

WEBTAN

Web Traffic Analyzer

system pro sledování návštěvnosti webů

© Miroslav Pecka, 2004-2006

<http://webtan.pavucina.com>

Obsah

1.	Slovo (před) úvodem	3
2.	Úvod.....	4
3.	Předpoklady pro instalaci systému.....	5
4.	Instalace a konfigurace	6
4.1	Instalace a základní konfigurace systému	6
4.2	Konfigurační soubory a proměnné	8
4.3	Pokročilá konfigurace	9
4.3.1	Nastavení masky webu.....	9
4.3.2	Rozpoznání návštěv.....	11
4.3.3	Rozpoznání návratu.....	11
4.3.4	Ignorované parametry v URL sledovaného webu.....	12
4.3.5	Konfigurační proměnné – rady a doporučení.....	12
4.3.6	Zobrazování názvu stránky	13
5.	Ovládání systému	15
5.1	Zobrazení statistik (reportů)	15
5.2	Kategorie reportů.....	16
5.2.1	Souhrny	16
5.2.2	Stránky	16
5.2.3	Návštěvy.....	17
5.2.4	Hodiny a dny	17
5.2.5	Odkazovače	17
5.2.6	Systémová nastavení	17
5.3	Obecně o reportech.....	18
5.3.1	Absolutní a relativní četnosti.....	18
5.3.2	Nezobrazení reportu kvůli nedostatku dat.....	18
5.3.3	Neznámé a nezjištěné údaje	18
5.3.4	Nezobrazení číselné hodnoty v reportech	19
5.4	Administrace	20
6.	Další funkce a vlastnosti systému.....	21
6.1	Kontrola umístění měřicího kódu	21
6.2	Ignorování „prokliknutí“ z Webtanu.....	21
7.	Chybové stavy a jejich řešení	22
7.1	Nemožnost zobrazení systému Webtan.....	22
7.2	Nemožnost zobrazení reportů.....	22
8.	Otázky a odpovědi	23
8.1	Stránka s měřícím kódem hlásí chybu javascriptu. Co to znamená?	23
8.2	Kam vložit měřicí kód?	23
8.3	Bude měřicí kód fungovat ve všech prohlížečích?.....	24
9.	Slovníček pojmů a zkratk	25
10.	Schéma databáze systému	27
11.	Schéma provázanosti souborů v systému.....	28

1. Slovo (před) úvodem

Systém WEBTAN byl vytvářen jako součást bakalářské práce. Manuál, který právě čtete, byl vytvářen jako příloha bakalářské práce. Proto je psán značně formálně a možná poněkud rozvlekle.

Přesto věřím, že poslouží dobře i samostatně každému uživateli systému WEBTAN.

V případě zájmu o kompletní text bakalářské práce navštivte a <http://webtan.pavucina.com>.

2. Úvod

Webtan (WEB Traffic ANalyzer) je systémem pro sledování návštěvnosti webů, který dokáže zjistit o návštěvnosti a návštěvnících sledovaného webu mnohé užitečné informace. Ty je pak možno analyzovat a využít je pro optimalizaci webu nebo jen jako informaci o tom, „jak si sledovaný web stojí“ – jak se vyvíjí návštěvnost v čase apod. Přehled jednotlivých zjišťovaných ukazatelů a jejich statistické shrnutí (dále jen „reporty“) najdete níže v kapitole 5.2 Kategorie reportů.

Systém používá k prezentaci veškerých informací o návštěvnosti webů webové rozhraní. Pro bezproblémové používání systému (z pohledu uživatele – nikoliv správce systému) je třeba standardní webový prohlížeč s podporou HTML 4.0, kaskádových stylů (minimálně CSS 1) a javascriptu (doporučeno – javascript není podmínkou).

3. Předpoklady pro instalaci systému

Systém využívá technologii PHP ve spolupráci s MySQL. Pro správné fungování je třeba (na straně serveru):

- PHP 4.1 a vyšší
- v PHP zapnutou podporu GD knihovny ve verzi vyšší než 1.6
- v PHP zapnutou podporu BCMath funkcí
- MySQL 4.0 a vyšší

