



jacek szlachetny

Istnieją skuteczne techniki filtrowania niechcianych e-maili

# Polowanie na SPAM

Batalia z pocztowymi śmieciami trwa. I choć ich nadawcy uciekają się do różnych sztuczek, internauci nie stoją na straconej pozycji. Naukowcy i programiści opracowali wiele metod walki ze spammerami. Metod skutecznych – trzeba dodać.

**Marcin Nowak**

Niedawno jeden z redakcyjnych kolegów zwrócił mi uwagę, że często piszemy o spamie. Ten temat rzeczywiście powraca od czasu do czasu, ale ma to uzasadnienie. Choć coraz więcej firm i instytucji dostrzega problem niechcianych maili, ich plaga nie traci na intensywności. Początkowo wiadomości reklamowe wyłącznie denerwowały internautów co bardziej wrażliwych na punkcie swojej prywatności. Teraz są postrzegane także jako problem gospodarczy. I choć firmy zajmujące się masowym wysyłaniem listów elektronicznych także płacą podatki, to te ostatnie są zdecydowanie mniejsze od strat spowodowanych przez spam. Efekt jest taki, że z wyjątkiem spamerów nikt nie wypowiada się pozytywnie o tym zjawisku.

Jakie szanse ma użytkownik w starciu z firmami, które na wysyłaniu e-maili zarabiają? Przede wszystkim po naszej stronie stoi prawo. Od niedawna obowiązuje ustawa, która zabrania wysyłania spamu. Niestety – w praktyce dotyczy ona tylko nadawców polskich listów. Możemy też dbać o „higienę” naszego adresu i nie umieszczać go w jawnej formie na stronach WWW i grupach dysku-

syjnych. Jednak nawet jeśli będziemy chronili nazwę naszej skrzynki, nie możemy być pewni sukcesu – spamerzy często wysyłają listy na losowe adresy.

Co zrobić, gdy śmieci zaczną docierać do naszej skrzynki? Na początku je kasujemy, później próbujemy stosowania różnych filtrów, instalowania specjalnych programów i zaczynamy się interesować, czy dostawca naszego konta pocztowego wykorzystuje mechanizmy obronne. Rozwiązań antyspamowych jest bardzo dużo, ale wszystkie one stosują jedną z kilku opracowanych technik. Zanim zdecydujemy się na zainstalowanie programu lub wybór danego dostawcy skrzynki pocztowej, warto zapoznać się z wykorzystywanymi mechanizmami, ich skutecznością, wadami oraz zaletami.

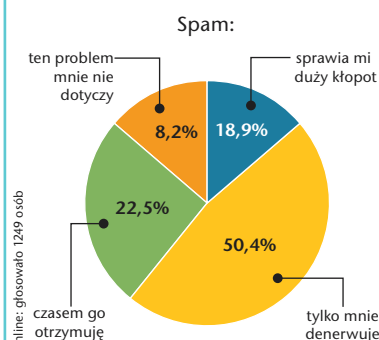
## Viagra, teen i money

Pierwsza metoda jest najbardziej intuicyjna i najłatwiejsza do zastosowania. Wykorzystuje ona fakt, że w znacznej części wiadomości spamowych możemy znaleźć te same, charakterystyczne frazy: sex, free, unsubscribe, viagra, loan itp. Mechanizm, który umożliwia korzystanie z tej techniki, to filtry,

obecne prawie w każdym programie pocztowym. Przeglądanie treści jest bardzo często wykorzystywane w programach antyspamowych. Ich skuteczność (a także samej metody) zależy od kilku szczegółów. Samo wyszukiwanie pewnych ciągów znaków nie jest zbyt efektywne. Często wymagane staje się sprawdzanie dodatkowych warunków. Wiele aplikacji rozróżnia małe i wielkie litery (spamerzy wykorzystują te ostatnie, by przyciągnąć naszą uwagę: WIN, FREE itp.). Popularna jest także analiza, czy dany ciąg znaków rozpoczyna lub kończy dany wyraz. Reguła typu „seks\*” zadziała przy słowach „seks”, „seksowna”, „seksu” itp., czasem stosowane są wyrażenia regularne. Bardzo dobre rezultaty daje zwracanie uwagi na to, w której części wiadomości występuje dana fraza. Rozróżniane są: treść wiadomości, temat, nadawca, odbiorca oraz inne nagłówki.

