

Rozdział

Internetowy system informacji o seminariach

Ireneusz SZCZEŚNIAK
Instytut Informatyki Teoretycznej i Stosowanej PAN
iszczesniak@iitis.gliwice.pl

Streszczenie

W rozdziale prezentujemy niezależny do platformy sprzętowej i programowej system do zautomatyzowanego generowania statycznej strony internetowej, która udostępnia informacje o seminariach. Opisujemy ogólną metodę zautomatyzowanego tworzenia stron internetowych, które wymagają umiarkowanie częstych uaktualnień. Na końcu rozdziału został krótko opisany sposób użycia systemu.

1. Wprowadzenie

Wiele organizacji posiada stronę internetową ze szczegółami dotyczącymi zaplanowanych seminariów, dalej nazywaną stroną seminaryjną. Takie strony są zwykle aktualizowane każdego tygodnia, aby np. dodać ogłoszenie o seminarium. Najczęściej aktualizacja tych informacji wykonywana jest poprzez edycję strony, co jest udręką administratora strony. Prowadzenie pewnych stron seminaryjnych jest zautomatyzowane, np. [1,2]. Autor tego rozdziału znalazł tylko jeden pakiet oprogramowania do prowadzenia stron seminaryjnych, przy którego użyciu jest prowadzona strona [2]. Ten pakiet nie spełnia jednak wielu warunków, głównie warunku bezpieczeństwa.

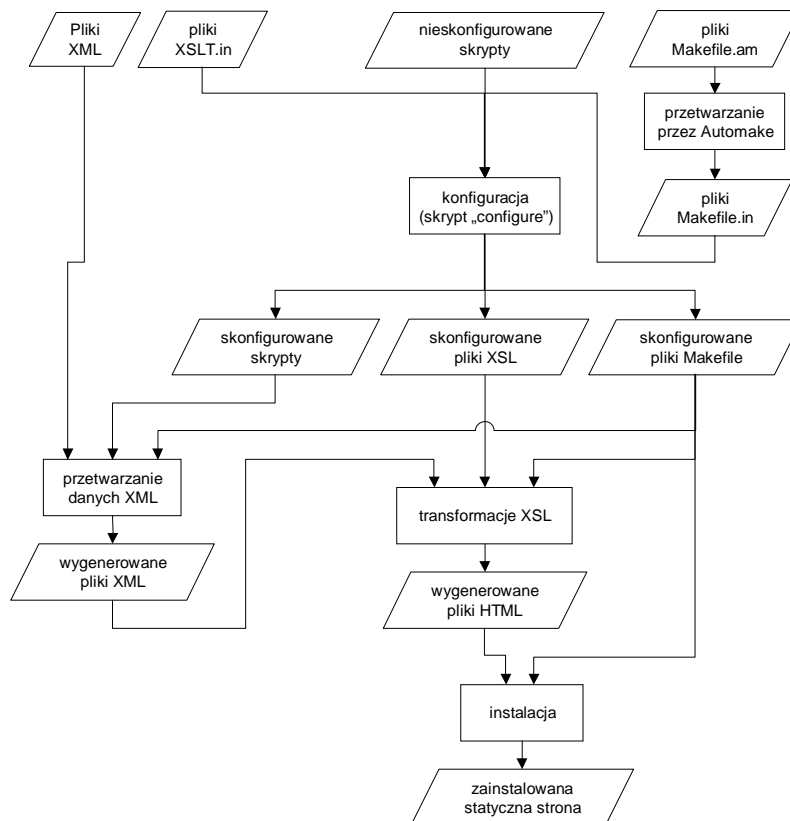
Dlatego został zaimplementowany pakiet oprogramowania do zautomatyzowanego prowadzenia stron seminaryjnych. Metoda używana w tym oprogramowaniu jest ogólna i może zostać zastosowana do prowadzenia innych stron internetowych, niekoniecznie seminaryjnych, które wymagają częstych aktualizacji. Prezentowana metoda jest oryginalna i raczej niekonwencjonalna.

Duże strony internetowe są tworzone dynamiczne (tzn. serwer uruchamia proces, który tworzy dokument w momencie otrzymania żądania od użytkownika) przez skrypty języka PHP, Python albo Perl. Takie strony są jednak podatne na włamania z powodu błędów i niedoskonałości w skryptach. Nasz pakiet został zaprojektowany z myślą o zapewnieniu największego bezpieczeństwa przed włamaniami do serwera.