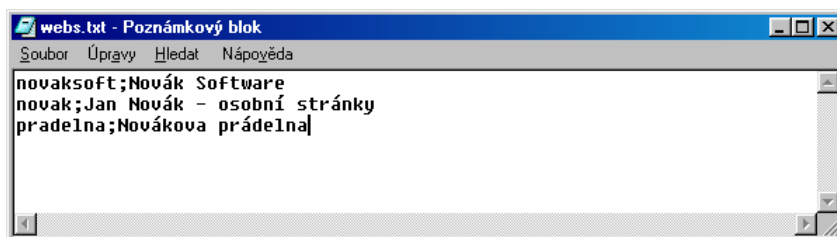
4. Instalace a konfigurace

4.1 Instalace a základní konfigurace systému

1. Do adresáře, v němž bude systém Webtan na serveru umístěn, rozbalíme archiv ZIP obsahující systém.
2. Otevřeme soubor *common.php*, který se nachází v kořenu (rootu) tohoto adresáře a zde upravíme obsah proměnně `$wta_url` na adresu odpovídající umístění systému v rámci serveru, na který systém instalujeme (např. <http://www.novaksoft.cz/statistika/>).
3. Vytvoříme novou prázdnou databázi v níž budeme uchovávat data systému Webtan.
4. V subadresáři *common* upravíme soubor *connect.php*. Do proměnných `$server_name`, `$db_user`, `$db_pass` a `$db_name` vyplníme databázový server, uživatelské jméno a heslo a jméno databáze (kterou jsme vytvořili v bodu 3). Uživatel databáze (pro systém Webtan) musí mít práva: CREATE, DROP, SELECT, INSERT, UPDATE, DELETE.
5. V subadresáři *config* upravíme soubor *webs.txt* a to tak, že pro každý web (který chceme sledovat) zapíšeme do souboru jeden řádek. Na řádek odpovídající příslušnému webu zapíšeme identifikátor (například `novaksoft`) webu a jeho „lidsky srozumitelný“ název (například `Novák Software`). Platí zde tato pravidla:
 - Tyto dvě hodnoty (identifikátor a název) musí být odděleny středníkem.
 - Identifikátor webu musí mít délku 3 až 10 znaků a smí se skládat pouze z malých písmen anglické abecedy (tedy znaky bez diakritiky) a číslic.
 - Tzv. „lidsky srozumitelný“ název webu se smí skládat pouze z malých a velkých písmen české abecedy, číslic, mezer a znaku pomlčka (v žádném případě nesmí obsahovat znak středník).

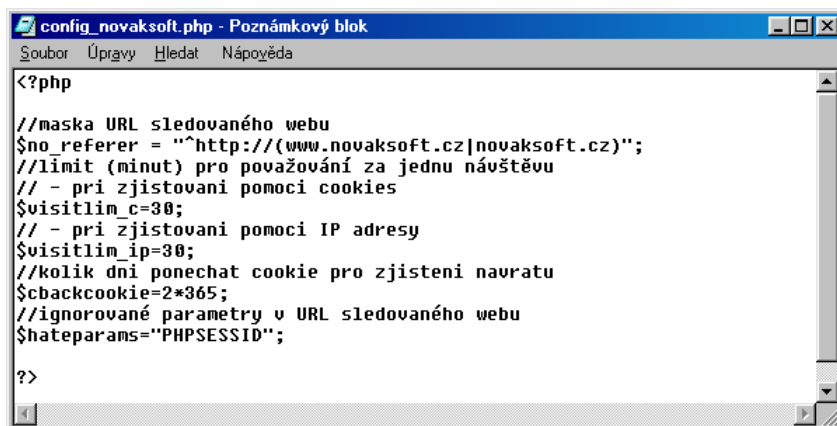
- V souboru *webs.txt* nesmí být kromě textu ve výše uvedené formě žádný další text.

Možné nastavení ukazuje obrázek.



6. Pro každý jednotlivý web (který jsme zapsali do souboru *webs.txt*) zkopírujeme (stále v rámci adresáře *config*) vzorový konfigurační soubor *config_sample.php* pod názvem *config_identifikator.php*, kde *identifikator* je identifikátor webu (např. *novaksoft*) uvedený v souboru *webs.txt*. V tomto souboru upravíme proměnnou *\$no_referer* tak, aby tato maska (resp. regulární výraz) pokrývala všechny URL stránek sledovaného webu. Více informací viz kapitola 4.3 Pokročilá konfigurace.

Možné nastavení ukazuje obrázek.



7. Nyní otevřeme ve webovém prohlížeči úvodní stránku systému (*index.php*), která se například nachází na adrese <http://www.novaksoft.cz/statistika/index.php> (pokud jsme do proměnné *\$wta_url* nastavili <http://www.novaksoft.cz/statistika/>,

resp. pokud jsme celý systém zkopírovali do příslušného adresáře na serveru). Zde klikneme na odkaz „Administrace“.

8. Nyní bychom měli mít na výběr všechny (v bodu 5) definované weby a je třeba vytvořit pro tyto weby tabulky, v nichž se budou shromažďovat získaná data.

Z roletového menu zvolíme libovolný web.

- a. Na zobrazené stránce nejdříve v rámečku „Společné tabulky“ klikneme na odkaz „Vytvořit všechny tabulky“. O úspěšném vytvoření tabulek a jejich naplnění iniciálními záznamy budeme informováni zprávou.
- b. Poté klikneme v rámečku označeném „Specifické tabulky pro...“ též na odkaz „Vytvořit všechny tabulky“. O úspěšném vytvoření tabulek a jejich naplnění iniciálními záznamy budeme informováni zprávou.

- c. Pro každý další web (který vybereme z roletového menu) opakujeme bod 8 b.

9. Když máme již vytvořeny všechny tabulky, nezbyvá než vložit do stránek webů (které si přejeme sledovat) měřicí kód. Stačí vybrat web, pro nějž chceme získat kód a v rámci administrace kliknout na odkaz „Kód pro stránky“. Vygenerovaný měřicí kód je třeba vložit do každé stránky, kterou hodláme sledovat a to do jejího těla, tj. mezi tagy <BODY> a </BODY>.

4.2 Konfigurační soubory a proměnné

V přecházející kapitole byl stručně popsán postup základní konfigurace. Kapitola následující (4.3) popisuje podrobně rozšířené možnosti konfigurace. Proto na tomto místě uvedme stručný přehled souborů a konfiguračních proměnných, které slouží ke konfiguraci systému Webtan.

Adresář	Soubor	Proměnná	Popis
kořenový adresář	common.php	\$wta_url	základ URL systému Webtan
common	connect.php	\$server_name	jméno databázového serveru

config	config_identifikator.php	\$db_user	uživatel databáze
		\$db_pass	heslo uživatele
		\$db_name	jméno databáze pro systém Webtan
		\$no_referer	maska URL sledovaného webu
		\$visitlim_c	pro rozeznání jednotlivých návštěv
		\$visitlim_ip	pro rozeznání jednotlivých návštěv
		\$cbcookie	pro rozeznání opakovaných návštěv (návratů)
		\$hateparams	ignorované parametry v URL sledovaného webu
	webs.txt		seznam sledovaných webů (identifikátor a název webu)

Poznámka: Text *identifikator* zastupuje konkrétní identifikátor webu (např. novaksoft) uvedený v souboru *webs.txt*.

