



SPAM?

Szinte már el is felejtettem, mi az...

Sütő János (sj@acts.hu)

Miért érdemes statisztikai spamszűrőt használni?



Fals pozitív hibaarány (FP): <0.1%

Fals negatív hibaarány (FN): <0.5%

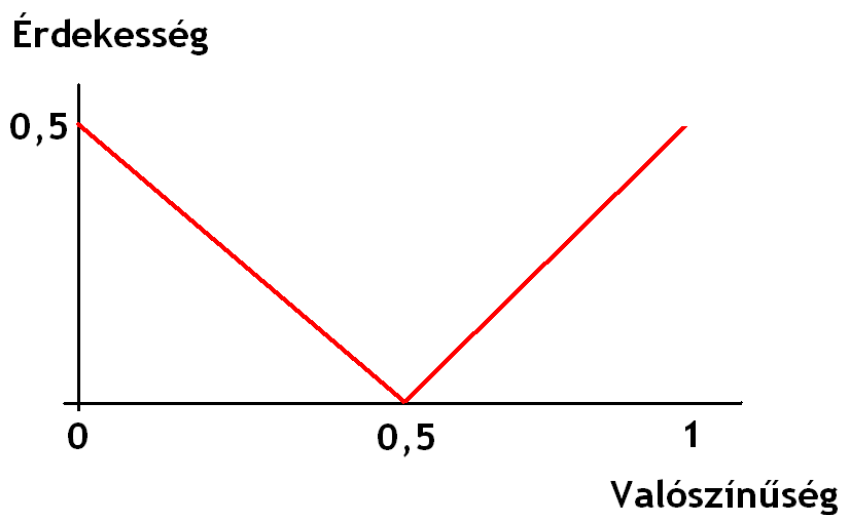
Graham: Plan for spam



Hogyan működik?



Érdekesség és valószínűség

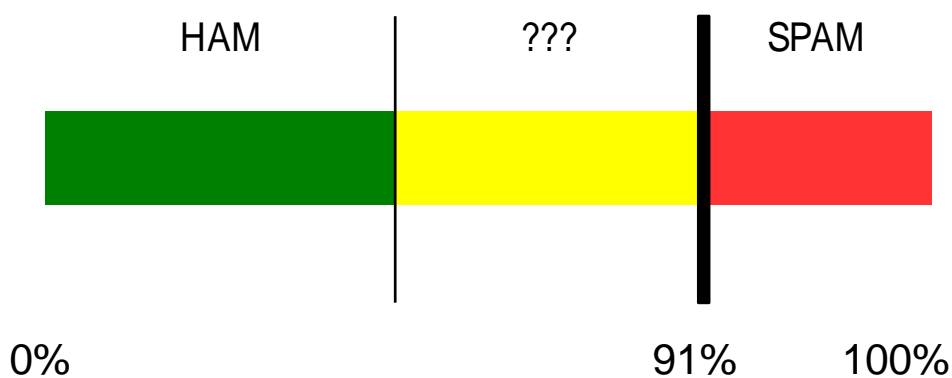


Döntési mátrix

enhanced+libido	0.9999
Subject*impress	0.9934
triples	0.9311
absolutely	0.8933
...	...



Statisztikai összefoglalás



Received: from [188.55.47.22] (**unknown** [188.55.47.22])
From: <sj@acts.hu>
 Date: Mon, 6 Sep 2010 03:00:45 +0300
 Subject: Seems **you like** it **all**
 Reply-To: <sj@acts.hu>
 MIME-Version: 1.0
 Content-Type: text/plain; charset="**ISO-8859-1**"
 Content-Transfer-Encoding: 8bit
 Message-Id: <20100906000046.6F72D15002@mx.acts.hu>

We meet at 8 p.m. <http://www.nodream.ru/>



URL*ru	0.9990
Subject*like	0.9879
Subject*you	0.9709
Subject*all	0.9591
UNKNOWN_CLIENT*	0.9999
FROM*<sj@acts.hu>	0.9983
HEADER*iso-8859-1	0.8774

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 9 [ms]



Received: from 82.subnet110-138-138.speedy.telkom.net.id
 (**unknown** [110.138.138.82])
 Received: from [110.138.138.82] (port=4990 helo=USERE5)
 by mx-exchange.cbeyondonline.net with smtp
 From: "**Herbert** Rodrigues" <sexagenarianf60@rossjohnsonweb.com>
 To: nekem
 Subject: **The male** dignity corresponds to **what is hidden** within the
pants. Make your device 6 **inches** long **really fast**
 X-Priority: 3
 X-MSMail-Priority: Normal
 X-Mailer: Microsoft Windows Mail 6.0.6001.18000
 X-MimeOLE: Produced By Microsoft MimeOLE V6.0.6001.18049
 X-Spam: Not detected

Dont wait for miracle, do it now.. Be successful due to long item.

