

SiteMap V.16.02

V případě že máte na Internetu množství .html stránek asi byste chtěli aby byly Vaše stránky nalezeny přes Google. Vyhledávač Google neví o Vaší aktivitě. Jednoduché je registrovat si Vaši stránku např.

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

Google pak používá vyhledávací program tzv. **Spider**, který se podívá na Váš soubor index.html a tam hledá odkazy na další Vaše .html stránky.

Je dobrým zvykem mít v hlavním tj. kořenovém adresáři ještě soubor **robots.txt**, kde necháte pro Spider nějaký odkaz. Tento soubor může vypadat takto :

```
User-agent: *
```

```
Allow: /
```

V tomto případě mu pan Jankovič sděluje, aby se podíval na všechny stránky které najde příkazem **Allow**. Jsou i jiné možnosti, ty si dohledejte na Internetu.

Může se ale stát, že používáte v menu JavaScript a Spider ve Vašem souboru index.html nenajde použitelný odkaz. Já jsem narazil na free hosting, kde se serveru "spiderování" jednoduše nelíbí a tak nenajde a nenajde. Protože je služba free, není se kde stěžovat, dokonce ani na lampárně.

Pro tyto účely zavedl Google pomůcku pro Spidera jakýsi seznam stránek kam se má Spider podívat. Protože Spider si načte stránku jako Internet-Explorer, najde spolehlivě vše co mu nabídnete, tedy pokud takový soubor na Internetu je. Upravený soubor robots.txt pak vypadá takto:

```
User-agent: *
```

```
Allow: /
```

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

Existují také programy které online nebo z Vašeho PC "spiderují" Vaše stránky a potřebný soubor se jménem **sitemap.xml** vytvoří na Vašem PC. Jenže i ty mohou narazit a potom nastane problém. No dal by se soubor sitemap.xml vytvořit i ručně, akorát takový jako jsem já potřeboval by měl více než 154 řádků.

1. Omezení programu

Program Sitemap.exe který jsem vytvořil má určitá omezení ale mně docela vyhovuje. Abychom předešli zklamání tak jejich zde uvádím.

Bod 1.

Na mém PC mám stejnou strukturu adresářů jako mám na Internetu. Je třeba dodržovat velké a malé písmena. Tedy i když to dosud nebylo nutné, pro použití programu musí být názvy podadresářů a souborů s dodržením malých a velkých písmen. V každém adresáři je soubor index.html

Bod 2.

Program neprohledává adresáře na Internetu ale na Vašem PC. V každém adresáři i podadresáři hledá soubor index.html. Pokud se v podadresáři soubor index.html nenajde předpokládá se, že takový podadresář není ani na Internetu.

Bod 3.

Jiné .html soubory jako index.html si program nevšímá. Pokud máte v nějakém podadresáři více .html souborů předpokládá se, že se jedná o pokračování. Není v zájmu věci aby Google indexoval např. druhý díl příběhu, neboť tím by zájemce nečetl příběh od začátku.

Takže teď názorný příklad :

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

Řádek 1. bude mít adresu <http://jankovic.azet.com/index.html>

Řádek 2. **nebude indexovaný** - jméno souboru

Řádek 3. bude mít adresu <http://jankovic.azet.com/Kvety/index.html>

Řádek 4. a řádek 5. bude indexován pokud o to požádáte, v takovém případě to bude

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

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

Řádek 6. **nebude indexovaný** neboť adresář neobsahuje soubor index.html

Řádek 8. bude mít adresu <http://jankovic.azet.com/Zvierata/index.html>

Řádek 9. bude indexován pokud o to požádáte, v takovém případě to bude

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

Řádek 10. a 11. **nebude indexovaný** - jméno souboru

Řádek 12. bude indexován pokud o to požádáte, v takovém případě to bude <http://jankovic.azet.com/Zvierata/Medved/index.html>

Řádek 13. **nebude indexovaný** - jméno souboru

2. Parametry pro program

Protože program Sitemap.exe nemá žádnou grafickou nastavbu kde je nějaké menu, musí se nějak dozvědět kde je na Internetu je Vaše hlavní stránka. Program hledá soubor **sitemap.txt** kde musí být minimálně jeden řádek. Pokud soubor sitemap.txt nenajde nebo bude soubor sitemap.txt prázdný nevytvoří nic.

Soubor **sitemap.exe** a **sitemap.txt** musí být v kořenovém adresáři na Vašem PC, tedy tam kde máte Váš první (úvodní) soubor index.html.

