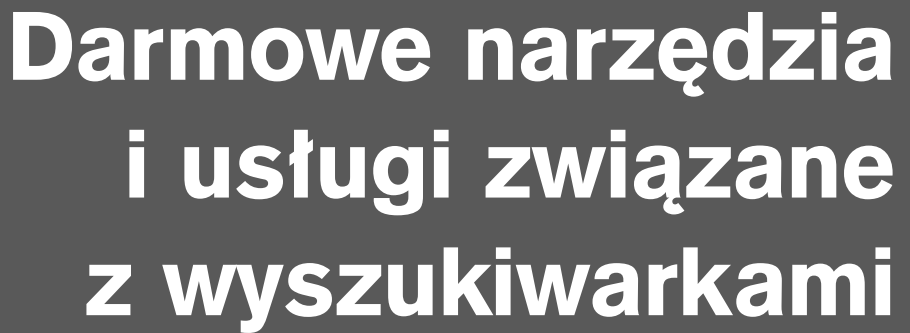
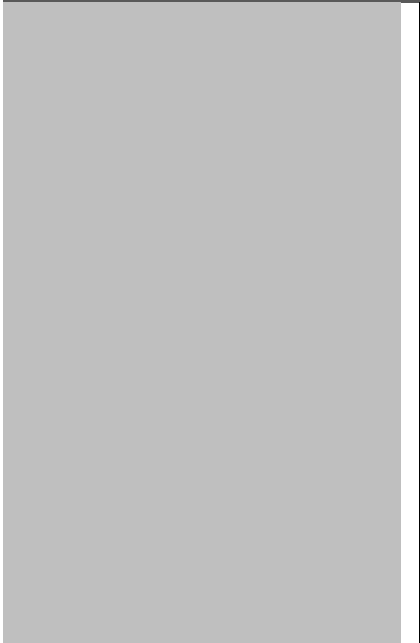




# 10



## **Darmowe narzędzia i usługi związane z wyszukiwarkami**



**Możesz wysłać informacje o swojej witrynie do wyszukiwarek oraz zobaczyć ją z ich perspektywy, wykorzystując darmowe narzędzia Yahoo! i Google.**

**Yahoo! i Google oferują kilka darmowych i bardzo przydatnych narzędzi oraz usług ułatwiających zgłoszenie witryny w wyszukiwarkach i uzyskanie bardziej szczegółowych informacji na temat jej indeksowania. Te narzędzia umożliwiają również zdobycie wiedzy o linkach przychodzących, słowach kluczowych stosowanych przy wyszukiwaniu i innych danych istotnych z punktu widzenia poprawy wyszukiwalności.**

**Zanim zapoznasz się ze statystyką, musisz upewnić się, że wyszukiwarki indeksują zawartość Twojej witryny. Możesz ułatwić zadanie robotowi wyszukiwarki poprzez opracowanie mapy witryny.**

## Utworzenie i zgłoszenie pliku `sitemap.xml`

Z historycznego punktu widzenia komunikacja pomiędzy webmasterami a wyszukiwarkami uległa znacznemu ograniczeniu. W przeszłości po zbudowaniu witryny następowała rejestracja jej adresu URL w indeksach głównych wyszukiwarek, aby zachęcić ich roboty do zindeksowania stron należących do witryny. Dysponując jedynie adresem URL strony głównej, roboty wyszukiwarek mogą nie zauważyć niektórych stron, zwłaszcza tych, które są dostępne tylko za pośrednictwem systemu wyszukiwania. Kilka lat temu specjaliści z firmy Google zdali sobie sprawę, że ten problem może niekorzystnie wpłynąć na wszechstronność indeksu i znaleźli proste rozwiązanie zwane *sitemap.xml*.

W czerwcu 2005 firma Google wprowadziła protokół mapy witryny w standardzie XML, umożliwiając webmasterom przekazanie struktury witryny do wyszukiwarek w celu zapewnienia dokładnej indeksacji (<http://sitemaps.org>). Dzisiaj, ponieważ protokół *sitemap.xml* jest również uznawany przez wyszukiwarki Yahoo!, Ask i MSN Live Search, ten sam plik XML jest wykorzystywany przez wszystkie główne wyszukiwarki, ułatwiając im zindeksowanie witryny.

Protokół *sitemap.xml* umożliwia także webmasterom dołączenie informacji o każdej stronie, w tym datę jej ostatniej aktualizacji, częstotliwość zmian oraz pozycję strony w hierarchii całej witryny. Dzięki tym dodatkowym informacjom roboty wyszukiwarki może lepiej wykonać swoją pracę.

Struktura pliku *sitemap.xml* jest stosunkowo prosta. Oto krótki przykład zawierający znaczniki typowe dla tego protokołu:

```
<?xml version="1.0" encoding="UTF-8"?>
<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://przyklad.pl/</loc>
    <priority>1.0</priority>
    <changefreq>daily</changefreq>
  </url>

  <url>
    <loc>http://przyklad.pl/o_witrynie/</loc>
    <priority>0.5</priority>
    <changefreq>monthly</changefreq>
  </url>

</urlset>
```

Chociaż przykład zawiera tylko dwa adresy URL stron należących do witryny, protokół *sitemap.xml* może opisywać witryny w dowolnej skali. Na początku pliku znajduje się wstępny zapis XML oraz definicja użytego schematu. Nie jest to jednak część najistotniejsza. Zwróć uwagę na dwie pary znaczników `<url>`, z których każda zawiera inne informacje. Każdy znacznik `<url>` dotyczy pojedynczej strony, jej położenia, priorytetu i częstotliwości zmian. Ponieważ strona główna jest najważniejszą stroną na witrynie, jej priorytet również jest najwyższy.

Definicja priorytetu i częstotliwości zmian strony w obrębie znacznika `<url>` nie jest obowiązkowa. Znacznik `<priority>` zawiera liczbę dziesiętną z przedziału od 0,0 do 1,0, przy czym 1,0 oznacza najwyższy priorytet. Dzięki temu roboty wyszukiwarek wiedzą, które strony są dla Ciebie najważniejsze, i przyjmują podobne priorytety w swoim indeksie.

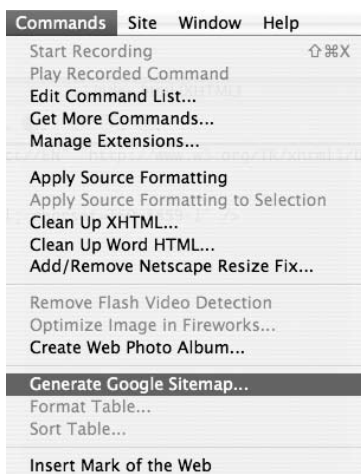
Znacznik `<changefreq>` zawiera ogólną informację na temat częstotliwości zmian strony, ale nie ma to bezpośredniego przełożenia na częstotliwość indeksacji. Listę możliwych wartości znacznika `<changefreq>` oraz inne informacje o znacznikach, które mogą wystąpić w pliku *sitemap.xml*, znajdziesz pod adresem <http://www.sitemaps.org/protocol.php>.

Plik jest dość prosty i stosunkowo łatwo można go zbudować. Gdyby Twoja witryna miała setki lub tysiące stron, ręczne tworzenie takiego pliku byłoby jednak zbyt czasochłonne.

Na szczęście istnieje szereg aplikacji desktopowych i internetowych, które pozwalają na automatyzację tego procesu i generację pliku *sitemap.xml*, zwalniając Cię od żmudnej pracy nad tworzeniem dokumentów XML. Są też dostępne skrypty w różnych językach, systemy zarządzania treścią i aplikacje internetowe wspomagające różne prace z tym związane.