Bližší vysvětlení funkce a možností nastavení jednotlivých konfiguračních proměnných objasňuje následující kapitola.

4.3 Pokročilá konfigurace

4.3.1 Nastavení masky webu

Jak již bylo uvedeno, proměnná `$no_referer` musí obsahovat regulární výraz odpovídající všem URL stránek sledovaného webu. Pokud tvorbu regulárních výrazů neovládáte, vše potřebné pochopíte z několika příkladů možných nastavení této konfigurační proměnné.

Příklad 1:

Všechny URL stránek webu začínají stejně. Potom stačí do proměnné napsat část URL, která je vždy společná a těsně před tuto URL napsat znak „^“. Obsah proměnné tedy může být například:

`^http://www.novaksoft.cz` nebo

`^http://www.freeweb.cz/novak/` nebo

`^http://pradelny.cz/praha/novak/`

Příklad 2:

Existuje několik (málo) možných začátků URL stránek sledovaného webu. Tato situace nejčastěji nastává, je-li možno na stránky přistupovat pomocí URL s „www.“ i bez „www.“.

V takovém případě je třeba na začátek proměnné opět zapsat znak „^“, za nímž následují kulaté závorky, které obsahují jednotlivé možné začátky URL oddělené znakem „|“ („svislítko“). Obsah proměnné tedy může být například:

`^(http://www.novaksoft.cz|http://novaksoft.cz) nebo`

`^(http://www.freeweb.cz/novak/|`

`http://www.novak.cz|http://novak.cz) nebo`

`^(http://www.pradelny.cz/praha/novak/|`

`http://pradelny.cz/praha/novak/)`

Pro úplnost je třeba doplnit, že:

- V příkladech jsme se (v zájmu zjednodušení) dopustili záměrně nepřesnosti.

V URL stránek v regulárním výrazu by samozřejmě každému znaku tečka mělo předcházet zpětné lomítko (neboť znak tečka má v regulárních výrazech zvláštní význam). Přesto v absolutní většině případů by i tento nepřesný zápis neměl způsobovat problémy.

- Regulární výraz musí být vždy zapsán vcelku na jednom řádku.
- Regulárními výrazy jsou míněny regulární výrazy dle normy POSIX 1003.2¹.
- Výše uvedené regulární výrazy by bylo možno zapsat efektivněji. To by však vyžadovalo exkurz do teorie tvorby regulárních výrazů.

¹ POSIX 1003.2 regular expressions – <http://www.die.net/doc/linux/man/man7/regex.7.html>

- Pokud máte zájem o hlubší proniknutí do tajů regulárních výrazů, navštivte web www.regularnivrazy.info, kde se dozvíte více.

Účel nastavování masky webu je blíže vysvětlen v kapitole 6.1 Kontrola umístění měřicího kódu.

4.3.2 Rozpoznání návštěv

Do jedné návštěvy jsou počítána všechna zobrazení stránek sledovaného webu, uskutečněná z jednoho počítače během určité relativně krátké doby. Přesněji řečeno, je dán časový interval mezi zobrazením jednotlivých stránek, který nesmí být překročen, má-li být zobrazení stránky považováno za zobrazení náležející do aktuální návštěvy. Tento interval je možno nastavit samostatně pro každý sledovaný web. Doporučená hodnota je 30 minut. Tuto hodnotu je možno nastavit pomocí konfiguračních proměnných `$visitlim_c` a `$visitlim_ip`. Obě proměnné jsou iniciálně nastaveny na hodnotu 30 (minut). Proměnná `$visitlim_c` je použita v případě, že návštěvník stránek je identifikován pomocí cookies. V případě, že není možno návštěvníka stránek identifikovat pomocí cookies, provádí se identifikace pomocí IP adresy – v tom případě je použita hodnota proměnné `$visitlim_ip`. Doporučeno je nastavit obě proměnné na stejnou hodnotu nebo proměnnou `$visitlim_c` nastavit na větší hodnotu než hodnotu `$visitlim_ip`.

4.3.3 Rozpoznání návratu

Systém v reportu návštěv udává celkový počet návštěv a (z toho) počet návratů. Za návrat je považována návštěva návštěvníka, který již někdy v minulosti sledovaný web navštívil. Jak dlouho si chceme „pamatovat minulost“, přesněji řečeno jaký smí být nejdelší časový interval (počet dní) mezi návštěvami návštěvníka, aby každá jeho další návštěva byla považována za návrat, udává proměnná `$backcookie`. Iniciální nastavení je na hodnotu 730 (dní), tj. 2 roky.

4.3.4 Ignorované parametry v URL sledovaného webu

Možnost ignorování vybraných parametrů v URL stránek sledovaného webu je jednou ze specifických funkcí systému, kterou lze využít zvláště jsou-li sledovány stránky dynamicky generovaného webu, kde URL dynamicky generované stránky obsahuje mnoho parametrů. Čím více je v URL parametrů, tím větší je počet možných kombinací těchto parametrů a potažmo počet unikátních zobrazitelných URL sledovaného webu. Mohlo by se tak stát, že výsledkem statistiky zobrazení stránek bude seznam několika set unikátních URL, přičemž četnost zobrazení jednotlivých URL bude v řádu jednotek. Pro zjištění, jaké stránky jsou navštěvovány, nás většinou některé parametry vůbec nemusí zajímat (různé pomocné parametry obsahující například informace o uživatelském nastavení vzhledu webu apod.). Proto je možno v systému Webtan pro každý sledovaný web definovat libovolné množství parametrů, které se mají v URL sledovaného webu ignorovat (resp. z URL extrahovat). Názvy těchto parametrů je možno nastavit pomocí konfigurační proměnné `$hateparams`, v níž je seznam parametrů, které se mají ignorovat. Jednotlivé parametry je třeba oddělit čárkou. V názvech parametrů není brán ohled na rozdíl v malých a velkých písmenech (case insensitive). Pokud tuto funkci nevyužijete, můžete ponechat iniciální nastavení této konfigurační proměnné nebo ji zcela smazat. Proměnná `$hateparams` je jedinou nepovinnou konfigurační proměnnou.

