

Statisztikai spamszűrők

Hatékony védelem a spam ellen



Sütő János (sj@acts.hu)

Graham - Plan for spam

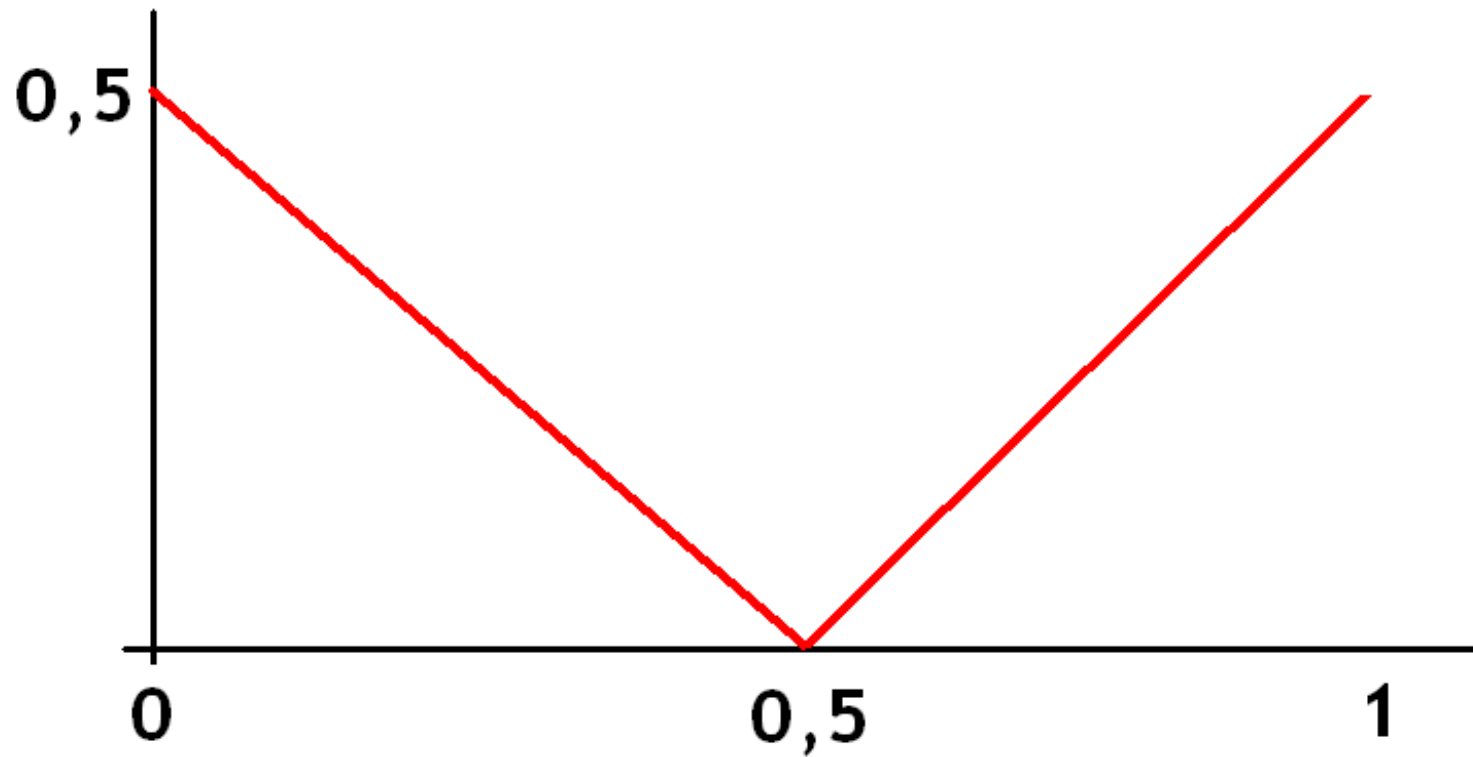