Prezentowany pakiet jest oparty wyłącznie na przenośnych (tzn. niezależnych od platformy sprzętowej i programowej) narzędziach (np. Python) i dlatego sam jest przenośny. Pakiet można używać w systemach typu UNIX, takich jak Linux, Mac OS X i inne systemy tego typu (np. AIX, HP-UX), ponieważ korzysta on z typowych narzędzi tych systemów (np. narzędzia Make). Pakiet także można stosować w różnych wersjach systemu Microsoft Windows, ale wtedy musi być dostępne środowisko Cygwin.

2. Zastosowana metoda

Pakiet używa następujących dobrze udokumentowanych i przetestowanych narzędzi: Automake, Autoconf, Make, Python, xsltproc, pdflatex i CVS, które są oprogramowaniem GNU [3]. Schemat przetwarzania danych jest przedstawiony na rys. 1. Pakiet zaprojektowano z myślą o prostocie, niezawodności i bezpieczeństwie.



Rys. 1. Schemat przetwarzania danych przez prezentowane oprogramowanie.

2.1. Prostota

Pakiet jest całkowicie oparty na plikach tekstowych, aby był on możliwie najprostszy w budowie. Pliki konfiguracyjne, skrypty, pliki danych (pliki XML) są plikami tekstowymi. Ta prostota pomaga w przenośności i niezawodności, ponieważ

administrator strony może rozszerzać pakiet i usuwać z niego błędy. Pakiet nie zawiera binarnych programów ani binarnych plików.

Każdy dokument HTML jest generowany przez pakiet przy użyciu polecenia „xsltproc” [4]. Administrator strony umieszcza dane na temat seminariów w plikach XML, natomiast instrukcje dotyczące formatowania strony w plikach XSL. Dlatego wygląd strony jest rozdzielony od informacji zawartych na stronie.

Pliki o nazwie „Makefile” (pliki konfiguracyjne oprogramowania Make) określają sposób oraz kolejność generowania plików wynikowych. Tworzenie plików „Makefile” jest bardzo pracochłonne i monotonne, ponieważ ich zawartość zazwyczaj jest łatwa do określenia. Z tego powodu do tworzenia plików „Makefile” jest użyte oprogramowanie Automake, które tworzy je w oparciu o informacje zapisane w swoich plikach konfiguracyjnych o nazwie „Makefile.am”, które są znacznie prostsze w porównaniu do plików „Makefile”. Przykładowy plik „Makefile.am” jest przedstawiony poniżej. Ten plik zawiera tylko 5 linii, natomiast jego plik wynikowy zawiera kilkadziesiąt linii.

```
SUBDIRS = data posters info archives images download
datadir = $(prefix)
data_DATA = index.html main.css
index.html: index.xsl index.xml ./data/.p/future.xml
$(XSLTPROC) --xinclude index.xsl index.xml > index.html
```

2.2. Niezawodność

Stacyczna strona z najmniejszym ryzykiem może przestać funkcjonować. Zależy nam jednak nam nad szybkim odtworzeniem strony, jeżeli taka konieczność zajdzie, na przykład po przypadkowym usunięciu jej z dysku. Z tego powodu pakiet jest przechowywany na serwerze CVS (Concurrent Versions System), skąd może zostać pobrany i użyty do odtworzenia strony [5]. Przechowywanie pakietu na serwerze CVS jest wydajne, ponieważ oprogramowanie CVS zostało zaprojektowane do wydajnej obsługi plików tekstowych. Co więcej, serwer CVS przechowuje historię zmian pakietu i dlatego jest możliwe odtwarzanie stanu strony z dowolnej chwili w przeszłości.

Oprogramowanie CVS nie jest nieodzowne dla pakietu, jest one jedynie zwykłą wygodą. W naszym przypadku regularne tworzenie kopii zapasowych potrafi zastąpić serwer CVS, ponieważ pakiet jest rozwijany przez tylko jedną osobę. Mimo tego, oprogramowanie CVS jest użyte, ponieważ jest ono niezawodne, proste w obsłudze i oszczędnie korzysta z dysku.