4.3.5 Konfigurační proměnné – rady a doporučení

Kromě proměnné `$hateparams` jsou všechny ostatní proměnné povinné, tzn. jejich neexistence či nenastavení by způsobilo nefunkčnost systému.

Všechny konfigurační proměnné je třeba nastavit před vložením měřicího kódu do stránek a po dobu sledování webu hodnoty těchto proměnných již neměnit. Změna hodnot konfiguračních proměnných by mohla vést ke zkreslení výsledků měření, resp. k nemožnosti kvalitního srovnání výsledků naměřených před a po změně hodnot.

4.3.6 Zobrazování názvu stránky

V reportu „Stránky“ se místo nepříliš lidsky vyhlížejících URL stránek můžou zobrazovat jejich názvy (například: „Ceník“, „Nabídka služeb“ apod.). Aby se tak mohlo stát, je třeba, aby měřicí kód vložený do každé konkrétní stránky obsahoval také „lidský“ (resp. obyčejnému uživateli srozumitelný) název konkrétní stránky. Toto je možno snadno realizovat především u dynamicky generovaných stránek, kde se také měřicí kód může dynamicky generovat. Standardně vygenerovaný měřicí kód vypadá například takto:

```
<!-- webtan -->
<script language="JavaScript" type="text/javascript">
<!--
document.write("<img
src=\"http://www.novaksoft.cz/statistika/counter.php?wid=pok&pgtitle=&refer
er=\" + escape(top.document.referrer) + "&screenres=\" + screen.width + "x" +
screen.height + "&colordepth=\" + screen.colorDepth + "\" width=\"1\"
height=\"1\" alt=\"Webtan\">");
// -->
</script>
<noscript>

</noscript>
<!-- KONEC - webtan -->
```

Podstatný je parametr `pgtitle` (v kódu zvýrazněn), který je standardně prázdný. Pokud parametr `pgtitle` „naplníme“ názvem stránky, předáme tak tuto informaci systému a ten již bude stránku „znát“ i pod tímto názvem. Využití možnosti dynamicky generovat obsah parametru `pgtitle` předpokládá znalost některého skriptovacího jazyka (typicky PHP či ASP). Název stránky předávaný jako `pgtitle` by měl být ve shodném kódování, jaké používá systém Webtan (tj. Windows-1250) a měl by být nejdříve zakódován pomocí PHP funkce `urlencode()` nebo jí ekvivalentní.

Pomocí PHP dynamicky vygenerovaný měřicí kód může vypadat následovně:

```
<!-- webtan -->
<script language="JavaScript" type="text/javascript">
<!--
document.write("<img
src=\"http://www.novaksoft.cz/statistika/counter.php?wid=pok&pgtitle=<?php
echo urlencode($nazevstranky); ?>&referer=" + escape(top.document.referrer)
+ "&screenres=" + screen.width + "x" + screen.height + "&colordepth=" +
screen.colorDepth + "\" width=\"1\" height=\"1\" alt=\"Webtan\">");
// -->
</script>
<noscript>

</noscript>
<!-- KONEC - webtan -->
```

5. Ovládání systému

WWW rozhraní systému je možno rozdělit na dvě části – zobrazení statistik (reportů) a administraci. Výběr, kterou část systému si přejeme zobrazit, nám umožní úvodní stránka systému (soubor *index.php* v kořenovém adresáři systému), kde stačí zvolit příslušný odkaz („Zobrazení statistik“ či „Administrace“).

5.1 Zobrazení statistik (reportů)

V prvním kroku je třeba z menu vybrat web, jehož reporty budeme chtít zobrazit. Nyní se zobrazí nabídka nabízených reportů, resp. nabídka šesti kategorií do nichž jsou reporty rozděleny.

Reporty můžeme rozdělit na základní dvě skupiny – a to „Souhrny“ a jednotlivé reporty. Souhrny obsahují číselná vyjádření některých sledovaných veličin za jednotlivé dny v rámci zvoleného měsíce (souhrn za měsíc) nebo za jednotlivé měsíce v rámci zvoleného roku (souhrn za rok). Pro souhrn za měsíc vybereme příslušný rok a měsíc. Pokud místo čísla měsíce zvolíme „--“, zobrazí se souhrn za celý rok.

Jednotlivé („nesouhrnné“) reporty obsahují specifické textové a číselné informace pro libovolně zvolené období (ode dne, měsíce, roku – do dne, měsíce, roku). Datумы udávající zobrazený časový interval se zadávají do textových polí „Od:“ a „Do:“ ve tvaru RRRR-MM-DD (R=rok, M=měsíc, D=den).

Rozdělení reportů do jednotlivých kategorií zobrazuje tabulka.