Technologia ta ma kilka zalet. Przede wszystkim jest intuicyjna, więc użytkownicy wykorzystujących ją programów sami mogą dodawać nowe reguły. Jest także dość skuteczna, choć tylko na początkowym etapie walki ze spamem. Umożliwia łatwą wymianę reguł (da się je opisać w pliku tekstowym). Powstał nawet do tego specjalny język o nazwie Sieve ([www.cyrusoft.com/sieve/](http://www.cyrusoft.com/sieve/)). Niestety, przeglądanie treści ma też wady. Największą jest jej ograniczona skuteczność. Spamerzy stosują wiele technik, które skutecznie „oślepiają” filtry (patrz: ramka „Spamerzy kontratakują”). Powoduje to, że bardzo często trzeba dodawać nowe reguły, co wraz z kolejną wadą – brakiem możliwości automatycznego tworzenia reguł – sprawia, że obsługa tego mechanizmu zajmuje czasem tyle, co ręczne wyłowienie i usunięcie spamu. Dodatkowo, jeśli dbamy o skuteczność aplikacji, filtrów przybywa i ich sprawdzanie trwa coraz dłużej. Wreszcie ostatnia wada – działanie tej metody zwykle ma charakter zerojedynkowy. Wystarczy jeden zły wyraz, a list zostaje uznany za spam. Jednak np.

## % Nie lubimy śmieci

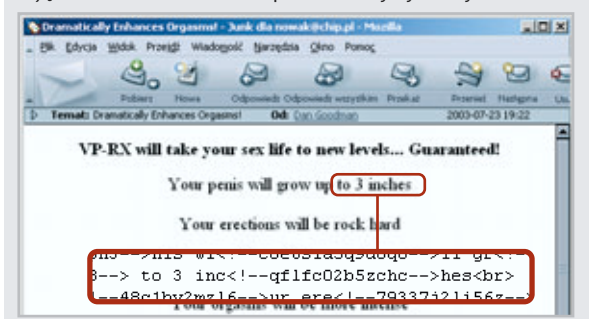


**Na PONAD POŁOWĘ INTERNAUTÓW spam działa irytująco.**

## i Spamerzy kontratakują

Nad opisanymi w tekście technologiami pracują tysiące programistów, ale podobny ruch trwa w obozie przeciwników. Nadawcy pocztowych śmieci opracowali wiele metod utrudniania pracy aplikacjom antyspamowym. Jednymi z pierwszych były wstawianie fałszywego adresu nadawcy (w ten sposób oszukiwane są czarne listy) i bardzo częsta zmiana numerów IP maszyn wysyłających wiadomości. Aby przechytrzyć technologię przeglądania treści, e-maile są kodowane (np. Base64), zapisywane w HTML-u, w treść są wpłatane niewidoczne dla użytkowników wyrażenia (np. `Via<!--GH-->gra` – patrz: ilustracja), wyrazy są zmieniane (np. `m0ney`) itp. Podstawą działania wielu technologii jest identyfikacja kopii tego samego spamu. By utrudnić ten proces, wiadomości są zmieniane („win money”, „win big money”, „You can win prize” itp.) albo dodawane są do nich losowe ciągi znaków.

Spamerzy często używają cudzych serwerów pocztowych (jeśli te nie są zabezpieczone) oraz wysyłają listy na adresy tworzone na podstawie słownika. Powstały też wirusy, które wysyłają maile reklamowe z komputerów zwykłych użytkowników.



SpamAssassin wymaga do tego obecności kilku obciążających słów lub innych przesłanek (program korzysta z wielu technik).

## Czarno na białym

Przeglądanie treści to nieskomplikowana metoda, ale istnieje inny sposób walki ze spamem, który opiera się na jeszcze łatwiejszych zasadach. Otóż mowa o czarnych i białych listach. Zasada działania pierwszej opiera się na obserwacji, że spamerzy wysyłają wiele e-maili oraz robią to często. Wystarczy zidentyfikować adresy, które pojawiają się w polu From, i ignorować kolejne przesyłki od „namierzonych” osób. Sposób ten kiedyś działał doskonale, ale teraz jego skuteczność jest ograniczona (patrz: ramka „Spamerzy kontratakują”). Skoro wiemy, jak działa czarna lista, łatwo wyjaśnić, czym jest jej biała odmiana. Otóż zawiera ona adresy zaufane, zwykle znajomych i kontrahentów (a także list dyskusyjnych). Przesyłki pochodzące z tych źródeł akceptowane są zawsze. Oczywiście stosowanie białej listy nabiera sensu, w chwili gdy używamy też innych metod filtrowania spamu.

