet.com/index.html>

Riadok 2. **nebude indexovaný** - meno súboru

Riadok 3. bude mať adresu <http://jankovic.azet.com/Kvety/index.html>

Riadok 4. a riadok 5. bude indexovaný ak o to požiadate, v takom prípade to bude

<http://jankovic.azet.com/Kvety/Orchidea/index.html>

<http://jankovic.azet.com/Kvety/Tulipan/index.html>

Riadok 6. **nebude indexovaný** lebo adresár neobsahuje súbor index.html

Riadok 8. bude mať adresu <http://jankovic.azet.com/Zvierata/index.html>

Riadok 9. bude indexovaný ak o to požiadate, v takom prípade to bude <http://jankovic.azet.com/Zvierata/Pes/index.html>

Riadok 10. a 11. **nebude indexovaný** - meno súboru

Riadok 12. bude indexovaný ak o to požiadate, v takom prípade to bude <http://jankovic.azet.com/Zvierata/Medved/index.html>

Riadok 13. **nebude indexovaný** - meno súboru

## 2. Parametre pre program

Pretože program Sitemap.exe nemá žiadnú grafickú nadstavbu kde je nejaké menu, musí sa nejako dozvedieť kde je na Internete Vaša hlavná stránka. Program hľadá súbor **sitemap.txt** kde musí byť minimálne jeden riadok. Ak súbor sitemap.txt nenajde, alebo bude súbor sitemap.txt prázdny nevytvorí nič.

Súbor **sitemap.exe** a **sitemap.txt** musia byť v koreňovom adresári na Vašom PC, teda tam kde máte Váš prvý (úvodný) súbor index.html.

Pre horeuvedený príklad stačí taký obsah súboru sitemap.txt

<http://jankovic.azet.com>

S takýmto súborom ale nejde program sitemap.exe "do hĺbky", teda najde len :

<http://jankovic.azet.com/index.html>

<http://jankovic.azet.com/Kvety/index.html>

<http://jankovic.azet.com/Zvierata/index.html>

Vy rozhoduje o tom, či ponúknete pre Google v súbore sitemap.xml aj viac. Takže ak ponúkate podadresáre pod **Kvety** a podadresáre pod **Zvierata**, musí súbor sitemap.txt vyzerať takto:

<http://jankovic.azet.com>

Kvety

Zvierata

V distribučnom balíku máte súbor sitemap.txt ktorý ovšem musíte upraviť podľa Vašej stránky. Potom už len stačí kliknúť na súbor sitemap.exe a dielo je hotové. Keď neskôr niečo zmeníte, stačí len znova kliknúť na sitemap.exe .

## 3. Skúška správnosti

Stále ste sa spoliehali, že štruktúra na Vašom PC a na Internete je zhodná. Tak ako v škole aj tu sa odporúča skúška správnosti. Stačí ju urobiť len raz a potom pri každej zmene v štruktúre Vašich adresárov. Napríklad keď pridáte podadresár Sport.

Na ten účel si ešte do koreňového adresára nakopírujte z distribučného balíka súbor **sitemap.xsl** . Potom otvorte vygenerovaný súbor sitemap.xml v internetovom prehliadači.

# XML Sitemap

Total URLs in this sitemap file: 6	
	URL
1	<a href="http://stofko.zohosites.com/index.html">http://stofko.zohosites.com/index.html</a>
2	<a href="http://stofko.zohosites.com/Galeria/index.html">http://stofko.zohosites.com/Galeria/index.html</a>
3	<a href="http://stofko.zohosites.com/Giftab/index.html">http://stofko.zohosites.com/Giftab/index.html</a>
4	<a href="http://stofko.zohosites.com/GoogleTel/index.html">http://stofko.zohosites.com/GoogleTel/index.html</a>
5	<a href="http://stofko.zohosites.com/HomePhone/index.html">http://stofko.zohosites.com/HomePhone/index.html</a>
6	<a href="http://stofko.zohosites.com/Vizitka/index.html">http://stofko.zohosites.com/Vizitka/index.html</a>
Sitemap freeware from	

Teraz postupne preklikajte všetky riadky, musí Vám nabehnúť Vaša stránka z Internetu. Ak nie hľadajte chybu u seba. Keď raz Vašu štruktúru podadresárov odskúšate, nemusíte to už robiť znova. Program sitemap.exe Vám vždy vygeneruje správny súbor sitemap.xml . Pri zmenách stačí urobiť skúšku správnosti len pre nové položky.

Súbor **sitemap.xml** a **robots.txt** pošlite na Internet.  
Súbor sitemap.xml a sitemap.exe nie je na Internete potrebný.