Souhrny	Jednotlivé reporty				
Souhrny	Stránky	Návštěvy	Hodiny a dny	Odkazovače	Systémová nastavení
Měsíční souhrny nebo Roční souhrny	Zobrazené stránky	Shrnutí	Přístupy po hodinách dne	Přístupy z odkazujících serverů	Prohlížeče
	Vstupní body	Domény	Přístupy po dnech v týdnu	Vyhledávače a vyhledávací fráze	Operační systémy
Statistické charakteristiky		Kraje, města, oblasti			Rozlišení obrazovky
					Barevná hloubka

5.2 Kategorie reportů

5.2.1 Souhrny

Souhrny – souhrnný report umožňuje zobrazit několik základních ukazatelů návštěvnosti za období vybraného měsíce či roku. Zobrazeny jsou údaje: počet zobrazených stránek, počet realizovaných návštěv, počet návratů, počet unikátních IP adres (označeno „UIP“) a poměrový ukazatel počet zhlédnutých stránek v rámci jedné návštěvy webu (označeno „Str./Návštěva“). Pro každou skupinu těchto údajů jsou vypočteny některé statistické ukazatele (minimum, maximum, medián, aritmetický průměr, rozptyl a směrodatná odchylka).

5.2.2 Stránky

Zobrazené stránky – „lidský“ (resp. obvyčejnému uživateli srozumitelný) název či URL stránky (více viz kapitola 4.3.6 Zobrazování názvu stránky), absolutní a relativní počet zobrazení stránky

Vstupní body – „lidský“ název či URL vstupního bodu, absolutní a relativní počet zobrazení stránky

5.2.3 Návštěvy

Shrnutí – počet zobrazených stránek, návštěv, návratů a unikátních IP (označeno „UIP“) adres návštěvníků

Domény – absolutní a relativní počet návštěv, stránek a UIP při přístupu z dané top-level domény

Kraje, města, oblasti – absolutní a relativní počet návštěv, stránek a UIP při přístupu z daného kraje, města či oblasti

5.2.4 Hodiny a dny

Přístupy po hodinách dne – počet zobrazených stránek, uskutečněných návštěv a návratů a UIP v jednotlivých hodinách dne (součet za všechny dny ve sledovaném období)

Přístupy po dnech v týdnu – počet zobrazených stránek, uskutečněných návštěv a návratů a UIP v jednotlivých dnech týdne (součet za všechny dny resp. týdny ve sledovaném období)

5.2.5 Odkazovače

Přístupy z odkazujících serverů – názvy (resp. webové adresy) webů, z nichž návštěvník přistoupil (přišel) na stránky sledovaného webu, absolutní a relativní počet takových přístupů

Vyhledávače a vyhledávací fráze – názvy (resp. webové adresy) vyhledávačů a klíčová slova (fráze), pomocí nichž návštěvník přistoupil (přišel) na stránky sledovaného webu, absolutní a relativní počet takových přístupů

5.2.6 Systémová nastavení

Prohlížeče – typy a verze prohlížečů, které návštěvníci webu použili při prohlížení webu, absolutní a relativní počet zobrazených stránek a uskutečněných návštěv

Operační systémy – typy a verze operačních systémů návštěvníků webu, absolutní a relativní počet zobrazených stránek a uskutečněných návštěv

Rozlišení obrazovky – rozlišení obrazovky nastavené na počítačích návštěvníků, kteří navštívili web, absolutní a relativní počet zobrazených stránek a uskutečněných návštěv

Barevná hloubka – barevná hloubka nastavená na počítačích návštěvníků, kteří navštívili web, absolutní a relativní počet zobrazených stránek a uskutečněných návštěv

5.3 Obecně o reportech

5.3.1 Absolutní a relativní četnosti

Ve většině reportů jsou jednotlivé ukazatele udávány v absolutních číslech (například: „počet návštěv z domény CZ: 20“) i v procentech vzhledem k celkovému počtu (například: „počet návštěv z domény CZ: 10%“ – tzn. celkový počet návštěv ze všech domén musel být 200). Pokud je v záhlaví tabulky (reportu) text „[abs.][%]“, znamená to, že levý sloupec obsahuje absolutní hodnoty a pravý sloupec hodnoty relativní (percentuální vyjádření).

5.3.2 Nezobrazení reportu kvůli nedostatku dat

Pokud v období, které je nastaveno jako období pro zobrazení reportů, nebyla získána data potřebná pro vytvoření konkrétního reportu pro toto období, je zobrazena informační zpráva „Pro dané období nejsou k dispozici žádné informace.“. Zobrazení reportu pravděpodobně dosáhnete, zvětšíte-li časový interval („Od:“ – „Do:“) pro zobrazení reportů.

5.3.3 Neznámé a nezjištěné údaje

Za určitých okolností nemusí být systém schopen zjistit zcela všechny údaje. Jedná se především o tyto údaje: URL zobrazené stránky, doménu 1. řádu z níž návštěvník přišel, návštěvníkem používaný prohlížeč a operační systém nebo nastavené rozlišení obrazovky a nastavenou barevnou hloubku.

Blíže k jednotlivým případům:

- URL zobrazené stránky se nezobrazí především v případě, že měřicí kód byl zobrazen bez jeho vložení do stránky nebo nebylo možno získat URL stránky do níž byl vložen. Takový případ je v reportech označen jako „(nezjištěno, přímé volání počítadla)“.
- Doména 1. řádu nemusí být zjištěna, pokud nelze zjistit doménový název návštěvníkova počítače. Takový případ je v reportech označen jako „?? (Nezjištěno)“.
- Používaný prohlížeč a operační systém nemusí být identifikován, jedná-li se o prohlížeč či operační systém, který systém Webtan neumí zařadit do některé z definovaných skupin. Tato situace je však výjimečná. Takový případ je v reportech označen jako „Jiný, nezjištěno“.
- Rozlišení obrazovky a barevná hloubka nastavená na počítači návštěvníka je zjišťována pomocí javascriptu. Pokud návštěvníkův prohlížeč javascript nepodporuje nebo ho má návštěvník ve svém prohlížeči vypnutý, není možno požadované údaje zjistit. Takový případ je v reportech označen jako „Jiné, nezjištěno“ (resp. „Jiná, nezjištěno“).