Wyczerpującą listę dostępnych programów znajdziesz na stronie [http://code.google.com/sm\\_thirdparty.html](http://code.google.com/sm_thirdparty.html). Niektóre z nich są darmowe, inne kosztują kilka dolarów. Jeśli jesteś użytkownikiem Dreamweavera, może Cię zainteresować darmowa wtyczka z generatorem mapy witryny w standardzie Google opracowana przez George'a Petrova i udostępniona na witrynie DMXZone.com (<http://www.dmxzone.com/showDetail.asp?TypeId=3&NewsId=10538>). Ilustruje to rysunek 10.1. Po pobraniu i zainstalowaniu wtyczki musisz uruchomić ponownie Dreamweavera i zdefiniować witrynę, która ma być poddana analizie. W tym celu wybierz polecenie *Site* (witryna), następnie *New Site* (nowa witryna) i wpisz niezbędne informacje.

**RYСУNEK 10.1.** Generator mapy witryny w standardzie Google George'a Petrova jest darmową wtyczką do Dreamweavera, która znacznie upraszcza tworzenie plików *sitemap.xml*



W następnym kroku wybierz *Commands* (polecenia), *Create Google Sitemap* (utwórz mapę witryny w standardzie Google) i nakładka rozpocznie swoją pracę. Wynikowy plik *sitemap.xml* zostanie umieszczony w głównym folderze Twojej witryny.

Jeżeli korzystasz z języka skryptowego po stronie serwera w celu budowania systemu szablonów stron, utworzenie mapy witryny stanie się problematyczne. W tym przypadku będzie potrzebne inne narzędzie do tworzenia pliku *sitemap.xml*.

Taką aplikację znajdziesz na stronie XML-Sitemaps.com (<http://www.xml-sitemaps.com>). Jeżeli Twoja witryna ma nie więcej niż 500 stron, XML-Sitemaps.com dokona analizy wszystkich stron i utworzy bezpłatnie plik *sitemap.xml*. Dla większych witryn trzeba dokupić skrypt w języku PHP, który stanie się integralną częścią projektu. Skrypt ten generuje również mapy w formatach RSS i HTML. Mapy witryn w formacie HTML są szczególnie przydatne użytkownikom, którzy stracili orientację wewnątrz witryny lub chcą zapoznać się z jej zakresem informacyjnym. Wyszukiwarki, które nie obsługują protokołu *sitemap.xml*, będą mogły wykorzystać mapę w formacie HTML do nawigacji po Twojej witrynie.

Po uruchomieniu skryptu nastąpi automatyczne powiadomienie wyszukiwarki Google, że plik został zaktualizowany, aby zminimalizować czas oczekiwania na wizytę robota. Skrypt w języku PHP o nazwie XML-Sitemaps.com sprawdzi także, czy linki są aktualne, i wskaże te, które wymagają poprawienia.

Czytelnikom szukającym darmowego skryptu PHP o podobnych funkcjach polecam sprawdzenie programu phpSitemapNG (<http://enarion.net/google>). Możesz zaprogramować regularne odświeżanie pliku *sitemap.xml* za pomocą skryptu php-SitemapNG i Crona, narzędzia do automatycznego uruchamiania zadań w systemach operacyjnych Unix i Linux. Użytkownicy systemu Windows mogą natomiast wykorzystać funkcję *Zaplanowane zadania w Panelu sterowania*, aby włączyć uruchamianie skryptu PHP.

Niektóre aplikacje desktopowe umożliwiają szybkie utworzenie pliku *sitemap.xml*. Firma RAGE oferuje Google Sitemap Automator dla komputerów Mac (<http://www.ragesw.com/products/googlesitemap.html>). Program analizuje witrynę i generuje plik *sitemap.xml*, zapisuje go na Twoim serwerze i wysyła powiadomienie do wyszukiwarki Google (rysunek 10.2). Jego intuicyjny interfejs bardzo ułatwia cały proces. Chociaż możesz zastosować wersję demonstracyjną programu Google Sitemap Automator do wygenerowania i zapisania pliku *sitemap.xml*, będziesz musiał wykupić licencję, by skorzystać z funkcji powiadamiania wyszukiwarki Google.



Jeżeli pojęcie

Crona w systemach Unix i Linux nie jest Ci znane, zapoznaj się z artykułem Aarona Brazella na stronie <http://www.sitepoint.com/article/introducing-cron>. Natomiast użytkownikom Windows Servera polecam artykuł poświęcony Task Schedulerowi, umieszczony na stronie <http://www.iopus.com/guides/winscheduler.htm>.

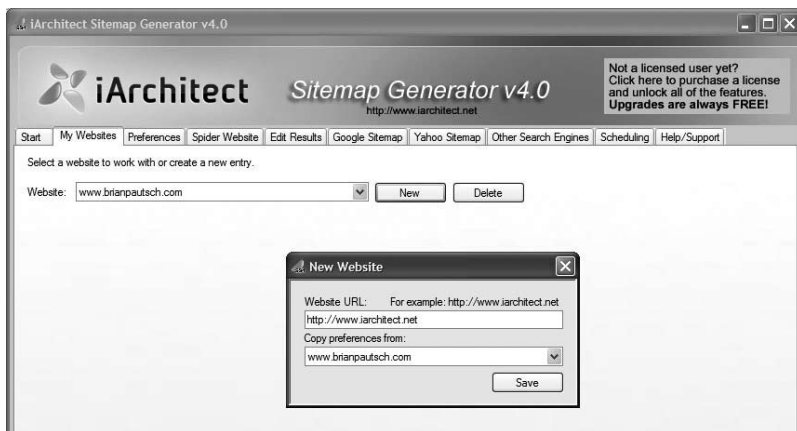
**RYSUNEK 10.2.**

Program Google Sitemap Automator firmy RAGE (<http://www.ragesw.com/products/googlesitemap.html>) szybko utworzy plik *sitemap.xml*, zapisze go na Twoim serwerze i wyśle powiadomienie do wyszukiwarki Google



Firma iArchitect oferuje świetny program pracujący w systemie operacyjnym Windows o nazwie Sitemap Generator (<http://www.iarchitect.net/Products/Sitemap-Generator>), który przedstawiono na rysunku 10.3. Oprócz tworzenia i zapisu mapy witryny oraz funkcji powiadamiania wyszukiwarek Sitemap Generator tworzy także mapy w formacie HTML i udostępnia opcje, które umożliwiają automatyzację czynności związanych z aktualizacją pliku *sitemap.xml*.

**RYSUNEK 10.3.** Sitemap Generator dla Windows oferuje więcej funkcji niż programy konkurencyjne, między innymi tworzenie mapy witryny w formacie HTML i harmonogram aktualizacji pliku *sitemap.xml*



Jeśli nie dysponujesz programem do generowania mapy witryny i automatycznej podmiany pliku *sitemap.xml*, będziesz musiał zrobić to sam. Na ogół plik ten znajduje się w głównym katalogu serwera WWW, na przykład *public\_html* lub *www*. Po utworzeniu nowej wersji pliku będziesz musiał wysłać informację o ścieżce dostępu do wyszukiwarek, aby sprowokować jak najszybsze wizyty ich robotów.

## Powiadomienie wyszukiwarek o pliku *sitemap.xml*

Istnieją trzy sposoby powiadamiania wyszukiwarek o nowej wersji pliku *sitemap.xml*:

- *robots.txt*,
- *ping*,
- zgłoszenie standardowe.