Czarna i biała lista mają swoje odmiany. Nie muszą one koniecznie zawierać pełnych adresów. Czasami odrzucane lub akceptowane są wszystkie wiadomości z wybranych domen, a podstawą do odmowy odbioru może być adres IP czy domena nadawcy albo jego serwera pocztowego. To ostatnie dotyczy głównie maszyn określanych jako Open Relay, czyli takich, przez które e-maila może wysłać każdy i do każdego. Podobnie traktowane są przesyłki nadchodzące z niezabezpieczonych serwerów proxy. Takie maszyny są często wykorzystywane przez spamatorów chcących zachować anonimowość i dlatego administratorzy serwerów bojkotują je.

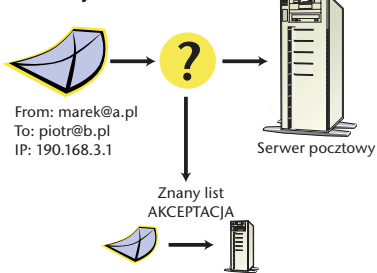
Oczywiście czarna lista niekoniecznie musi być wykorzystywana przez klient pocztowy. Równie często, a może częściej używają ich administratorzy serwerów pocztowych. 130 »

## INTERNET

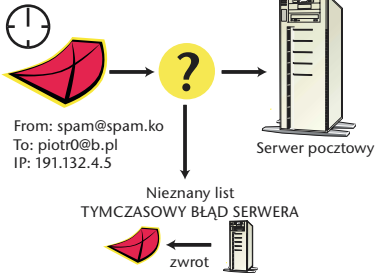
## Techniki walki ze spamem

## Zasada działania szarej listy

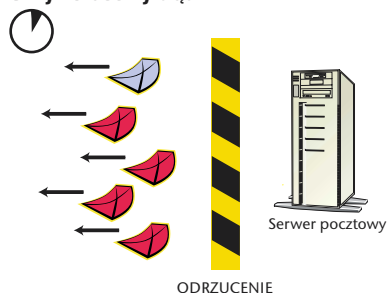
## 1. Nadejście listu



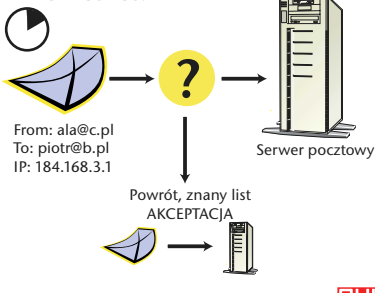
## 2. Odrzucenie listu



## 3. Tymczasowy błąd



## 4. Powrót listu



Jeśli nadejdzie list od **NIEZNANEGO NADAWCY** (2), przesyłka ta oraz wszystkie, które przyjdą przez następną godzinę, są odrzucane (3). Po pewnym czasie serwery pocztowe nadawców ponowią próbę transmisji (4). Spamerzy nie troszczą się jednak o swoje e-maile i one nie wracają.

Aby zwiększyć skuteczność tej metody walki ze spamem, informacje o czarnych wpisach przesyłają i pobierają oni ze specjalnych baz danych (jedną z nich był PolSpam – patrz: 4/2003, 196), które są tworzone przez wielu administratorów i zwykłych użytkowników.

## Szara lista

Dość oryginalną metodę walki ze spamem opracował Evan Harris. Szara lista to mechanizm, który opiera się na następujących za-

łożeniach: spamerzy wysyłają wiele wiadomości naraz i nie troszczą się o ich dalszy los. Szara lista musi być zaimplementowana na serwerze pocztowym. W tej metodzie wszystkie e-maile są identyfikowane dzięki trzem informacjom: numerze IP komputera dostarczającego pocztę, adresie nadawcy oraz odbiorcy. Jeżeli do serwera dotrze przesyłka o nieznaną triadzie, następuje odmowa przyjęcia e-maila wraz z komunikatem o tymczasowym błędzie. Podobnie traktowane są wszystkie przesyłki, które nadejdą

przez zadany czas. Ponieważ – jak pisałem – spamerzy zwykle wysyłają wiele listów, dzięki takiemu zachowaniu serwera kolejne niechciane listy są odrzucane bez niepotrzebnego sprawdzania ich triad. „Odbite” wiadomości reklamowe zwykle nie są ponownie wysyłane (ich nadawcy stosują metodę „wyslij i zapomnij”) – w przeciwieństwie do zwykłych e-maili, które po określonym czasie przyjdą ponownie.