5.3.4 Nezobrazení číselné hodnoty v reportech

Za určitých okolností se v tabulce, kde by bylo možno očekávat číslo, žádné číslo nezobrazí.

Takové situace jsou dvě:

- „Nemá smysl počítat.“ – Tato situace je signalizovaná zobrazením textu „---“ (s popiskem „Nemá smysl počítat.“) a vyskytuje se pouze v tabulce statistických ukazatelů v reportu „Souhrny“. Zde je na místě, neboť je zobrazen v řádku se součty hodnot jednotlivých ukazatelů a sloupci poměrového ukazatele „Str./Návštěva“. Je zřejmé, že součet hodnot tohoto poměrového ukazatele by neměl žádný smysl.

- „Není možno spočítat!“ – Tato situace signalizovaná zobrazením textu „[?]“ (s popiskem „Není možno spočítat!“) odpovídá stavu, kdy by při pokusu o výpočet hodnoty došlo k dělení nulou. K takové situaci může dojít, pokud v určitém časovém období (které je nastaveno pro zobrazení reportů) bylo zaznamenáno zobrazení stránky, ovšem nebyla zaznamenána (resp. započata) žádná návštěva. Potom by při pokusu o výpočet počtu zobrazených stránek na jednu návštěvu (ukazatel „Str./Návštěva“) i při výpočtu procentuálního vyjádření počtu návštěv docházelo k dělení nulou.

5.4 Administrace

O administrační části (modulu) systému již byla zmínka v kapitole 4.1 Instalace a základní konfigurace systému (bod 7 a 8). Na tomto místě uveďme všechny funkce administračního modulu, který můžeme rozdělit na tyto 3 části:

- Správa tabulek – umožňuje vytvářet a mazat tabulky s daty; umožňuje sledovat velikost jednotlivých tabulek náležejících každému sledovanému webu. Pro každou tabulku je uváděn počet záznamů v tabulce, velikost dat v tabulce samotné, velikost indexu pro danou tabulku a součet těchto dvou velikostí. Jedná se o údaje ryze technického charakteru, které „obyčejného uživatele“ systému Webtan v podstatě nemusí zajímat.
- Nastavené parametry – umožňuje zobrazení aktuálního nastavení konfiguračních proměnných. Nastavení (resp. změnu) těchto parametrů je však třeba provést pomocí konfiguračního souboru – více viz kapitola 4 Instalace a konfigurace.
- Kód pro stránky – umožňuje zobrazení měřicího kódu, který je třeba vložit do stránek sledovaného webu. Pro odstranění případných problémů s měřícím kódem nahlédněte do kapitoly 8 Otázky a odpovědi.

6. Další funkce a vlastnosti systému

V kapitole 4.3 Pokročilá konfigurace je popsáno nastavování funkcí a vlastností systému. Na tomto místě podrobněji popíšeme dvě vlastnosti. Zatímco v kapitole 4.3.1 (Nastavení masky webu) je popsán postup nastavení konfigurační proměnné `$no_referer` – v kapitole 6.1 (Kontrola umístění měřicího kódu) je detailněji vysvětlen účel jejího nastavování. V kapitole 6.2 (Ignorování „prokliknutí“ z Webtanu) je popsána specifická funkce systému, která má za účel eliminovat případné zkreslení statistik.

6.1 Kontrola umístění měřicího kódu

Měřicí kód má za úkol zaznamenávat zobrazení pouze sledovaného webu, resp. souboru stránek, jejichž URL odpovídá masce URL zadané pomocí konfigurační proměnné `$no_referer` (více viz kapitola 4 Instalace a konfigurace). Při každém načtení měřicího kódu se provede kontrola, zda je měřicí kód vložen do stránky, jejíž URL odpovídá konfiguraci pro daný sledovaný web. Pokud tomu tak není, je zobrazení takové stránky systémem zcela ignorováno. Je tak zamezeno tomu, aby mohl někdo cizí vložením „vašeho měřicího kódu“ do svých stránek zkreslit výsledky měření.

6.2 Ignorování „prokliknutí“ z Webtanu

V reportu „Stránky“ jsou zobrazené URL stránek (event. jejich názvy) přímo jako odkazy na příslušnou stránku. Při sledování statistik webu je přímá možnost zobrazit určitou konkrétní stránku sledovaného webu jistě užitečná, ovšem takovými návštěvami bychom zkreslovali celkové statistiky. Pravděpodobně nás totiž nebudou zajímat „naše vlastní zobrazení“, která provedeme v souvislosti s analýzou reportů v systému Webtan. Proto je systém nastaven tak, aby zobrazení stránky sledovaného webu, na kterou bylo přistoupeno odkazem z reportu „Stránky“, ignoroval.

7. Chybové stavy a jejich řešení

7.1 Nemožnost zobrazení systému Webtan

Statistická část (zobrazení reportů) či administrační část systému se nezobrazí v případě, že je v souboru `connect.php` v adresáři `common` špatně nastaveno připojení k databázovému serveru či špatné jméno databáze. Pro správné nastavení viz kapitolu 4 Instalace a konfigurace.