Najlepsze rezultaty osiągniesz, stosując jednocześnie każdy z tych sposobów, aby zwiększyć do maksimum prawdopodobieństwo szybkiej indeksacji.

### **robots.txt**

W rozdziale 3., „Strategie po stronie serwera” opisaliśmy protokół *robots.txt* wykorzystywany przez wyszukiwarki w celu ustalenia, które pliki i katalogi powinny być pominięte przy indeksacji. Ten plik tekstowy znajduje się w głównym katalogu serwera WWW i jest odczytywany automatycznie przez wszystkie roboty wyszukiwarek.

Plik *robots.txt* może być także zastosowany do powiadamiania wszystkich wyszukiwarek o ścieżce dostępu do pliku *sitemap.xml*. Wystarczy tylko umieścić następujący zapis na końcu pliku *robots.txt*:

```
Sitemap: sitemap.xml
```

Zaletą metody opartej na wykorzystaniu pliku *robots.txt* jest to, że każda wyszukiwarka, która interpretuje ten protokół, znajdzie wpis dotyczący mapy witryny. Wyszukiwarki nie są jednak powiadamiane i nie otrzymują zaproszenia do zindeksowania witryny. Jeżeli uruchamiasz nową witrynę, dobrze jest wysłać powiadomienie bezpośrednio do wyszukiwarek w oparciu o jedną z dwóch metod, które teraz przedstawimy.

### **Ping**

*Ping* jest to krótka wiadomość przesyłana między dwoma komputerami. Po opublikowaniu nowej witryny możesz w ten sposób zaprosić wyszukiwarki do jej zindeksowania, wysyłając im informację o położeniu pliku *sitemap.xml*.

Google, Ask i Yahoo! oferują możliwość zastosowania tego mechanizmu za pośrednictwem przeglądarki. Zastąp przykładowy adres URL w poniższych przykładach ścieżką bezwzględną do swojego pliku *sitemap.xml*.

**■ Google:**

<http://www.google.com/webmasters/sitemaps/ping?sitemap=http://przyklad.pl/mapa.xml>

**■ Ask:** <http://submissions.ask.com/ping?sitemap=http://przyklad.pl/mapa.xml>**■ Yahoo!:**

<http://search.yahooapis.com/SiteExplorerService/V1/ping?sitemap=http://przyklad.pl/mapa.xml>

Zastanawia nieobecność na tej liście wyszukiwarki MSN Live Search. Chociaż interpretuje ona protokół *sitemap.xml*, w czasie pisania tej książki nie oferowała jeszcze żadnego narzędzia do rejestracji witryny. Jedynym sposobem jej powiadomienia jest zatem plik *robots.txt*.

Podane adresy URL są także bardzo przydatne do budowania systemu zarządzania treścią lub aplikacji internetowej, która w sposób ciągły zgłasza nowe strony do wyszukiwarek. Twoja aplikacja mogłaby aktualizować plik *sitemap.xml*, a następnie zastosować mechanizm pingowania.



W celu zgłoszenia pliku *sitemap.xml* do wyszukiwarek za pomocą usługi *ping*, utwórz nowy folder w grupie zakładek ulubionych na przeglądarce, a następnie dodaj adresy stron z usługą *ping* pliku *sitemap.xml*. Za każdym razem, gdy zaktualizujesz swój plik *sitemap.xml*, po prostu kliknij te zakładki, a wiadomość o zmianie zawartości pliku zostanie wysłana do odpowiedniej wyszukiwarki.

## Rejestracja manualna

Możesz również poinformować wyszukiwarki o nowym pliku *sitemap.xml*, wchodząc na ich adresy internetowe i dokonując ręcznego zgłoszenia adresu URL swojej mapy witryny. Niestety, tylko wyszukiwarki Google i Yahoo! posiadają aktualnie narzędzia do rejestracji manualnej *sitemap.xml*. Jest to niewątpliwie najbardziej uciążliwy sposób zgłaszania mapy witryny. Jednak warto z niego skorzystać w przypadku uruchomienia nowej witryny.

Pod koniec tego rozdziału dowiesz się, że przy ręcznym powiadomieniu wyszukiwarek Google i Yahoo! o nowej witrynie mapa witryny zostanie sprawdzona pod względem składniowym i w przypadku wystąpienia błędów zostaniesz o tym poinformowany. Otrzymasz także informację, kiedy po raz ostatni nastąpił odczyt tego pliku, dzięki czemu zorientujesz się, czy najnowsza treść znajduje się już w indeksie wyszukiwarki.

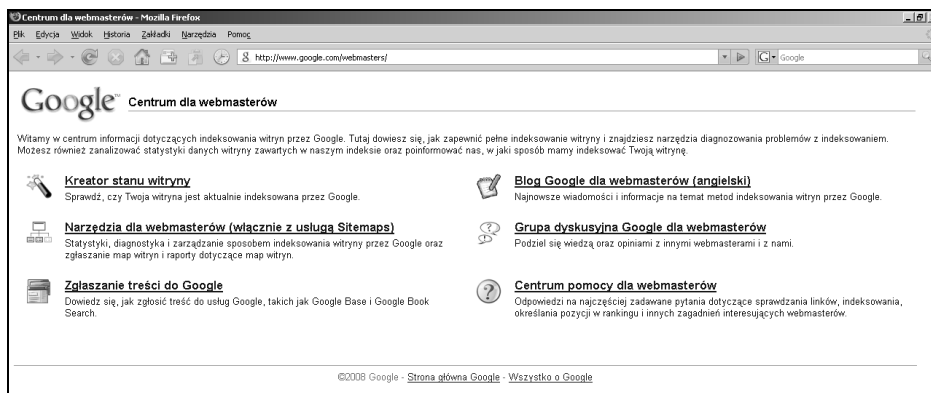


## UWAGA

Jeżeli potrzebujesz wygenerować mapę witryny dla bloga WordPress, przeczytaj sekcję rozdziału 5. zatytułowaną „Automatyczne generowanie XML-owej mapy witryny”.

## Google — Centrum dla webmasterów

Google oferuje szeroki wachlarz darmowych programów statystycznych, diagnostycznych i zarządzających w swoim Centrum dla webmasterów (<http://www.google.com/webmasters>), co widać na rysunku 10.4. Z tego miejsca możesz wystać swój adres URL do Google i sprawdzić stan witryny, by przekonać się, czy została już zindeksowana. Jednak funkcjonalności Centrum dla webmasterów są bardzo zróżnicowane.



**RYСУNEK 10.4.** Centrum dla webmasterów Google zawiera wiele darmowych i użytecznych programów statystycznych, diagnostycznych i zarządzających

Główny zestaw darmowych programów nazywa się Narzędzia dla webmasterów (<https://www.google.com/webmasters/tools/?hl=pl>). Zanim zaczniesz posługiwać się tymi narzędziami, zarejestruj się lub zaloguj na swoje konto. Możesz zarządzać mapami witryn i przeglądać statystyki dotyczące wielu witryn z poziomu jednego konta, dzięki czemu możesz utrzymywać zakładki dla witryn różnych klientów.