Pro výše uvedený příklad stačí takový obsah souboru sitemap.txt

<http://jankovic.azet.com>

S takovým souborem ale nejde program sitemap.exe "do hloubky", tedy najde jen :

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

Vy rozhoduje o tom, zda nabídnete pro Google v souboru sitemap.xml i více. Takže pokud nabízíte podadresáře **Kvety** a podadresáře **Zvierata**, musí soubor sitemap.txt vypadat takto :

<http://jankovic.azet.com>
Kvety
Zvierata

V distribučním balíku máte soubor sitemap.txt který ovšem musíte upravit podle Vaší stránky. Pak už jen stačí kliknout na soubor sitemap.exe a dílo je hotovo. Když později něco změníte, stačí jen znovu kliknout na sitemap.exe .

3. Zkouška správnosti

Stále jste se spoléhali, že struktura na Vašem PC a na Internetu je shodná. Tak jako ve škole i zde se doporučuje zkouška správnosti. Stačí ji udělat jen jednou a pak při každé změně ve struktuře Vašich adresářů. Například když přidáte podadresář Sport.

Na ten účel si ještě do kořenového adresáře nakopírujte z distribučního balíku soubor **sitemap.xsl** . Pak otevřete vygenerovaný soubor sitemap.xml v internetovém prohlížeči.

XML Sitemap

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

Nyní postupně proklikejte všechny řádky, musí Vám naběhnout Vaše stránka z Internetu. Pokud ne hledejte chybu u sebe. Když jste jednou Vaši strukturu podadresářů odzkoušeli, nemusíte to už dělat znovu. Program sitemap.exe Vám vždy vygeneruje správný soubor sitemap.xml. Při změnách stačí udělat zkoušku správnosti pouze pro nové položky.

Soubor **sitemap.xml** a **robots.txt** pošlete na Internet.
Soubor sitemap.xsl a sitemap.exe není na Internetu potřebný.