Môj malý súbor sitemap.xml vyzerá takto :

```
<?xml version='1.0' encoding='UTF-8'?>
<?xml-stylesheet type='text/xml' href='sitemap.xml'?>
<urlset xmlns='http://www.sitemaps.org/schemas/sitemap/0.9'
xmlns:xsi='http://www.w3.org/2001/XMLSchema-instance'
xsi:schemaLocation='http://www.sitemaps.org/schemas/sitemap/0.9
http://www.sitemaps.org/schemas/sitemap/0.9/sitemap.xsd'>

<url><loc>http://stofko.zohosites.com/index.html</loc></url>
<url><loc>http://stofko.zohosites.com/Galeria/index.html</loc></url>
<url><loc>http://stofko.zohosites.com/Giftab/index.html</loc></url>
<url><loc>http://stofko.zohosites.com/GoogleTel/index.html</loc></url>
<url><loc>http://stofko.zohosites.com/HomePhone/index.html</loc></url>
<url><loc>http://stofko.zohosites.com/Vizitka/index.html</loc></url>

</urlset>
```

## 4. Vedľajší produkt sitemap.html

V novej verzii som ešte doplnil generovanie súboru **sitemap.html** . Pozor, súbor sa generuje automaticky a môže prepísať Váš súbor s rovnakým menom. Taký súbor používajú aj renomované portály. Slúži na primitívnu orientáciu pre istý typ užívateľov.

Nedá mi vrátiť sa do hlbokej minulosti, keď som poznal jednu pani, ktorá písala služobnú korešpondenciu do jedného súboru Doklady.602 vo vtedajšom programe T602 v MS DOS-e. Potom keď potrebovala vytlačiť jednu A4 stranu, našla si ju v tom súbore a nechala tlačíť tú jednu stranu. Celé to skrachovalo v okamihu keď jej havaroval ten súbor Doklady.602 ... Možno by stačilo každý doklad písať do samostatného súboru a mať potom lepší prehľad a mať istotu.

Aby nemuselo byť všetko v jednom adresári a vôbec nie v jednom súbore môžeme si vytvoriť aj viac adresárov (priečinkov) čo zjednodušuje hľadanie v budúcnosti. Tak to potom aj vyzerá vo Vašej internetovej stránke. Majme ako príklad, že píšete niečo o orchideach a Vaše lokálne aj internetové adresáre vyzerajú takto :

index.html v koreňovom adresári

Zvierata - Podadresár  
Zvierata - Psy - Buldog - Podadresár  
Zvierata - Psy - **Buldog - index.html**

Priroda - Podadresár  
Priroda - Kvety - Podadresár  
Priroda - Kvety - Orchidea - Podadresár  
Priroda - Kvety - **Orchidea - index.html**

Pochopiteľne vo Vašich .html súboroch je odkaz na **Priroda**, v prirode na **Kvety** a tam na **Orchidea**. Takže klikaním sa tam dá ľahko dostať. Lenže keď stretnete na ulici Vašu susedku dôchodkynu ktorá má internet, ona sa tam z Vašej prvej stránky nedokliká a teda Váš text o tom ako pestujete orchideu nenajde a bude reklamovať.

Program sitemap.exe Vám automaticky vygeneruje súbor sitemap.html . Či ho použijete je len na Vás, môžete ale aj nemusíte. Ak ano, v súbore index.html v koreňovom adresári by mal byť ľahko nájateľný odkaz na tento súbor.

Súbor sitemap.html začne byť nepríjemne veľký keď má viac ako 200 riadkov, treba v ňom hľadať tak ako to Vaša susedka robí v papierovom telefonnom zozname (kniha). Napríklad internetové obchody, kde sú tisíce položiek obvykle sitemap.html preto nepoužívajú.

Aby ste mohli v súbore sitemap.html meniť jeho vzhľad je použitý aj súbor štýlov **sitemap.css** Tento si v takom prípade tiež nakopírujte do koreňového adresára. Treba si uvedomiť, že súbor sitemap.html sa generuje vždy automaticky a prepísal by Vaše dokonalé dielo. Naopak **súbor sitemap.css je fixný** a ten sa použije vždy v takom stave ako je. Výsledok môže vyzeráť takto :

<http://stofko.zohosites.com>

1. GALERIA Menu s referenciami na podadresáre
2. GIFTAB Súbor s referenciami na .jpg súbory
3. GoogleTel Telefonný zoznam z kontaktov GMAIL-u.
4. HomePhone Zotriedený telefónny zoznam
5. SiteMap Vytvorí súbor sitemap.xml pre Google
6. VIZITKA Generuje vizitku pre tlač

Na rozdiel od súboru .xml tu hľadá program sitemap.exe nielen cestu k súboru, ale aj to čo je napísané v HTML značke **title**. To potom ukáže v menu. Značka title je viacmenej povinná a text tam uvedený sa zobrazí vo Vašom internetovom prehliadači v hlavičke. V každom Vašom súbore to môže vyzeráť takto :

```
<html> <title>Červená orchidea</title>
```

alebo

```
<html>  
<title>Červená orchidea</title>
```

Text medzi značkami title musí byť pre program sitemap.exe v jednom riadku !