Po zalogowaniu się zostaniesz przeniesiony do panelu sterowania, gdzie możesz dodać adres URL swojej witryny, aby przeglądać dane na jej temat i ustawiać niektóre preferencje. Zostaniesz poproszony o sprawdzenie, czy masz prawo zarządzania witryną na jeden z dwóch sposobów. Możesz albo

dodać specjalny znacznik meta z unikalną wartością na swojej stronie głównej, albo umieścić w głównym katalogu na serwerze WWW specjalny plik przygotowany przez Google. Po wykonaniu jednego z tych zadań wystarczy kliknąć przycisk *Verify* (*Weryfikuj*), a robot wyszukiwarki Google natychmiast odwiedzi Twoją witrynę i potwierdzi obecność znacznika meta albo pliku. Po udowodnieniu, że jesteś webmasterem tej witryny, możesz rozpocząć posługiwanie się narzędziami Google.

## Narzędzia webmastera

Pakiet narzędzi Google dla webmasterów składa się z pięciu sekcji:

- *Diagnostics* (*Diagnostyka*),
- *Statistics* (*Statystyka*),
- *Links* (*Linki*),
- *Site maps* (*Mapy witryn*),
- *Tools* (*Narzędzia*).

Po zalogowaniu się na stronie narzędzi Google dla webmasterów i kliknięciu wybranego adresu URL witryny, którą chcesz zarządzać, wyświetla się strona z podstawowymi informacjami (rysunek 10.5).

The screenshot shows the Google Webmaster Tools interface. On the left is a sidebar with navigation links: Overview, Diagnostics, Statistics, Links, Sitemaps, and Tools. The main content area is titled 'Overview' and shows the website URL 'http://aaronwalter.com/'.

**Indexing | Top search queries »**

Home page crawl:	✓ Googlebot last successfully accessed your home page on Jul 15, 2007.
Index status:	✓ Pages from your site are included in Google's index. See <a href="#">Index stats</a> . [?]

**Web crawl errors**

Error type	Count	Action
HTTP errors	✓ 0	–
Not found	⚠ 1	<a href="#">Details »</a>
URLs not followed	✓ 0	–
URLs restricted by robots.txt	⚠ 1	<a href="#">Details »</a>
URLs timed out	✓ 0	–
Unreachable URLs	⚠ 4	<a href="#">Details »</a>
<b>Total:</b>	<b>6</b>	

**RYСУNEK 10.5.** Strona z podstawowymi informacjami na temat witryny w pakiecie narzędzi Google dla webmasterów umożliwia szybki dostęp do statystyk i innych informacji zawartych w podsekcjach

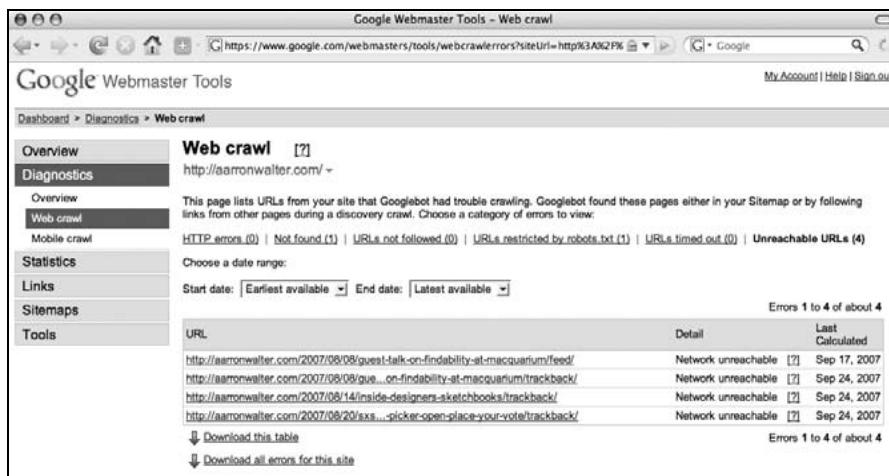
Na pierwszej stronie znajdują się informacje zasadnicze, na przykład strony z błędem 404, strony, których nie udało się załadować, błędy HTTP lub strony z zablokowanym dostępem do indeksacji ze względu na odpowiedni zapis w pliku *robots.txt*. W przypadku wykrycia nieprawidłowości pojawiają się linki do informacji szczegółowych.

Na tej stronie znajdziesz również datę ostatniej indeksacji Twojej strony głównej oraz wskazanie, czy treść Twojej witryny znajduje się w indeksie Google. Te informacje ułatwiają wykrycie poważniejszych problemów. Przyjrzyjmy się dokładniej niektórym funkcjom z pakietu narzędzi Google dla webmasterów.

## Diagnostics (Diagnostyka)

W sekcji *Diagnostics (Diagnostyka)* znajdziesz bardziej szczegółowe informacje o problemach napotkanych przez Google przy indeksowaniu Twojej witryny (rysunek 10.6). Istnieje sześć rodzajów błędów i tematów rejestrowanych przez Google:

- **Błędy HTTP:** błędy konfiguracji serwera, zabronione katalogi itp.
- **Not found:** błędy 404.
- **Adresy URL stron, które są przez Ciebie oznaczone, aby nie były indeksowane.**



The screenshot shows the Google Webmaster Tools interface for the 'Web crawl' section. The page title is 'Web crawl [7]'. Below the title, it says 'http://aaronwalter.com/'. The main content area lists URLs that Googlebot had trouble crawling. The list includes four URLs, all marked as 'Network unreachable'. The table has columns for 'URL', 'Detail', and 'Last Calculated'. Below the table, there are links to 'Download this table' and 'Download all errors for this site'.

URL	Detail	Last Calculated
http://aaronwalter.com/2007/08/08/guest-talk-on-findability-at-macquarium/feed/	Network unreachable [7]	Sep 17, 2007
http://aaronwalter.com/2007/08/08/guest-talk-on-findability-at-macquarium/trackback/	Network unreachable [7]	Sep 24, 2007
http://aaronwalter.com/2007/08/14/inside-designers-sketchbooks/trackback/	Network unreachable [7]	Sep 24, 2007
http://aaronwalter.com/2007/08/20/xxx...picker-open-place-your-vote/trackback/	Network unreachable [7]	Sep 24, 2007

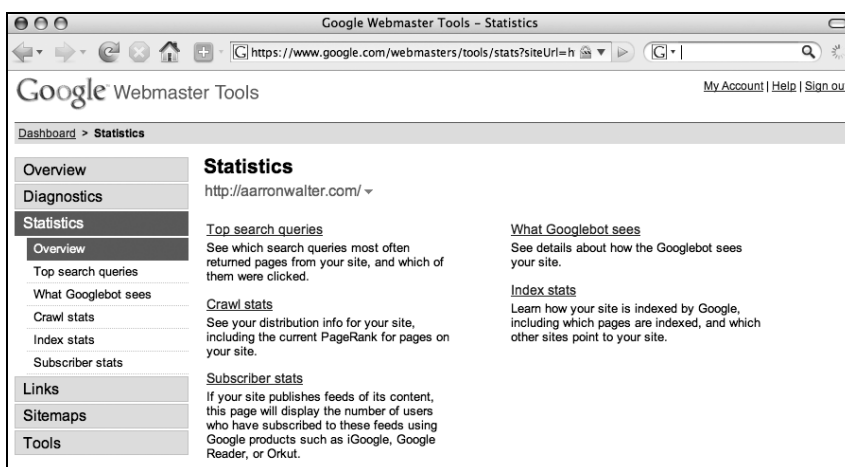
**RYSUNEK 10.6.** Możesz dokładnie zlokalizować problemy napotkane przez robota wyszukiwarki Google w czasie indeksowania witryny w sekcji *Diagnostics (Diagnostyka)* pakietu narzędzi dla webmasterów

- **Adresy URL zastrzeżone w pliku *robots.txt*.**
- **Adresy URL stron, których nie udało się zindeksować z powodu wolnego połączenia sieciowego, błędów w kodzie źródłowym itp.**
- **Nieosiągalne adresy URL:** strony wymienione w pliku *sitemap.xml*, które nie istnieją.