Dekódolás, tokenekre bontás



Valószínűség és érdekesség

Érdekesség



Valószínűség

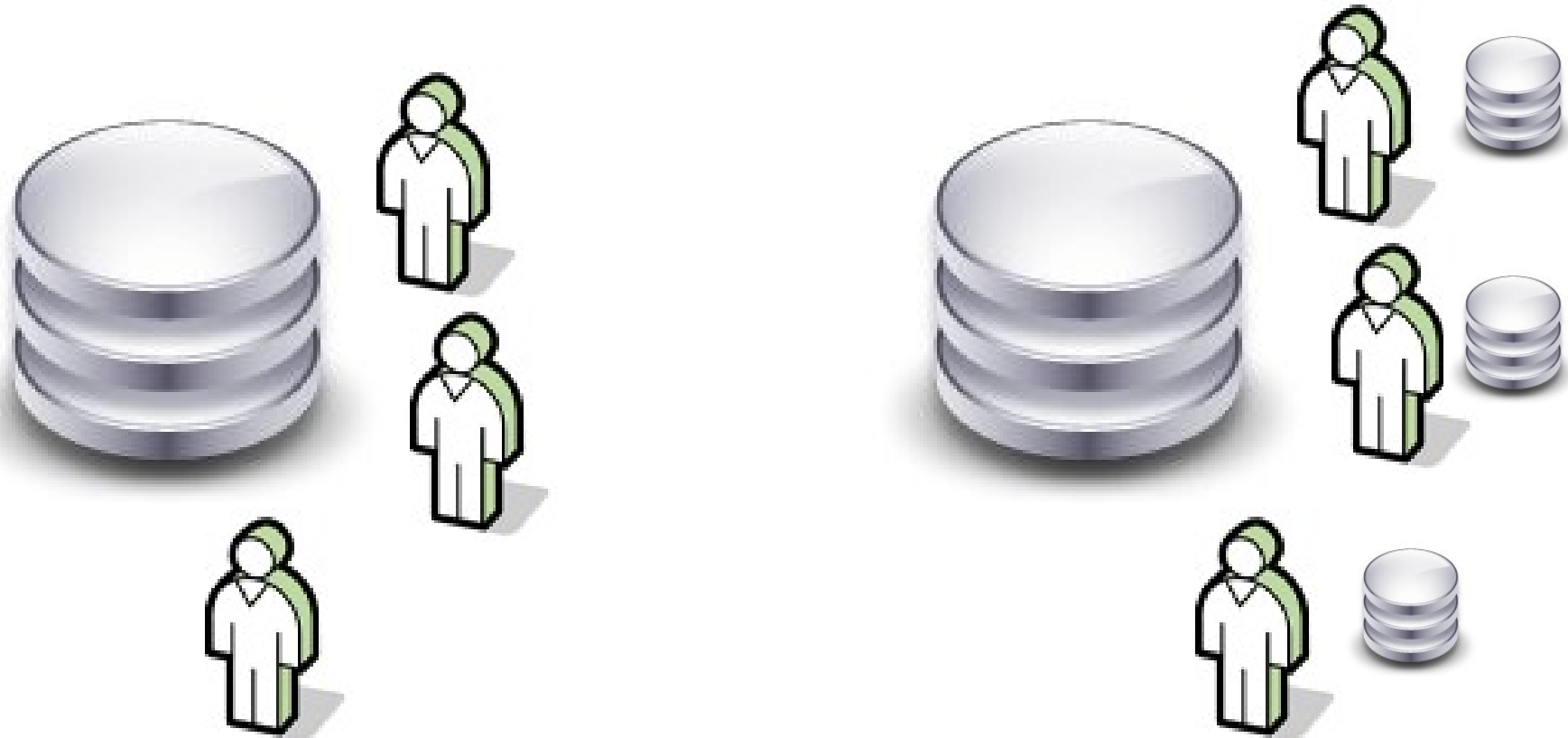
Döntési mátrix felépítése