Zaletami tej metody są – przynajmniej obecnie – spora skuteczność i brak interwencji użytkownika. Autor techniki szarej listy opisuje, jak mogą się zachować spamerzy chcący obejść to zabezpieczenie. Technika ta jest nieodporna tylko na jeden scenariusz – ponowne wysłanie list z tego samego komputera. Takie działanie zwiększy znacznie koszt nadania pojedynczej wiadomości reklamowej, ale nie wykluczone, że tak zachowują się spamerzy lub że znajdą inny trik. Szara lista ma jeszcze jedną zasadniczą wadę: opóźnia dostarczanie zwyczajnych listów. Evan Harris twierdzi, że ta niedogodność jest bardzo mała, ale niewykluczone, że w niektórych wypadkach może mieć duże znaczenie. Tym bardziej że przyzwyczailiśmy się i liczymy na to, iż e-maile dostarczane są zwykle w ciągu kilku minut.

## Pytanie-odpowieź

Szara lista wykorzystywała fakt, że spamerzy nie zajmują się swoimi e-mailami po ich wysłaniu. Ta sama właściwość legła u podstaw innej metody, którą można nazwać pytanie-odpowieź (ang. challenge-response). Adekwatna jest też inna nazwa: „tylko z pozwoleniem”. Osoba wykorzystująca tę metodę odgradza się od świata i przyjmuje wyłącznie e-maile od nadawców, którym ufa (w tym miejscu pytanie-odpowieź przypomina białą listę). Aby jednak nieznajomi mogli się z nią

## i Gdy internauci współpracują

Jaka jest największa wada spamu? To, że jest go dużo. Można tego faktu użyć przeciwko pocztowym śmieciom – wystarczy ignorować te przesyłki, których jest dużo. Na tym opiera się technika Distributed Checksum Clearinghouse. DCC składa się z wielu serwerów. Każdy identyfikuje przechodzące przez niego listy i notuje liczbę kopii poszczególnych mejli. W celu porównywania listów oczywiście nie przechowuje całych wiadomości, ale dla każdego tworzy sumę kontrolną. Jeśli do serwera dotrze list o znanym kodzie, przypisany jej licznik jest zwiększany o jeden. Gdy okaże się, że dana wiadomość pojawia się zbyt często, jest ona uznawana za spam.

System DCC składa się z serwerów gromadzących i wymieniających się informacjami o liczbie poszczególnych e-maili oraz klientów, które wysyłają swoje obserwacje do serwerów i pobierają dane sumaryczne.

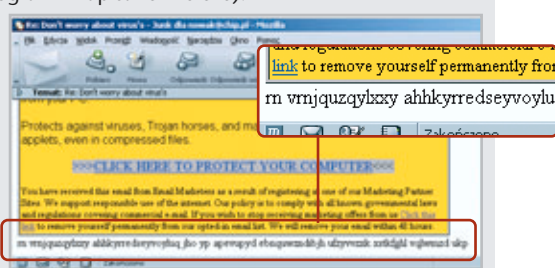
Klienci mogą być aplikacjami współpracującymi z serwerami pocztowymi lub programami zainstalowanymi u użytkowników. Dzięki temu, że całość jest siecią kooperujących ze sobą agentów, skuteczność zliczania spamu jest duża.

Ze współpracy internautów czerpie siłę także inne rozwiązanie zaimplementowane w programie SpamNet (znanym też jako Razor). Aplikacja ma formę wtyczki do programu pocztowego. Jego użytkownicy specjalnym przyciskiem zaznaczają spam. Jeżeli dana wiadomość zostanie uznana za niechcianą reklamę przez odpowiednio wiele osób, u pozostałych internautów zostanie automatycznie rozpoznana.