W sekcji *Diagnostics (Diagnostyka)* znajdziesz podsekcję *Web crawl (Indeksowanie Google Mobile)*. Jeżeli Twoja witryna ma format specyficzny dla urządzeń przenośnych, na przykład CHTML lub WML, Google zindeksuje jej treść. Na tej stronie znajdziesz informacje dotyczące tego tematu.

## Statistics (Statystyka)

*Statistics (Statystyka)* jest bodaj najciekawszą grupą narzędzi dla webmasterów (rysunek 10.7). Zawiera szereg podsekcji z interesującymi danymi na temat słów kluczowych wykorzystywanych przez internautów, obserwacji dokonanych przez robot wyszukiwarki Google w czasie indeksowania Twojej witryny, statystyki związane z indeksowaniem oraz subskrypcjami na kanały RSS. Jest to spora porcja informacji, dzięki którym możesz zorientować się, w jaki sposób wyszukiwarka postrzega Twoją witrynę ze swojej perspektywy.



**RYСУNEK 10.7.** Sekcja Statystyka narzędzi Google dla webmasterów zawiera dużo szczegółowych informacji na temat Twojej witryny. Między innymi możesz dowiedzieć się, jaka treść na Twoich stronach jest postrzegana przez Google jako ważna oraz które słowa i frazy kluczowe generują ruch na Twojej witrynie

## Top search queries (Najczęstsze zapytania)

*Top search queries (Najczęstsze zapytania)* to szczególnie ciekawy obszar do zbadania. Nie tylko możesz się dowiedzieć, jakie słowa lub frazy kluczowe generują ruch na Twojej witrynie, ale możesz także zobaczyć swój ranking dla każdego zapytania. Chociaż jest interesujące, które słowa kluczowe z Twojej listy słów kluczowych kierują internautów na Twoją witrynę, jeszcze ciekawsze jest poznanie innych słów kluczowych, których wcześniej nie brałeś pod uwagę. Prawdopodobnie poznasz kilka dziwacznych sformułowań, które internauci wpisali, a następnie kliknęli link do Twojej witryny na liście wyników wyszukiwania! Przyjrzyj się uważnie tym informacjom, gdyż mogą wskazać Ci drogę do dalszych usprawnień.

Dane mogą być filtrowane w oparciu o kryteria geograficzne oraz sposób wyszukiwania, w tym wyszukiwanie blogów, wyszukiwanie graficzne, wyszukiwanie Google Mobile lub zwykłe wyszukiwanie.

## What Googlebot sees (Informacje rejestrowane przez robota Googlebot)

W sekcji *What Googlebot sees (Informacje rejestrowane przez robota Googlebot)* możesz zobaczyć ranking słów i fraz kluczowych zawartych w treści Twojej witryny. Jeżeli nie widzisz swoich najważniejszych słów kluczowych w górnej części tej listy, oznacza to, że powinieneś przeczytać jeszcze raz fragment rozdziału 2., „Strategie tworzenia kodu HTML”, dotyczący strategii doboru słów kluczowych, oraz rozdziału 4., „Tworzenie treści przyciągającej internautów”, aby poprawić ich rozmieszczenie na swoich stronach.

Jeszcze bardziej intrygującą informacją w tej sekcji jest lista słów kluczowych, które zastosowali internauci w swoich linkach zaadresowanych do Twojej witryny. Krąży pogłoska, że jest to jeden z najważniejszych czynników uwzględnianych przez Google przy ustalaniu rankingu PageRank witryny. Niestety, brakuje w tym miejscu linków do witryn, na których znajdują się Twoje linki przychodzące, ale możesz je znaleźć w następnej sekcji narzędzi dla webmastera. W dalszej części tego rozdziału zajmiemy się innymi metodami uzyskiwania informacji na temat linków przychodzących.

Im większa korelacja pomiędzy głównymi słowami kluczowymi Twojej witryny a słowami kluczowymi zawartymi w etykietach linków przychodzących, tym większe są szanse na uzyskanie czołowych miejsc w listach rankingowych. Łatwo jest kontrolować tekst na swojej witrynie, ale nie możesz wpływać

na witryny innych ludzi. Jeżeli znasz niektórych właścicieli witryn, możesz zwrócić się do nich z prośbą o zmianę etykiet w linkach zaadresowanych do Twojej witryny.

Na tej samej stronie znajdziesz także informacje o trudnościach, jakie robot wyszukiwarki napotkał przy indeksowaniu treści Twojej witryny, uwzględniając formaty HTML, XML, PDF, Flash, tekstowy i inne.

### **Crawl stats (Statystyki indeksowania)**

Sekcja *Crawl stats (Statystyki indeksowania)* zawiera mniej informacji niż pozostałe, ale dane są równie interesujące. Znajdziesz tutaj średni ranking PageRank swoich stron oraz adres URL strony o najwyższym rankingu za okres ostatnich trzech miesięcy. Wiedząc, która strona ma najwyższy ranking, możesz określić, jaka jest tego przyczyna. Zazwyczaj powodem najwyższego rankingu jest ilość i jakość linków przychodzących. Jak już objaśnialiśmy w rozdziale 4., wartościowa treść jest zazwyczaj źródłem dużej liczby linków przychodzących. Jest to kolejna ważna informacja związana z ustaleniem, jaka treść jest najbardziej wartościowa dla Twojego audytorium.

### **Index stats (Statystyki indeksowania)<sup>1</sup>**

Sekcja *Index stats (Statystyki indeksowania)* zawiera kilka linków umożliwiających tworzenie zapytań przy użyciu operatorów zaawansowanych. Są to przykładowe wyniki pokazujące sposób zaindeksowania Twojej witryny. Te same rezultaty zobaczysz bezpośrednio w oknie wyszukiwarki Google, wpisując po prostu adres swojej witryny poprzedzony odpowiednim operatorem. W ten sposób dowiesz się nieco, w jaki sposób wyszukiwarka Google indeksuje Twoją witrynę:

- `site:przyklad.pl` — zindeksowane strony witryny,
- `allinurl:przyklad.pl` — strony odwołujące się do Twojej witryny,
- `link:przyklad.pl` — strony linkujące do Twojej witryny,
- `cache:przyklad.pl` — aktualny stan bufora,
- `info:przyklad.pl` — informacja, jaką Google posiada na temat Twojej witryny,
- `related:przyklad.pl` — strony podobne do Twojej witryny.

---

<sup>1</sup> W polskiej wersji serwisu zarówno *Crawl stats*, jak i *Index stats* tłumaczone są jako *Statystyki indeksowania* — *przyp. red.*

Małe ostrzeżenie: operator `link` nie gwarantuje kompletnej listy linków przychodzących. Niech więc Cię nie przerazi zaskakująco krótka lista wyników wyszukiwania po użyciu tego operatora. Narzędzie *Links (Linki)*, którym zajmamy się wkrótce, daje pełną listę bardzo przydatnych informacji. Również operator `allinurl` pozwala uzyskać obszerniejszą listę linków przychodzących.