Můj malý soubor sitemap.xml vypadá takto :

```
<?xml version='1.0' encoding='UTF-8'?>
<?xml-stylesheet type='text/xsl' href='sitemap.xsl'?>
<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. Vedlejší produkt sitemap.html

V nové verzi jsem ještě doplnil generování souboru **sitemap.html** . Pozor, soubor se generuje automaticky a může přepsat Váš soubor se stejným jménem. Takový soubor používají i renomované portály. Slouží na primitivní orientaci pro jistý typ uživatele.

Nedá mi vrátit se do hluboké minulosti, když jsem poznal jednu paní, která psala služební korespondenci do jednoho souboru Doklady.602 v tehdejšímu programu T602 v MS DOS-u. Pak když potřebovala vytisknout jednu A4 stranu, našla si ji v tom souboru a nechala tisknout tu jednu stranu. Celé to zkrachovalo v okamžiku kdy jí havaroval ten soubor Doklady.602 ... Možná by stačilo každý doklad psát do samostatného souboru a mít pak lepší přehled a mít jistotu.

Aby nemuselo být vše v jednom adresáři a vůbec ne v jednom souboru můžete si vytvořit i více adresářů (složek) což zjednodušuje hledání v budoucnu. Tak to potom i vypadá ve Vaší internetové stránce. Mějme jako příklad, že píšete něco o orchidejích a Vaše lokální i internetové adresáře vypadají takto :

index.html v kořenovém adresáři

Zvierata - Podadresář

Zvierata - Psy - Buldog - Podadresář

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

Priroda - Podadresář

Priroda - Kvety - Podadresář

Priroda - Kvety - Orchidea - Podadresář

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

Pochopitelně ve Vašich .html souborech je odkaz na **Priroda**, v přírodě na **Kvety** a tam na **Orchidea**. Takže klikáním se tam dá snadno dostat. Jenže když potkáte na ulici Vaši sousedku důchodkyni která má internet, ona se tam z Vaší první stránky nedokliká a tedy Váš text o tom jak pěstujete orchideje nenajde a bude reklamovat.

Program sitemap.exe Vám automaticky vygeneruje soubor sitemap.html. Zda ho použijete je jen na Vás, můžete ale také nemusíte. Pokud ano, v souboru index.html v kořenovém adresáři by měl být snadno najditelný odkaz na tento soubor.

Soubor sitemap.html začne být nepříjemně velký když má více než 200 řádků, třeba v něm hledat tak jak to Vaše sousedka dělá v papírovém telefonním seznamu (knize). Například internetové obchody, kde jsou tisíce položek obvykle sitemap.html proto nepoužívají.

Abyste mohli v souboru sitemap.html měnit jeho vzhled je použit i soubor stylů **sitemap.css** Tento si v takovém případě také nakopírujte do kořenového adresáře. Je třeba si uvědomit, že soubor sitemap.html se generuje vždy automaticky a přepsal by Vaše dokonalé dílo.

Naopak **soubor sitemap.css je fixní** a ten se použije vždy v takovém stavu jako je.

Výsledek může vypadat takto :

<http://stofko.zohosites.com>

1. GALERIA Menu s referenciami na podadresare
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 rozdíl od souboru .xml zde hledá program sitemap.exe nejen cestu k souboru, ale i to co je napsáno v HTML značce **title**. To pak ukáže v menu. Značka title je víceméně povinná a text tam uvedený se zobrazí ve Vašem internetovém prohlížeči v hlavičce. V každém Vašem souboru to může vypadat takto :

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

nebo

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

Text mezi značkami title musí být pro program sitemap.exe v jednom řádku.

Takhle by to nešlo, i když by to možná prohlížeč akceptoval :

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

Skoro bych zapomněl na významnou výhodu programu sitemap.exe. Když totiž dodatečně změníte ve značce title text, např.

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

tento text program při příštím generování souboru sitemap.html automaticky převezme. Tedy nemusíte něco měnit na dvou místech, což zmenšuje výskyt nesrovnalostí.

A nakonec i o diakritice. Já řeším diakritiku vždy nakonec. U souboru sitemap.xml je to vcelku fuk, ale v souboru sitemap.html třeba definovat diakritiku. Abyste nemuseli do automaticky generovaného souboru zasahovat a abyste ani náhodou na to nezapomněli, program sitemap.exe čte Váš soubor index.html v kořenovém adresáři a tam hledat definici diakritiky, např.

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

Z toho řádku si zjistí jakou diakritiku používáte (utf-8, windows-1250 nebo jinou). Zjištěnou hodnotu pak použije při generování souboru sitemap.html. Pokud definici nenajde, nastaví charset = windows-1250.

Je jedno zda ve Vašem souboru použijete :

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

5. Rozšířené vyhledávání

Od verze V.16.02 bylo v programu zavedeno rozšířené vyhledávání. V omezeních výše se jako bod 3 uvádí, že program hledá v každém adresáři jen soubor **index.html** a ostatní .html soubory ignoruje. Tak mi to vyhovovalo pro galerii fotografií, viz moje programy :

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

[GALERIA Menu s referenciami na podadresáre](#)

Jenže jsou i takové situace v případě spíše textových stránek, že v jednom podadresáři je také více .html souborů. Při řešení toho jak programu Sitemap sdělit jaké řešení chci jsem postupoval podobně jako u výše uvedených dvou programů.

Když potřebujete aby Sitemap akceptoval i jiné .html soubory jako index.html, musíte to "vpašovat" do souboru index.html v hlavním (kořenovém) adresáři. Někde, stačí těsně před koncem vložte nestandardní značku sitemap takto:

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

Syntakticky čistší by bylo napsat <sitemap all />. Taková nestandardní značka se nezobrazí. Slovo **all** v angličtině značí vše-všechny. Podle toho se změní, tedy rozšíří rozsah souboru sitemap.xml i sitemap.html. Toto nastavení platí pro všechny podadresáře kromě hlavního (kořenového), kde program najde vždy jen index.html. Pokud by Vám to vadilo, napište !!!

6. Oznamy o chybách

Mohou nastat následující kritické situace na které program reaguje takto :

Cannot find Sitemap.txt

Jak již bylo řečeno bez souboru sitemap.txt program nefunguje. Vkopírujte si z distribučního balíku soubor sitemap.txt a upravte dle Vašich přání.

Cannot find URL in Sitemap.txt

Soubor sitemap.txt je prázdný, neobsahuje ani jen první řádek. Syntaxe adresy Vaší stránky tedy URL program sitemap.exe nekontroluje, akceptuje to co tam napíšete.

Cannot open directory XYZ

Na Vašem disku je nějaký vážný problém. Souborový systém sice vidí adresář např. XYZ, ale neví ho otevřít a tedy podívat se jaké jsou tam soubory. Taková chyba by se ve fungujícím Windows neměla vyskytnout !!!

Title not found

Do souboru sitemap.html se vloží tento text pokud Váš .html soubor neobsahuje syntakticky správnou značku **<title>**. Otevřete si patřičný soubor v podadresáři a opravte. Takový nedokonalý .html soubor byste neměli pro Google nabízet.

7. Závěr

Program je FREEWARE. Autor si ale vyhrazuje právo rozhodnout o tom kde bude program uložen pro stažení. Pokud hodláte program ke stažení uložit na jiný server, prosím informujte autora.

Program Sitemap.exe má v aktuální verzi následující vlastnosti :

Délka 8 192 Bajtov
Datum 02.01.2016
Čas 10:49:48

Ing. Branislav Štofko b.stofko@zoho.com