Takto by to nešlo, aj keď by to možno prehliadač akceptoval :

```
<html>  
<title>  
Červená  
orchidea  
</title>
```

Skoro by som zabudol na významnú výhodu programu sitemap.exe . Keď totiž dodatočne zmeníte v značke title text, napr.

```
<html>  
<title>Červená orchidea je najkrajšia</title>
```

tento text program pri budúcom generovaní súboru sitemap.html automaticky preberie. Teda nemusíte nič meniť na dvoch miestach, čo znižuje výskyt nezrovnalostí.

A nakoniec aj o diakritike. Ja riešim diakritiku vždy nakoniec. U súboru sitemap.xml je to vcelku fuk, ale v súbore sitemap.html treba definovať diakritiku. Aby ste nemuseli do automaticky generovaného súboru zasahovať a aby ste ani náhodou na to nezabudli, program sitemap.exe číta Váš súbor index.html v koreňovom adresári a tam hľadá definíciu diakritiky, napr.

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

Z toho riadku si zistí akú diakritiku používate (utf-8, windows-1250 alebo inú). Zistenú hodnotu potom použije pri generovaní súboru sitemap.html . Ak definíciu nenajde, nastaví charset=windows-1250 .

Je jedno či vo Vašom súbore použijete :

- charset=utf-8">
- CHARSET=UTF-8">
- charset=utf-8 >
- CHARSET=UTF-8">
- charset = utf-8 ">
- CharSet =utf-8 ">

## 5. Rozšírené vyhľadávanie

Od verzie V.16.02 bolo v programe zavedené rozšírené vyhľadávanie. V obmedzeniach vyššie sa ako bod 3 uvádza, že program hľadá v každom adresári len súbor **index.html** a ostatné .html súbory ignoruje. Tak mi to vyhovovalo pre galeriu fotografií, pozri moje programy :

[GIFTAB Súbor s referenciami na .jpg súbory](#)

[GALERIA Menu s referenciami na podadresáre](#)

Lenže sú aj také situácie v prípade skôr textových stránok, že v jednom podadresári je aj viac .html súborov. Pri riešení toho ako programu Sitemap oznámiť aké riešenie chcem som postupoval podobne ako u horeuvedených dvoch programov.

Keď potrebujete aby Sitemap akceptoval aj iné .html súbory ako index.html , musíte to "vpašovať" do súboru index.html v hlavnom (koreňovom) adresári. Niekde, stačí tesne pred koncom vložte neštandardnú značku sitemap takto :

```
<sitemap all>
</body></html>
```

Syntakticky čistejšie by bolo napísať **<sitemap all />** . Takáto neštandardná značka sa nezobrazí. Slovo all v angličtine značí všetko-všetky. Podľa toho sa zmení, teda rozšíri rozsah súboru sitemap.xml aj sitemap.html . Toto nastavenie platí pre všetky podadresáre okrem hlavného (koreňového), kde najde vždy len index.html . Pokiaľ by Vám to vadilo, napíšte !

## 6. Oznamy o chybách

Môžu nastať nasledovné kritické situácie na ktoré program reaguje takto :

### Cannot find Sitemap.txt

Ako už bolo povedané bez súboru sitemap.txt program nefunguje. Vkopírujte si z distribučného balíku súbor sitemap.txt a upravte podľa Vašich želaní.



### **Cannot find URL in Sitemap.txt**

Súbor sitemap.txt je prázdny, neobsahuje ani len prvý riadok. Syntax adresy Vašej stránky teda URL program sitemap.exe nepreveruje, akceptuje to čo tam napíšete.

### **Cannot open directory XYZ**

Na Vašom disku je nejaký vážny problém. Súborový systém síce vidí adresár napr. XYZ , ale nevie ho otvoriť a teda pozreť sa aké sú tam súbory. Taká chyba by sa vo fungujúcom Windows nemala vyskytnúť !!!

### **Title not found**

Do súboru sitemap.html sa vloží tento text ak Váš .html súbor neobsahuje syntakticky správnu značku **<title>** . Otvorte si patričný súbor v podadresári a opravte. Takýto nedokonalý .html súbor by ste nemali pre Google ponúkať.

## **7. Záver**

Program je FREEWARE. Autor si ale vyhradzuje právo rozhodnúť o tom kde bude program uložený pre stiahnutie. Pokiaľ hodláte program na stiahnutie uložiť na iný server, prosím informujte autora.

Program Sitemap.exe má v aktuálnej verzii nasledovné vlastnosti :

Dĺžka 8 192 Bajtov  
Dátum 02.01.2016  
Čas 10:49:48

Ing. Branislav Štofko [b.stofko@zoho.com](mailto:b.stofko@zoho.com)