2.3. Bezpieczeństwo

Najbezpieczniejszą stroną jest zawsze strona statyczna i nasza metoda ma celu zapewnienie maksymalnego bezpieczeństwa. Pakiet generuje kompletną statyczną stronę, która potem jest kopiowana do określonego miejsca na dysku. Stąd, serwer HTTP nie uruchamia skryptów do dynamicznego generowania stron i dlatego nie istnieje ryzyko włamania się do serwera przez wykorzystanie słabych punktów w skryptach. Języki interpretowane służące tym celom (np. PHP, Python, Perl) są szczególnie podatne na te ataki. Co prawda pakiet zawiera kilka skryptów języka Python, ale są one uruchamiane

przez administratora strony, który nie ma celu w wykorzystaniu ich błędów i niedoskonałości.

3. Krótka instrukcja obsługi

Sposób użycia pakietu jest standardowy, ponieważ wykorzystuje on oprogramowanie Autoconf. Aby stworzyć stronę pierwsze należy uruchomić skrypt „configure”, co spowoduje skonfigurowanie pakietu. Następnie należy wykonać komendę „make” w celu stworzenia wszystkich plików strony. Na końcu należy wykonać komendę „make install” żeby zainstalować stronę.

Skrypt konfiguracyjny szuka lokalizacji programów „python” i „xsltproc” oraz dokonuje konfiguracji pakietu. Przy użyciu argumentu „--prefix” można przekazać lokalizację, gdzie należy zainstalować stronę, w przeciwnym razie zostanie domyślnie przyjęta wartość „/seminar”. Po konfiguracji pakietu następuje generacja plików strony, takich jak plików PDF z ogłoszeniami seminariów, czy dokumentów HTML z listą minionych referatów. Proces instalacji jedynie kopiuje określone pliki do katalogu docelowego.

Informacje na temat jednego seminarium przechowywane są w jednym pliku w formacie XML przechowywanym w podkatalogu „data”. Język Python [6] jest zastosowany do przetwarzania danych w formacie XML i tworzenie plików XML zawierających minione i nadchodzące seminaria. Przykładowy plik zawierający dane na temat jednego seminarium jest przedstawiony poniżej.

```
<?xml version="1.0" encoding="ISO-8859-2"?>
<presentation>
<datafile>jsmith_pres.ppt</datafile>
<date><year>2003</year><month>3</month><day>14</day></date>
<time><hour>10</hour><minute>00</minute></time>
<author>Prof<dot/>Jan Wiśniewski</author>
<title>Tytuł referatu</title>
<abstract>To jest streszczenie referatu.</abstract>
```

4. Zakończenie

Rozdział przedstawia pakiet oprogramowania do zautomatyzowanego prowadzenia strony seminaryjnej. Zastosowana metoda jest oryginalna i niekonwencjonalna: strona jest tworzona i instalowana jak zwykle oprogramowanie. Pakiet jednak nie tworzy i instaluje programów, lecz pliki składające się na stronę. To podejście jest niezawodne, bezpiecznie i proste.

Należy przyznać, że używanie pakietu będzie uciążliwe przy bardzo częstych aktualizacjach, ponieważ każda aktualizacja pociąga za sobą konieczność ponownej instalacji strony. Jeżeli aktualizacje występują umiarkowanie często, co jest sensownym założeniem dla strony seminaryjnej, to narzut pracy związany z prowadzeniem strony przy użyciu naszego narzędzia może zostać przyjęty w zamian za bezpieczeństwo serwera HTTP udostępniającego statyczną stronę.

Strona seminaryjna Instytutu Informatyki Teoretycznej i Stosowanej PAN [1] służy jako przykład strony prowadzonej przy użyciu prezentowanego pakietu.

LITERATURA¹

1. <http://www.iitis.gliwice.pl/seminaria>
2. <http://www.ma.utexas.edu/cgi-pub/seminar/calendar>
3. <http://www.gnu.org>
4. <http://www.xmlsoft.org>
5. <http://www.cvshome.org>
6. <http://www.python.org>
7. Lutz M.: Programming Python: Object-Oriented Scripting. O'Reilly & Associates, Inc., 2001.

¹ Podane są tylko adresy internetowe, ponieważ są to najbardziej aktualne i kompletne źródła na poruszane w tym rozdziale tematy. Dostępne są także książki oraz artykuły i dlatego np. pozycja [6] może zostać zastąpiona przez pozycję [7], jednak poprzednia pozycja jest bardziej kompletna i aktualna.