Enter without knocking **<http://countgrip.ru>**



UNKNOWN_CLIENT*	0.9999
FROM*herbert	0.9520
Subject*the	0.8780
Subject*male	0.9900
Subject*what	0.8814
Subject*hidden	0.9928
Subject*pants.	0.9200
Subject*make	0.9218
Subject*your	0.9534
Subject*inches	0.9985
Subject*really	0.9392
Subject*fast	0.9791
enter+without	0.9400
URL*ru	0.9990

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 13 [ms]



Received: from esimsbgez (**unknown** [212.45.112.169])
 Subject: Incredib1e **gains** in length of 3-4 **inches** to **yourPenis**,
 PERMANANTLY or6n
 From: "Maryellen Gricelda" <maryellen_griceldarw@cs.com>

You **Have Everything** To Gain!
 Incredib1e **gains in length** of 3-4 inches to yourPenis,
 PERMANANTLY

- : **Amazing increase in thickness** of yourPenis, up to 30%
- : BetterEjacu1ation control
- : **Experience Rock**-HardErecetions
- : Explosive, intenseOrgasns
- : **Increase volume** ofEjacu1ate
- : **Doctor designed and endorsed**
- : 100% herbal, 100% **Natural**, 100% **Safe**

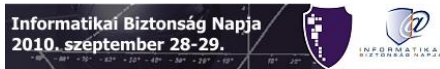
The proven NaturalPenisEnhancement that works!
 100% MoneyBack Guaranteed

<http://hardbodies.ru>



UNKNOWN_CLIENT*	0.9999
FROM*maryellen	0.9314
Subject*gains	0.9808
Subject*inches	0.9985
Subject*yourpenis	0.9888
have+everything	0.9314
gains+length	0.8800
increase+thickness	0.8800
amazing+increase	0.9314
experience+rock	0.8800
increase+volume	0.8800
doctor+designed	0.8800
and+endorsed	0.9314
natural+safe	0.8800
the+proven	0.9040
URL*ru	0.9990

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 13 [ms]



Received: from mchsi.com (97-64-142-210.client.mchsi.com
 [97.64.142.210])
 From: "RealPfizer's Shop" <seaasyxu2713@mchsi.com>
 Subject: **Hey sj up to 80% off. seen**

View as a web page

Merit pay systems are on **the rise** for teachers, paying teachers
 extra money based on excellent classroom evaluations, high **test
 scores and** for high success at their overall school.
 Numerous sets of **statistics have been proposed** to demonstrate
 the proportionate origins of **English vocabulary**.
 But the idea that as a **movement it was** in some way inimical to **the
 Church**, or to the conservative social order in **general is one** that
 has been **put forward for a century and more** without any
 substantial proof **being offered**.



ZOMBIE*	0.9999
general+one	0.9200
english+vocabulary	0.9467
movement+was	0.9520
the+church	0.9977
the+conservative	0.9845
century+and	0.9721
put+forward	0.9314
statistics+have	0.8800
the+rise	0.9088
been+proposed	0.9631
scores+and	0.8800
being+offered	0.9893
test+scores	0.9040
been+put	0.8800
for+century	0.9314

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 13 [ms]



Received: from 188.173.128.203 (**unknown** [188.173.128.203])
 From: "Balázs Péter" <balazs.peter2011@gmail.com>
 Subject: Legális címlisták, hírlevélküldés

Az **elektronikus Direkt Marketing** közismerten a leghatékonyabb reklámozási módszer.. Közvetlenül, **emailben keresheti meg** potenciális ügyfeleit, a címzettek jelentős része biztosan meg **fogja ismerni** a levélben bemutatott terméket, vagy szolgáltatást.

Ma Magyarországon tehát **csak a nem** természetes címekre küldhető tömeges reklámlevél (email) (és a legutóbbi törvényváltozás után ezek használhatók szabadon és LEGÁLISAN) - **vagyis olyan email címekre, melyek nem** tartalmazznak személynevet (példa:info@ceg.hu, office@ceg.hu **cím megfelel** a kritériumoknak).

Mi összesen 400.000 magyarországi, tömeges hírlevélküldésre alkalmas vállalkozás email címével rendelkezünk, **melyeket az idei** év folyamán nyilvános forrásokból állítottunk össze. A címek formátuma: info@ceg.hu, office@ceg.hu, ceg@ceg.hu

Az **email címlista** ára most kedvezményes: 17.000 Forint **Rendelje meg**



UNKNOWN_CLIENT*	0.9999
elektronikus+direkt	0.9467
direkt+marketing	0.9467
emailben+keresheti	0.9200
keresheti+meg	0.9200
fogja+ismerni	0.9200
office+cim	0.9200
email+cimekre	0.9200
vagyis+olyan	0.9200
olyan+email	0.9200
cimekre+melyek	0.9200
melyek+nem	0.9200
cim+megfelel	0.9200
melyeket+idei	0.9200
email+cimlista	0.9200
rendelje+meg	0.9631
csak+nem	0.0648
abban+esetben	0.0602

spaminess: 1.0000 in 41 [ms]