7.2 Nemožnost zobrazení reportů

Před samotným zobrazením reportů dochází ke kontrole konfiguračního souboru a tabulek potřebných pro zobrazení reportů. Pokud není kontrola úspěšná, tj. konfigurační soubor či potřebné tabulky nejsou v pořádku, je místo reportu zobrazeno hlášení „Klíčová chyba! / Fatal error!“ doprovázené popisem problému. Pro odstranění problému je třeba správně nastavit proměnné v konfiguračním souboru a vytvořit potřebné tabulky. Více viz kapitola 4 Instalace a konfigurace. Kontrolu nastavení konfiguračních proměnných a přítomnosti všech potřebných tabulek je možno provést též v administrační části systému Webtan.

Poznámka: Po dobu, kdy není možné zobrazení reportů, nejsou také pomocí měřicího kódu získávána žádná data, protože i měřicí kód vyžaduje bezchybný konfigurační soubor a tabulky příslušející sledovanému webu.

8. Otázky a odpovědi

8.1 Stránka s měřícím kódem hlásí chybu javascriptu. Co to znamená?

V zásadě jsou dvě možnosti proč se tak děje:

Podařilo se vám omylem poškodit vygenerovaný měřicí kód. V takovém případě vygenerujte měřicí kód znovu pomocí volby „Kód pro stránky“ v administrační části systému. Takový kód vložte znovu do stránek na místo původního měřicího kódu. Problémům předejdete, budete-li vkládat měřicí kód do zdrojového kódu stránek v „obyčejném“ textovém editoru. Programy typu Microsoft FrontPage apod. mohou při vkládání kódu přidávat nežádoucí formátovací znaky!

Pokud jsou na sledovaném webu používány rámy (frames) a stránka nejvyššího rámu je na jiném serveru (v jiné doméně 2. řádu) než stránka obsahující měřicí kód, javascript používaný v měřícím kódu pro zjištění URL odkazujícího serveru si s tím „neumí poradit“. Tato neschopnost je vlastností javascriptu jako takového, proto jedinou možností je vyhnout se situaci, kdy nejvyšší rám je na jiném serveru (například: <http://novak.zde.cz>) než stránka s měřícím kódem (např.: <http://mujweb.cz/novak/uvod.htm>).

8.2 Kam vložit měřicí kód?

Měřicí kód je třeba vložit do těla stránky (tj. mezi tagy `<BODY>` a `</BODY>`). V rámci těla stránky může být kód vložen téměř kdekoliv. Pokud chcete, aby byly zaznamenány jen stránky, které byly prohlížečem návštěvníka zcela načteny, vložte měřicí kód nejlépe až na úplný konec stránky. Měřicí kód se ve sledované stránce zobrazuje jako průhledný obrázek PNG o velikosti 1x1 pixel.

8.3 Bude měřicí kód fungovat ve všech prohlížečích?

Měřicí kód bude fungovat v každém prohlížeči, který podporuje zobrazení obrázků ve formátu PNG a nemá zobrazování obrázků direktivně vypnuto. V prohlížečích, které nepodporují javascript nebo jeho podporu mají vypnutou, bude měřicí kód fungovat také, pouze nebudou zjišťovány některé specifické údaje (rozlišení obrazovky, nastavená barevná hloubka obrazovky, odkazovače).

9. Slovníček pojmů a zkratek

DB

databáze

doména

unikátní označení webserveru složené z alfanumerických znaků, které je mu přiřazováno; rozlišujeme domény prvního řádu (*TLD – top level domain*), jimiž jsou (pevně dané) kombinace písmen vpravo od poslední tečky v adrese webové stránky (např. .cz, .com, .net, .org, .edu, .info atp.), domény druhého řádu (prodejceaut.cz) a třetího řádu (prazsky.prodejceaut.cz); existuje zeměpisné rozdělení domén prvního řádu např.: .cz – Česko, .de – Německo, .at – Rakousko nebo např. doména .edu znamená, že webserver takto označený patří nejspíše vzdělávací instituci (education)

IP adresa

unikátní číselné označení počítače (či sítě), přiřazené za účelem jednoznačné identifikace jinými počítači (resp. sítěmi)

kód pro stránky

viz *měřicí kód*

měřicí kód

HTML kód zajišťující sběr dat pro následné vytváření *reportů*

návrat

návštěva uskutečněná návštěvníkem, který sledovaný *web* již dříve navštívil

návštěva

zobrazení jedné nebo několika stránek daného *webu* po sobě v rámci relativně krátké doby

odkazovač

též referer; WWW stránka (nebo *web*), na níž se nachází odkaz, který byl použit k *návštěvě* stránky sledovaného *webu*

report

souhrn informací získaných pomocí *měřicího kódu*, zobrazený nejčastěji formou tabulky