enhanced+libido	0,9999
Subject*impress	0,9934
triples	0,9311
absolutely	0,8933

Statisztikai összegzés



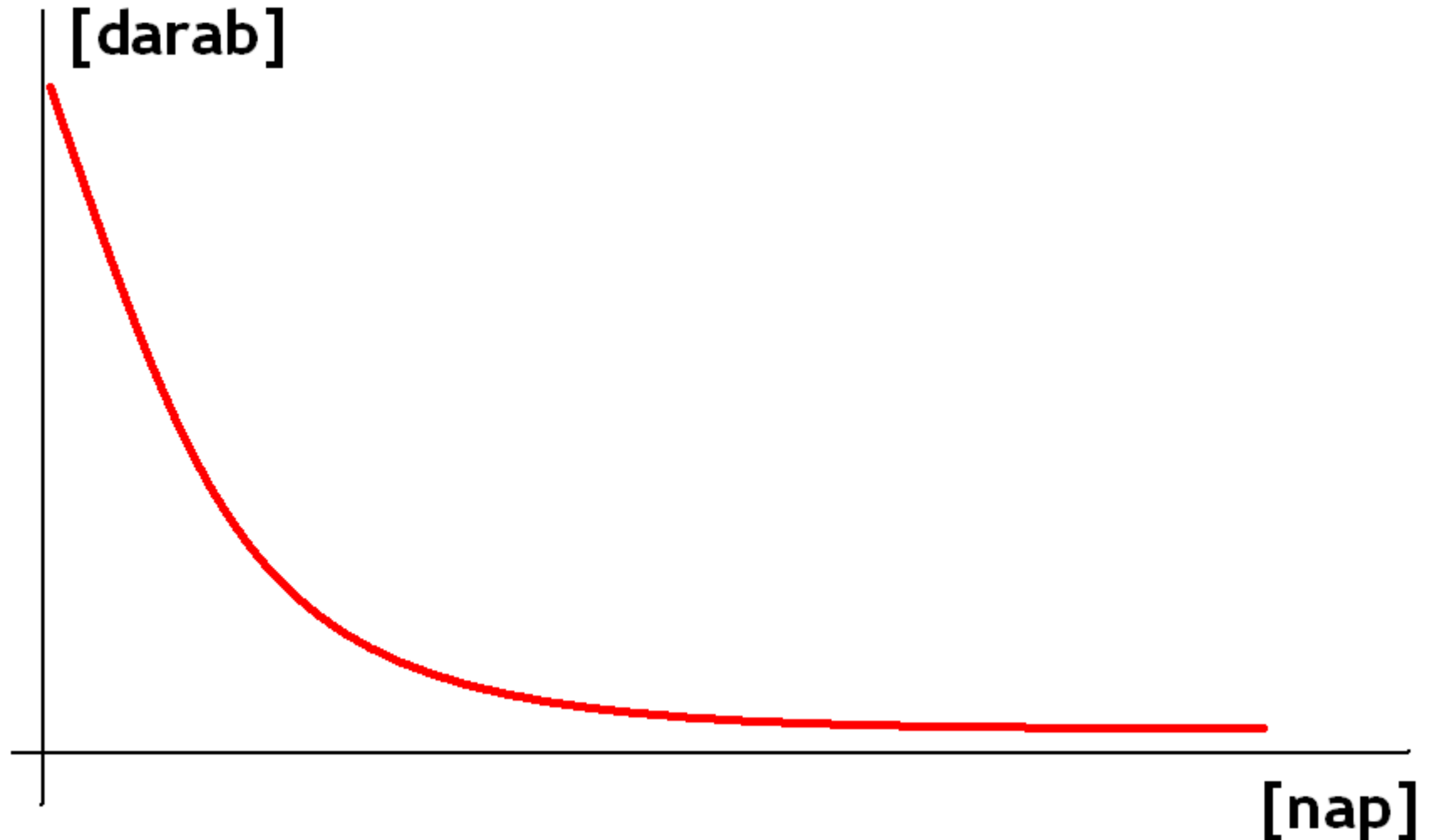
Token adatbázis