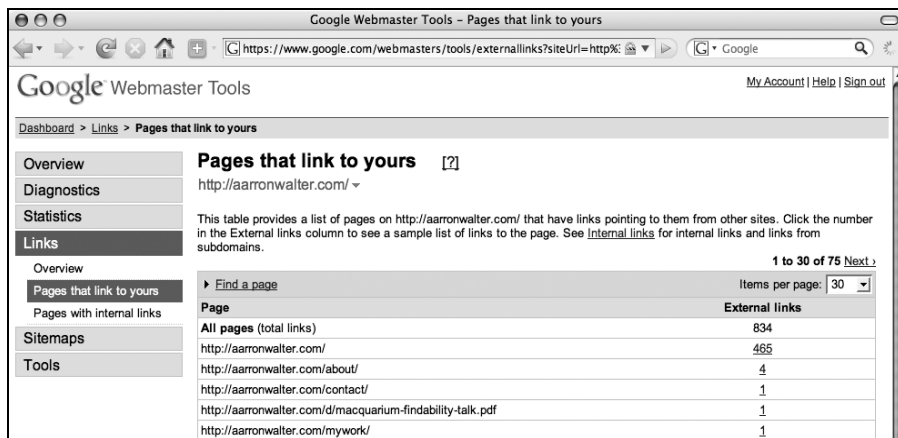
## Subscriber Stats (Statystyki subskrybentów)

Informacja, którą możesz uzyskać w sekcji *Subscriber stats (Statystyki subskrybentów)*, często nie obejmuje pełnej listy osób korzystających z kanałów RSS Twojej witryny. Wyszukiwarka Google śledzi jedynie subskrypcje zrealizowane za pomocą własnych agregatorów RSS. Jeżeli użytkownicy korzystają z Twoich kanałów RSS za pośrednictwem Bloglines (<http://bloglines.com>), Netvibes (<http://netvibes.com>) lub innego agregatora RSS, nie zobaczysz w tym miejscu ich danych.

Przyjmij zatem tę informację z przymrużeniem oka. W rozdziale 13. zatytułowanym „Analiza ruchu internautów” zajmemy się statystyką subskrypcyjną FeedBurnera (<http://feedburner.com>), która zawiera kompletny przegląd subskrybentów Twoich kanałów RSS.

## Links (Linki)

Sekcja *Links (Linki)* zawiera bardzo ciekawe dane na temat witryn zawierających linki do Twojej witryny oraz informacje o tym, które strony na Twojej witrynie posiadają najwięcej linków przychodzących (rysunek 10.8).



The screenshot shows the Google Webmaster Tools interface. The main heading is 'Pages that link to yours' with a question mark icon. Below it is the URL 'http://aaronwalter.com/'. A descriptive text states: 'This table provides a list of pages on http://aaronwalter.com/ that have links pointing to them from other sites. Click the number in the External links column to see a sample list of links to the page. See [Internal links](#) for internal links and links from subdomains.' The table has two columns: 'Page' and 'External links'. It shows 1 to 30 of 75 items. The data rows are as follows:

Page	External links
All pages (total links)	834
http://aaronwalter.com/	465
http://aaronwalter.com/about/	4
http://aaronwalter.com/contact/	1
http://aaronwalter.com/d/macquarium-findability-talk.pdf	1
http://aaronwalter.com/mywork/	1

**RYSUNEK 10.8.** Korzystając z narzędzia *Links (Linki)*, możesz ustalić, które strony na Twojej witrynie posiadają najwięcej linków przychodzących oraz które witryny zawierają te linki

Jest oczywiste, że strony z największą liczbą linków przychodzących zawierają treść, która najbardziej interesuje Twoje audytorium. Przyjrzyj się dokładnie tym danym, gdyż znajomość stopnia popularności poszczególnych stron ułatwi Ci taką rozbudowę witryny, by uzyskać jeszcze większą liczbę odwiedzin.

Sekcja *Links (Linki)* udostępnia również informacje na temat linków wewnętrznych na Twojej witrynie. Jest to przydatne dla oceny poziomu połączeń między stronami witryny. Dobrze rozbudowana sieć połączeń między stronami sprzyja dłuższym wizytom internautów, a tym samym zwiększa prawdopodobieństwo osiągnięcia celów biznesowych, takich jak dokonanie zakupu lub zapisanie się na listę dyskusyjną.

## Sitemaps (Mapy witryn)

W poprzedniej części tego rozdziału dowiedziałeś się, w jaki sposób utworzyć plik *sitemap.xml* i powiadomić o tym fakcie wyszukiwarki. Dzięki tej mapie wyszukiwarka może lepiej zindeksować treść Twojej witryny.

W sekcji *Sitemaps (Mapy witryn)* pakietu narzędzi Google dla webmasterów możesz wysłać adres URL swojego pliku *sitemap.xml* i monitorować jego status (rysunek 10.9). Jeżeli uruchamiasz nową witrynę, dobrze jest wysłać do wyszukiwarki Google adres pliku z mapą witryny i sprawdzić, czy nie zawiera błędów składniowych.

Google Webmaster Tools - Sitemaps

Dashboard > Sitemaps

**Sitemaps**  
http://aaronwalter.com/

Add a Sitemap

View: All Sitemaps OK

Sitemaps 1 to 1 of 1

Sitemap	Type	Submitted	Last Downloaded	Sitemap Status	URLs submitted
<input type="checkbox"/> sitemap.xml	SITEMAP	Web	Nov 6, 2006	Sep 27, 2007	OK 128

Delete Selected Resubmit Selected Download this table

Download Sitemap details for all of your sites as a .csv file

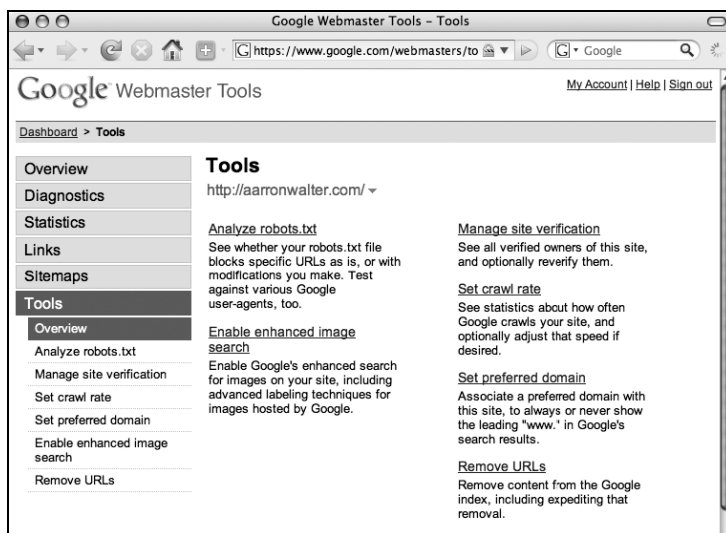
**RYСУNEK 10.9.** *Sitemaps (Mapy witryn)* to narzędzie umożliwiające wysłanie adresu URL Twojego pliku z mapą witryny i monitorowanie jego statusu. Wyszukiwarka Google powiadomi Cię o ewentualnych problemach napotkanych przy wczytywaniu tego pliku



Oprócz informowania o błędach składniowych w pliku z mapą witryny narzędzie Google udostępni Ci również datę ostatniego odczytu pliku przez robota wyszukiwarki oraz liczbę adresów URL zawartych w tym pliku. Jeżeli dysponujesz wersją swojej witryny dla urządzeń mobilnych, możesz również dostarczyć odrębny plik *sitemap.xml* dla tej wersji.