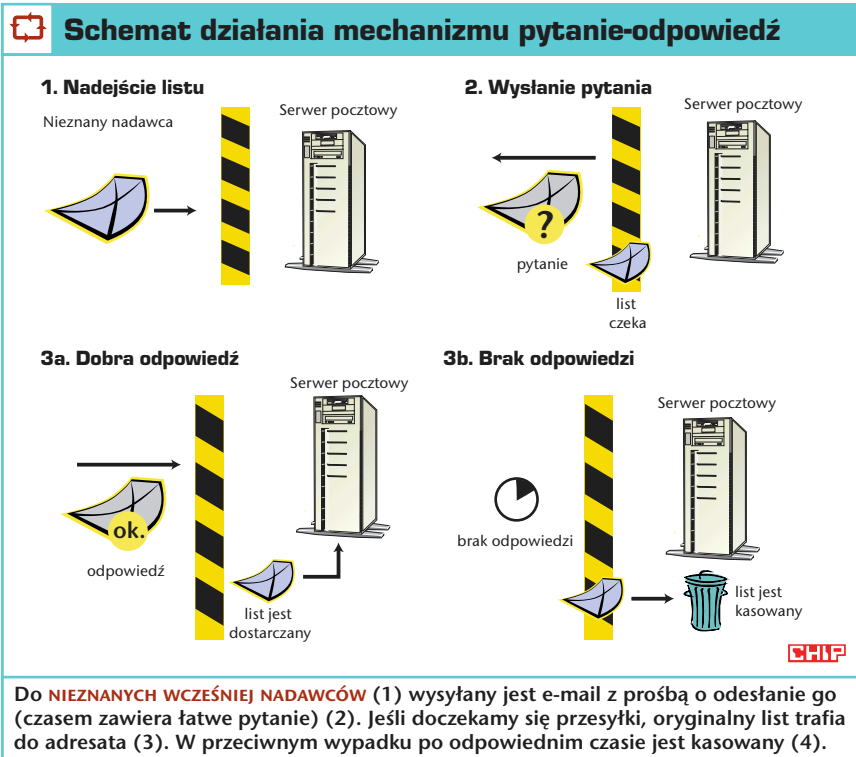
Twórcy tych technik mają wspólny problem: jak rozpoznawać spam, którego kopie są modyfikowane (patrz:

ilustracja poniżej)? Opracowano różne sposoby tworzenia sum kontrolnych, ale spamerzy próbują uodpornić na nie swoje listy. Choć cały czas trwa walka, SpamNet i DCC są skuteczne.

Ze współpracy internautów korzystają też Spam Assassin (wymiana reguł) oraz aplikacje firmy Vircom (członkowie VOP Anti-Spam Coalition tworzą filtry antyspamowe napisane w Sieve).







skontaktować, wszystkim wysyła wiadomość z łatwym pytaniem lub wręcz tylko z prośbą o odpowiedź. E-maile autorów, którzy napiszą do nas powtórnie, są przepuszczane, a reszta kasowana. Oczywiście człowiek bez trudu spełni polecenie, a automat spamera na tej próbie się „wyłoży”.

Idea jest jasna, ale jak zwykle diabeł tkwi w szczegółach. Nie zawsze korespondujemy z ludźmi. Wiele osób otrzymuje wiadomości od przeróżnych mechanizmów (potwierdzenia rejestracji, e-maile z list mailingowych itp.). Aby nasz „szlaban” działał prawidłowo, musimy takich nieożywionych nadawców dodać do białej listy. Nie każdy użytkownik jest w stanie to zrobić, a i nie zawsze wiemy, że automatyczna odpowiedź do nas przyjdzie.

Pozwolenie trzeba dać także naszemu adresowi – wiele osób wysyła sobie kopie listów albo artykuły z sieciowych archiwów. A tymczasem niektórzy spamerzy jako nadawcę ustawiają nas! W takim przypadku e-maile reklamowe przejdą przez naszą zaporę. Pomijając ten jeden przypadek, metoda pytanie-odpowieź jest skuteczna, a nawet – jak napisałem wyżej – zbyt skuteczna dla niektórych pożądaných listów. W niektórych rozwiązaniach technikę tę uelastyczniono. System TMDS ([tmda.net](http://tmda.net)) na podstawie określonych nagłówek może wiadomości od razu kasować (czarna lista) lub wpuszczać (biała).

### Kłania się statystyka

Wszystkie przedstawione techniki mają swoje zalety, ale i wady. Ciemnych stron nie jest pozbawione ostatnie podejście, jednak od kilku lat uznaje się je za najbardziej obiecujące.

Potwierdzeniem tych słów jest fakt, że zainteresowali się nim badacze Microsoftu, co zakończyło się uzyskaniem patentów.

Filtr bayesowski korzysta ze zdobyczy rachunku prawdopodobieństwa i statystyki. Idea opiera się na założeniu, że w wiadomościach reklamowych pewne wyrazy pojawiają się częściej niż w zwykłych listach. Pomysł, sprawiający w pierwszej chwili wrażenie naiwnego, okazuje się bardzo trafny. Badacze zajmowali się nim do dawna, ale do szerszej świadomości trafił on dzięki słynnemu artykułowi Paula Grahama „A Plan for Spam”.