Received: from ****travel.hu (i220-101.shosting.atw.hu [88.151.101.220])
 From: ****Travel <pipoly@****travel.com>
 Subject: új szállások új ajánlatok

< ... bla-bla-bla ... >

Tiszteletben tartva az 1997. évi LVIII. törvény 15. § (1), továbbá az **információs társadalommal összefüggő szolgáltatások egyes kérdéseiről szóló 2001. évi CVIII. törvényt**, az Ön e-mail címe egy nyilvánosan megvásárolható, illetve elérhető marketing adatbázisban rögzített, vagy az interneten elhelyezett, bárki számára hozzáférhető, nyilvános adatbázisból származik.



ZOMBIE*	0.9999
tiszteletben+tartva	0.9467
információs+társadalommal	0.8800
szolgáltatások+egyres	0.8800
szóló+évi	0.8800
évi+cviii	0.9200

...

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 15 [ms]

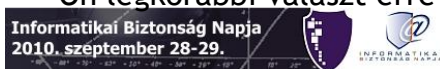


Received: from crossmanagement.net (**unknown** [220.99.116.221])
 From: "**patrick chan**"<mrpatrickchan66@yahoo.com.hk>
 Content-Type: text/plain; charset="**Windows-1251**"
To: undisclosed-recipients;
 Subject: üzleti lehetőség

Üzleti lehetőség
 Figyelem

Én **vagyok Mr. Patrick Chan** KW ügyvezető igazgató és gazdasági
 vezérigazgató-**Hang Seng Bank Ltd., Hong Kong**.
 Nekem van egy jövedelmező vállalkozás javaslatot kölcsönös
 érdeklődésre számot tartó **megosztani veletek;**
 ez magában foglalja az átadás egy nagy pénzüsszeg.
 Megkaptam az utalást én **keressen valakit, aki megfelel** az én
 javasolt üzleti kapcsolat.
 Ha érdeklő a **munka velem kapcsolatba velem** az én privát e-mail
 (patrik_kwchan009@**yahoo.com.hk**) további részletek

Ön legkorábbi választ erre a levélre, lesz méltányol.



UNKNOWN_CLIENT*	0.9999
FROM*patrick	0.9822
FROM*chan	0.9800
HEADER*windows-1251	0.9076
HEADER*undisclosed-recipients	0.9772
vagyok+patrick	0.9680
patrick+chan	0.9800
aki+megfelel	0.9680
hang+seng	0.9845
seng+bank	0.9845
hong+kong	0.9342
megosztani+veletek	0.9200
keressen+valakit	0.9200
munka+velem	0.9400
velem+kapcsolatba	0.9400
kapcsolatba+velem	0.9467

spam=1.000000, ham=0.000000
 spaminess: 1.0000 in 11 [ms]



Saját pontosságom

2010. augusztus:	FP: 2 / 3644 = 0.05%
	FN: 4 / 1790 = 0.22%
Legjobb hónap:	FP: 0 / 2913 = 0.00%
	FN: 1 / 1506 = 0.06%



Nagy teljesítmény

freemail.hu (2007): 18M levél/nap



=



⋮

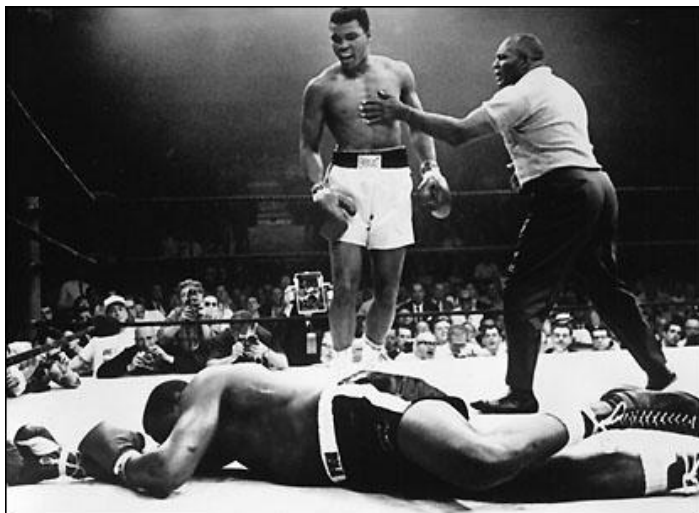
18 db HP DL 145*

⋮

4 db Core2Duo desktop PC



Hogyan lehet legyőzni a statisztikai spamszűrőket?



Informatikai Biztonság Napja
2010. szeptember 28-29.



Ha nem használjuk



Informatikai Biztonság Napja
2010. szeptember 28-29.





<http://clapf.acts.hu/>