## Tools (Narzędzia)

W sekcji *Tools (Narzędzia)* (rysunek 10.10) możesz dokonać analizy pliku *robots.txt*, zarządzać wybranym procesem weryfikacji witryny, ustawić częstotliwość indeksowania przez Google, zdefiniować preferowaną domenę, czyli wskazać, którego adresu URL należy użyć w celu indeksowania witryny, i usunąć niektóre strony z indeksu Google.



**RYСУNEK 10.10.**  
*Sekcja Tools (Narzędzia) oferuje szereg użytecznych preferencji i funkcji*

### Analyze robots.txt (Analizuj plik robots.txt)

Możesz zidentyfikować błędy składniowe w pliku *robots.txt*, korzystając z narzędzia do jego analizy. Jeżeli Googlebot odwiedził już Twoją witrynę i znalazł plik *robots.txt*, na tej stronie zobaczysz jego treść. To narzędzie umożliwi również wpisywanie adresów URL stron wyłączonych przez Ciebie z indeksacji, aby upewnić się, czy plik *robots.txt* skutecznie zapobiegnie ich zindeksowaniu.

### Manage site verification (Zarządzaj weryfikacją własności witryny)

Jeżeli z pewnych względów potrzebujesz sprawdzić zawartość znacznika meta lub nazwę pliku używanego przez Google do weryfikacji właściciela

witryny, znajdziesz te informacje w sekcji zarządzania weryfikacją witryny. Niestety, nie ma możliwości zmiany raz wybranej metody weryfikacji.

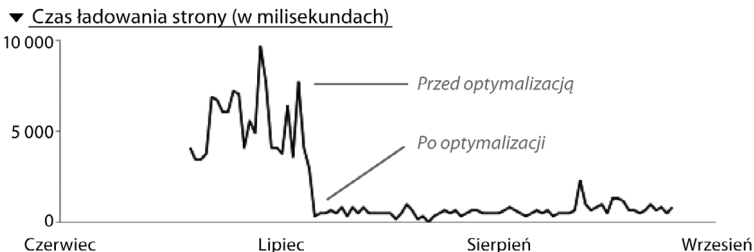
### Set crawl rate (Ustaw szybkość przemierzania)

W tej sekcji możesz utrzymywać zakładki z informacją o częstotliwości i szybkości indeksowania Twojej witryny przez Google. Możesz dowiedzieć się, jaka jest średnia liczba stron indeksowanych dziennie, średnia liczba kilobajtów pobieranych danych i czas ładowania.

Po wykonaniu optymalizacji witryny zgodnie ze wskazówkami zawartymi w rozdziale 3. zatytułowanym „Strategie po stronie serwera” możesz zauważyć wyraźne zmniejszenie czasu ładowania danych przy indeksowaniu witryny. Na rysunku 10.11 przedstawiam ogromną zmianę tego czasu, w jakim Googlebot indeksuje moją witrynę, przed i po dokonaniu optymalizacji.

#### RYSUNEK 10.11.

Po zoptymalizowaniu swojej witryny zauważyłem wzrost szybkości indeksowania przez Google o około 90 procent



Jeżeli Twoja witryna zdobędzie nagle dużą popularność, a szerokość pasma stanie się istotnym problemem, w tej sekcji możesz zmniejszyć częstotliwość indeksowania witryny.

### Set preferred domain (Ustawianie preferowanej domeny)

W rozdziale 3. dowiedziałeś się, że kiedy robot wyszukiwarki Google indeksuje witrynę oraz traktuje adresy zaczynające się na WWW i bez tego przedrostka jako całkowicie różne witryny. Ponieważ adres URL <http://www.mojawitryna.pl> może występować w linkach przychodzących w postaci <http://mojawitryna.pl>, Google może nadać jej niższy ranking PageRank, chociaż obydwa adresy odnoszą się do tej samej witryny.

W sekcji *Set preferred domain (Ustawianie preferowanej domeny)* możesz poinformować Google o wybranym formacie URL, aby nie tracić na rankingu PageRank. W rozdziale 3. znajduje się jeszcze inne rozwiązanie tego problemu, oparte na module `mod_rewrite` serwera Apache i remapowaniu wszystkich żądań do jednego skonsolidowanego formatu. Dobrym pomysłem jest zastosowanie obydwu metod.

## Enable enhanced image search (Rozszerzone wyszukiwanie grafiki)

Jeżeli Twoje pliki graficzne są wykrywane za pośrednictwem wyszukiwarki grafiki Google, możesz włączyć rozszerzone wyszukiwanie plików graficznych. Obrazy mogą być lokalizowane na podstawie nazwy pliku albo zawartości znacznika alt, ale Google oferuje jeszcze bardziej pomysłowe podejście, aby zwiększyć trafność wyników wyszukiwania grafiki.

Google Image Labeler (<http://images.google.com/imagelabeler>) to projekt identyfikacji elementów graficznych polegający na przedstawieniu ochotnikom przypadkowej serii obrazów zebranych z witryn, które mają włączoną opcję rozszerzonego wyszukiwania grafiki. Ochotnicy tworzą opisowe metadane, które są następnie wykorzystywane przez Google. Ponieważ wyszukiwanie grafiki staje się coraz popularniejsze, włączenie tej opcji może przyczynić się do zwiększenia liczby odwiedzin Twojej witryny.

## Remove URLs (Usuń adresy URL)

Jeśli w indeksie Google znajdują się niepożądane fragmenty treści Twojej witryny, możesz zażądać ich usunięcia, korzystając z sekcji *Remove URLs (Usuń adresy URL)*. Oczywiście możesz zablokować indeksowanie w pliku *robots.txt* (zajrzyj do rozdziału 3., „Strategie po stronie serwera”) lub zastosować znacznik meta *noindex* (zajrzyj do rozdziału 2., „Strategie tworzenia kodu HTML”). Jednak żadna z tych metod nie pozwala na natychmiastowe usunięcie strony z indeksu Google. Trzeba poczekać do najbliższej reindeksacji witryny. Natomiast sekcja *Remove URLs (Usuń adresy URL)* daje taką możliwość.

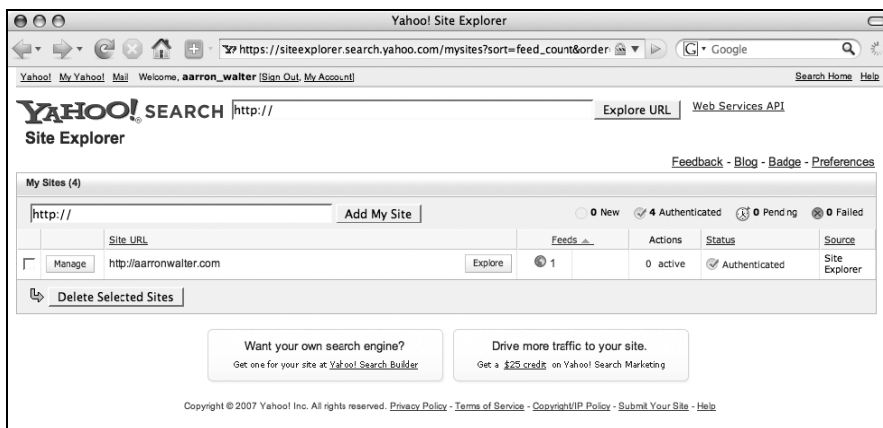
# Informacje o witrynie — Yahoo! Site Explorer

Site Explorer jest darmowym zestawem narzędzi, które odślaniają kulisy indeksowania Twojej witryny przez Yahoo! (<https://siteexplorer.search.yahoo.com>) oraz informują, kto dolinkował się do Twoich stron. Site Explorer umożliwia także zgłoszenie pliku *sitemap.xml* do Yahoo! po uruchomieniu nowej witryny.