TLD – top level domain

doména 1. řádu – viz *doména*

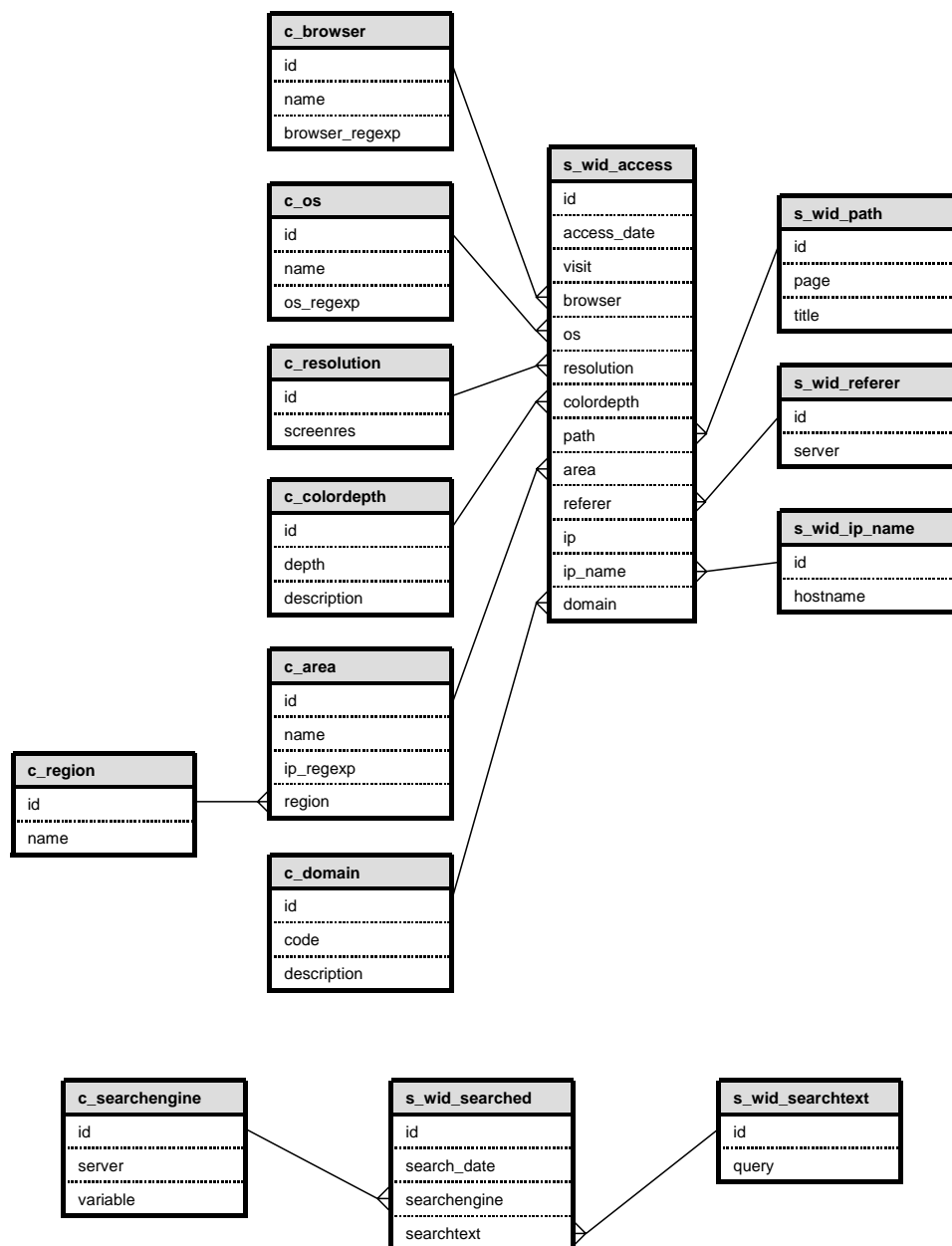
UIP

unikátní *IP adresa*

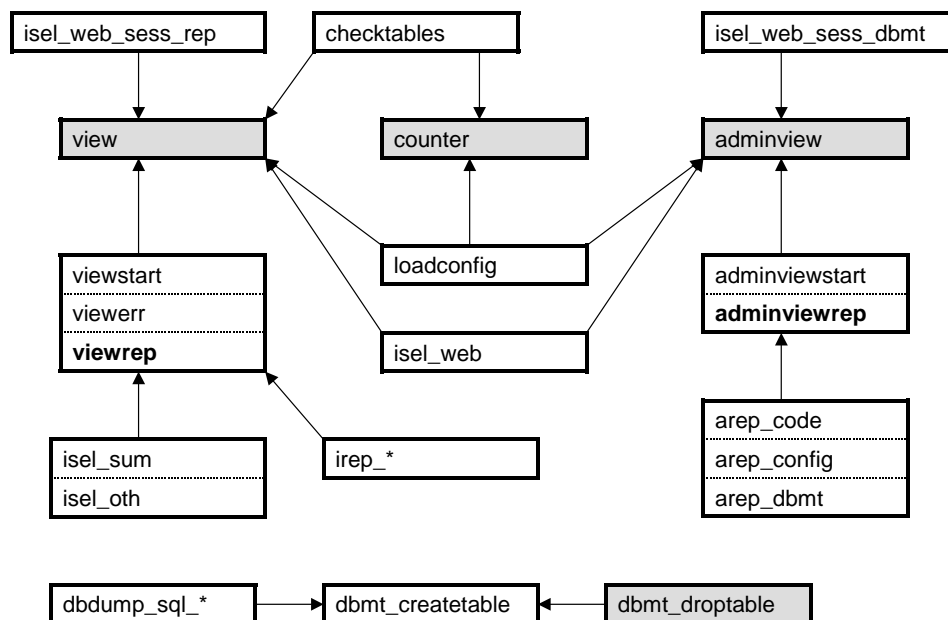
web

soubor WWW stránek tvořící logický celek; většinou mají stránky téhož webu shodnou *doménu* 2. či 3. řádu

10. Schéma databáze systému



11. Schéma provázanosti souborů v systému



Poznámky:

- Všechny soubory ve schématu jsou PHP skripty a jejich názvy mají extenzi „.php“.
- Provázaností je míněno vzájemné vkládání PHP skriptů – „includování“.
- Schéma zobrazuje pouze základní soubory systému (tj. nezobrazuje všechny soubory „.php“, které jsou nutné k fungování systému).

Vysvětlivky:

- název souboru zvýrazněný tučným písmem identifikuje soubor, do nějž se „includuje“ soubor
- do souborů zvýrazněných šedým pokladem se ještě „includují“ soubory: *common.php* a *connect.php*
- ze souborů oddělených tečkovanou linkou se „includuje“ pouze jeden

- ze skupiny souborů označených pomocí „*“ se „include“ pouze jeden