Nie wdając się w szczegóły, działanie filtru można opisać następująco: rozpoczyna się ono od przeanalizowania pewnej porcji „dobrych” oraz „złych” e-maili. Efektem tej operacji jest tablica, w której każdemu wyrażeniu z listów przypisano prawdopodobieństwo, że zawierająca go wiadomość jest oraz że nie jest spamem. Gdy pojawi się nowy list, wyławiana jest określona liczba (np. 15) wyrażań, którym odpowiadają największe prawdopodobieństwa z przygotowanej wcześniej tablicy. Wtedy na podstawie wzoru Bayesa ([www.paulgraham.com/naivebayes.html](http://www.paulgraham.com/naivebayes.html)) obliczane jest prawdopodobieństwo, że dany e-mail jest lub nie jest spamem.

Przedstawione podejście jest tylko podstawą do tworzenia skutecznych mechanizmów obronnych. Niezwykle istotne są szczegóły, na przykład to, co jest traktowane za wyrażenie, czy rozróżniane są duże/małe litery, czy bierzemy pod uwagę miejsce wystąpienia danej frazy (temat listu, nadawca, treść) itp. Można również badać pary lub trójki wyrazów, ignorować lub nie znaczniki

HTML-a, wagę ma też to, jak traktujemy wyrażenia, których wcześniej nie widzieliśmy. Badacze Microsoftu biorą również pod uwagę porę dnia, w której e-mail dotarł do odbiorcy.

Filtry Bayesa dowiodły swojej skuteczności, a poświadczyć mogą o tym także ja. Mechanizm ten zaimplementowano w Mozilli 1.3 oraz Netscape’ie 7.1. W ciągu kilku tygodni prawidłowo wyłapał kilkaset niechcianych maili. Pomylił się tylko kilka razy, pomijając początkowy okres, kiedy się uczył. Zdarza mu się też zaliczać do śmieci listy, które wyglądają jak spam, choć nim nie są (newslettery). Zaletą tej techniki jest automatyczne budowanie statystyk – w przeciwieństwie np. do metody przeglądania treści, gdzie sami musimy tworzyć reguły. Filtr Bayesa jest odporny na większość prób zmylenia go. Doskonale wyłapuje dziwne wyrazy (v1agra, se<df>x), które przecież nie występują w zwyczajnych e-mailach.

### Łączmy je

Żadna z opisanych metod nie jest idealna. Najlepsze rezultaty daje ich połączenie, a białe i czarne listy są wręcz zalecane jako uzupełnienie każdej z opisanych technik. Takie podejście prezentuje np. SpamAssassin, który wykorzystuje prawie wszystkie opisane metody.

### Więcej informacji

#### TMDS

<http://tmda.net/>

#### SPAMASSASSIN

<http://www.spamassassin.org/>

#### DISTRIBUTED CHECKSUM CLEARINGHOUSE

<http://www.rhyolite.com/anti-spam/dcc/>

#### SZARA LISTA

<http://projects.puremagic.com/greylisting/>

#### SPAMNET

<http://www.spamnet.com/>

<http://razor.sourceforge.net/>

#### BADANIA MICROSOFTU (FILTR BAYESA)

<http://research.microsoft.com/~horvitz/junkfilter.htm>

#### ARTYKUŁ PAULA GRAHAMA (FILTR BAYESA)

<http://www.paulgraham.com/spam.html>

#### FILTR BAYESA

<http://news.bbc.co.uk/1/hi/technology/3014029.stm>

#### METODY WALKI ZE SPAMEM

<http://www.computerbits.com/archive/1998/0600/lan9806.html>

<http://www.mspalliance.com/articles/?artid=272>

#### PORÓWNIANIE METOD ZWALCZANIA SPAMU

<http://www-106.ibm.com/developerworks/linux/library/l-spamf.html>

#### PROGRAMY I SYSTEMY ANTYSZPAMOWE

<http://dmoz.org/Computers/Internet/Abuse/Spam/Filtering/>

#### PYTANIE-ODPOWIEŹ

<http://www.politechbot.com/p-04746.html>

<http://www.templetons.com/brad/spam/challengeresponse.html>

#### JĘZYK SIEVE

<http://www.cyrusoft.com/sieve/>