Aby skorzystać z funkcji Site Explorera, będziesz musiał utworzyć własne konto na witrynie Yahoo!, o ile jeszcze go nie posiadasz, a następnie uwierzytelnić swoją witrynę według procedury podobnej do procesu

uwierzelniania w pakiecie narzędzi Google dla webmasterów. Możesz udowodnić, że witryna należy do Ciebie, dodając specjalny znacznik meta na swojej stronie głównej lub zapisując plik dostarczony przez Yahoo! w głównym folderze serwera WWW. Po wykonaniu jednego z tych zadań możesz zasygnalizować wyszukiwarce Yahoo! gotowość do sprawdzenia, że jesteś właścicielem witryny.

Na poziomie głównego obszaru panelu sterowania o nazwie *My Sites* (moje witryny) możesz dodawać adresy URL stron do eksploracji i utrzymywać zakładki dotyczące stanu uwierzelnienia Twojej witryny (rysunek 10.12). Możesz także poinformować Yahoo! o położeniu pliku *sitemap.xml* w sekcji *Feeds* (kanały) Site Explorera po kliknięciu przycisku *Manage* (zarządzaj), który znajduje się obok pola tekstowego z adresem URL Twojej strony.

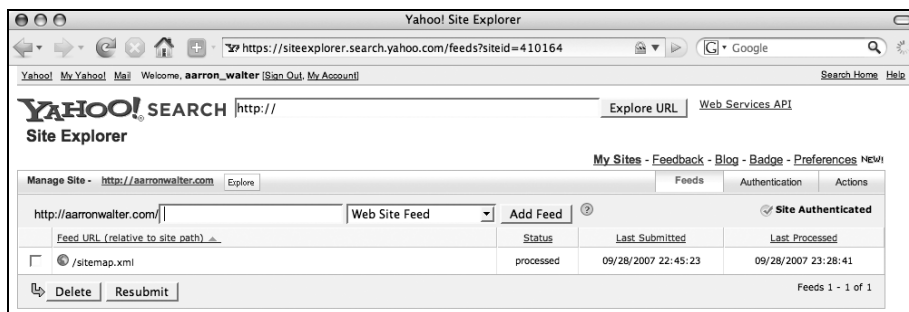


Za zgodą Yahoo! Inc. © 2007. YAHOO! i logo YAHOO! są znakami zastrzeżonymi przez firmę Yahoo! Inc.

**RYСУNEK 10.12.** Możesz zarządzać i eksplorować dowolną liczbę witryn w programie Yahoo! Site Explorer

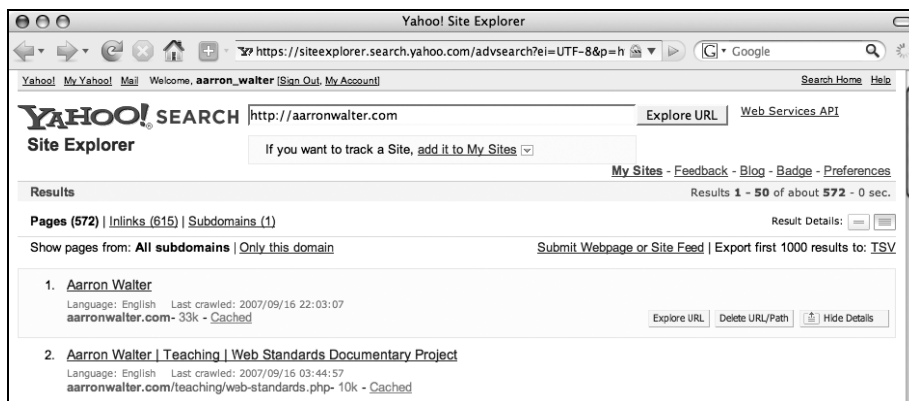
W sekcji *Feeds* wpisz ścieżkę dostępu do pliku z mapą witryny na Twoim serwerze, wtedy Yahoo! odczyta plik, a następnie zindeksuje Twoją witrynę (rysunek 10.13). Możesz także zdefiniować mapę witryny dla wersji mobilnej, jeżeli taką posiadasz. W tym samym oknie możesz odczytać datę i czas ostatniego wczytania i przetworzenia przez Yahoo! Twojego pliku z mapą witryny.

Sercem Site Explorera jest narzędzie *Explore* (eksploruj), do którego wchodzimy ze strony *My Sites*, klikając przycisk *Explore* z prawej strony Twojego adresu URL. Site Explorer listuje wszystkie strony poindeksowane przez Yahoo! na Twojej witrynie (rysunek 10.14). Jest to szczególnie



Za zgodą Yahoo! Inc. © 2007. YAHOO! i logo YAHOO! są znakami zastrzeżonymi przez firmę Yahoo! Inc.

**RYSUNEK 10.13.** Możesz powiadomić Yahoo! o położeniu swoich plików `sitemap.xml` dla wersji standardowej i mobilnej w sekcji *Feeds Site Explorera*



Za zgodą Yahoo! Inc. © 2007. YAHOO! i logo YAHOO! są znakami zastrzeżonymi przez firmę Yahoo! Inc.

**RYSUNEK 10.14.** Site Explorer podaje listę wszystkich stron Twojej witryny zindeksowanych przez Yahoo!. Jest także dołączona kompletna lista witryn zawierających linki do Twojej witryny

przydatne w przypadku nowej witryny, gdyż można sprawdzić, co znajduje się w indeksie Yahoo!. Możesz także usunąć dowolną stronę z indeksu Yahoo!, klikając przycisk *Delete URL/Path* (usuń stronę/ścieżkę), który znajduje się powyżej zapisu na temat pojedynczej strony.

Po kliknięciu w link *Inlinks* (linki przychodzące) zobaczysz pełną listę witryn zawierających link do Twojej witryny. Jest to ważna informacja, ponieważ im więcej posiadasz linków na renomowanych witrynach, tym

wyższy jest Twój ranking. Jeżeli poprosiłeś przyjaciół, kolegów i znajomych o utworzenie linka do Twojej strony, w tej sekcji programu Site Explorer sprawdzisz, kiedy wyszukiwarka Yahoo! zauważyła nowe linki przychodzące.

Niektóre witryny próbują w sposób nieuczciwy zdobywać linki przychodzące poprzez komentarze na blogach witryn o wysokim rankingu i wykorzystanie protokołu trackback. Jak już objaśniałem w rozdziale 5., „Budowanie wyszukiwalnego bloga”, trackback to automatycznie generowany wpis wyświetlany pod artykułem, gdy inny blog zawiera link do tej strony. Trackback zazwyczaj zawiera linka do witryny, która go wygenerowała.

Niektórzy blogerzy publikują bezsensowne i niezwiązane z danym tematem wpisy, dołączając linki do Twojej witryny, aby uzyskać w sposób nieuczciwy linki przychodzące i wzrost rankingu w wyszukiwarkach. Jeżeli zauważysz takie spammerskie linki przychodzące w czasie przeglądania danych w programie Site Explorer, możesz to zgłosić do Yahoo!, zaznaczając właściwy zapis i klikając przycisk *Report Spam* (zgłoś spam).

Dzięki programowi Site Explorer możesz zdobyć wiele cennych informacji. Warto regularnie monitorować zindeksowane strony i linki przychodzące, aby zapewnić widoczność swojej witryny dla internautów korzystających z wyszukiwarki Yahoo!.