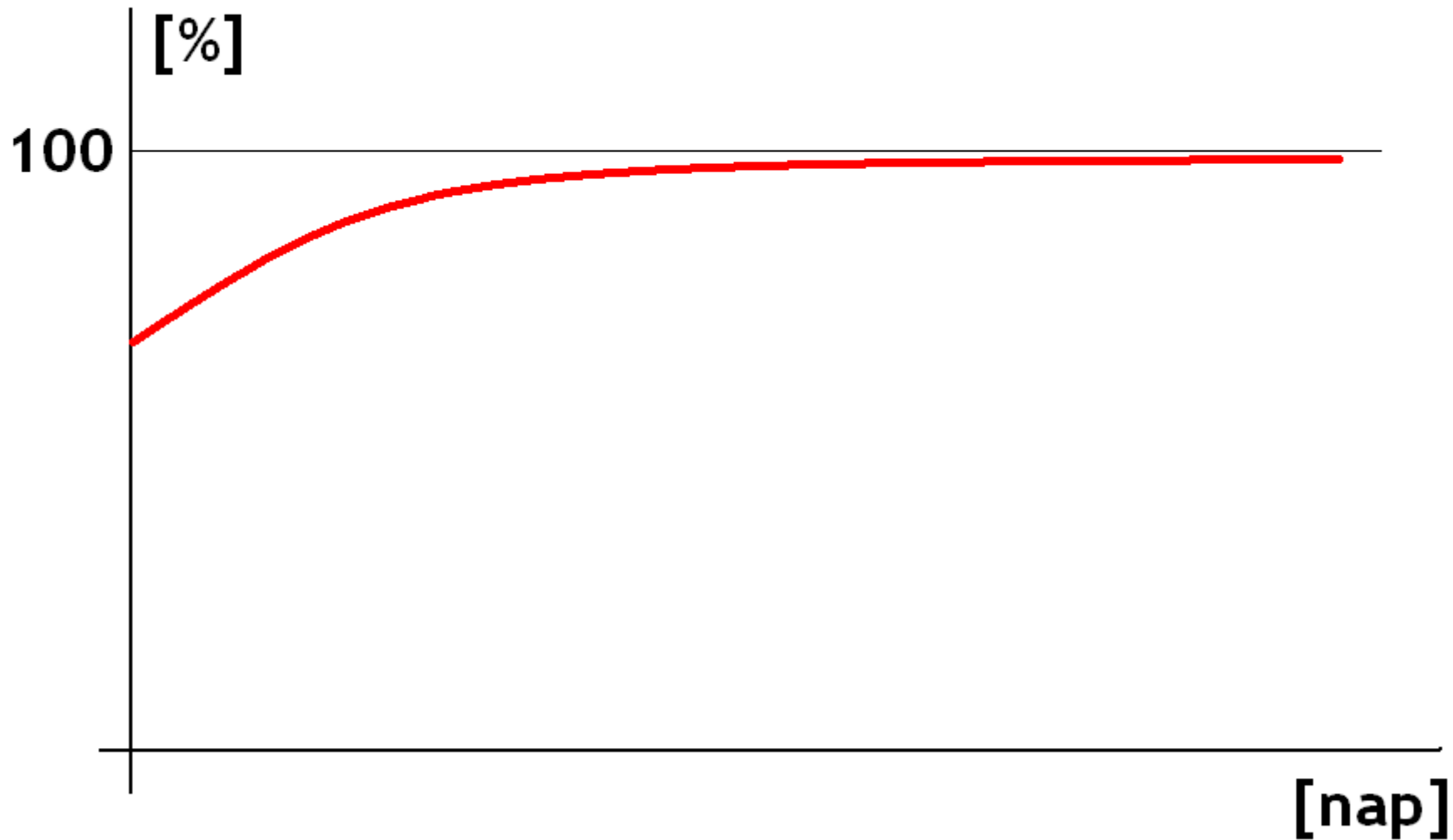
Tanítás



Tanítás mennyisége vs. idő



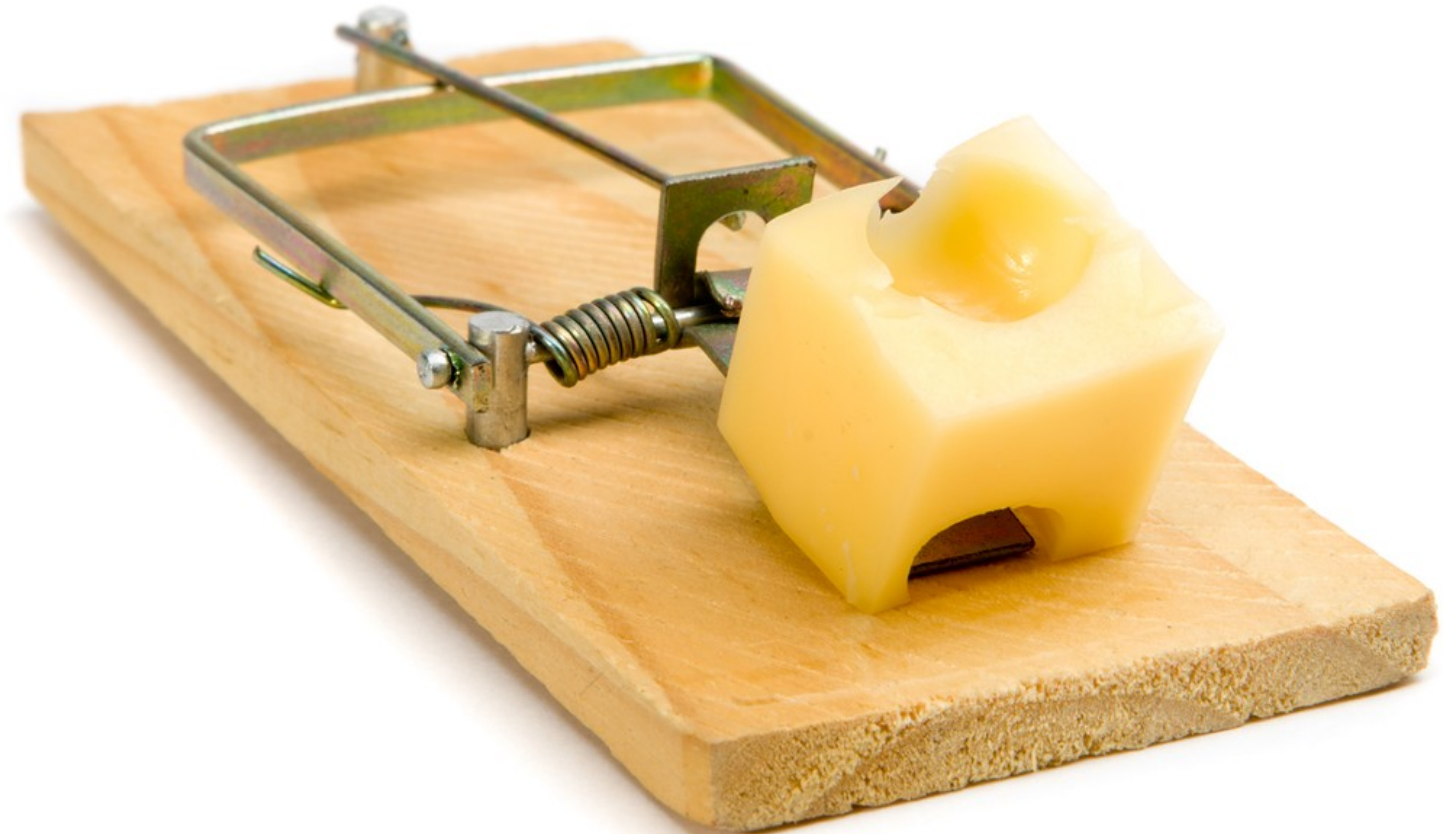
Pontosság vs. idő



Fehérlista

0. SQL táblában tárolt minták
1. Név, email, domain, IP-cím, FQDN
2. „From:” sor alapján

Csapda email cím + aknamező =
intelligens feketelista

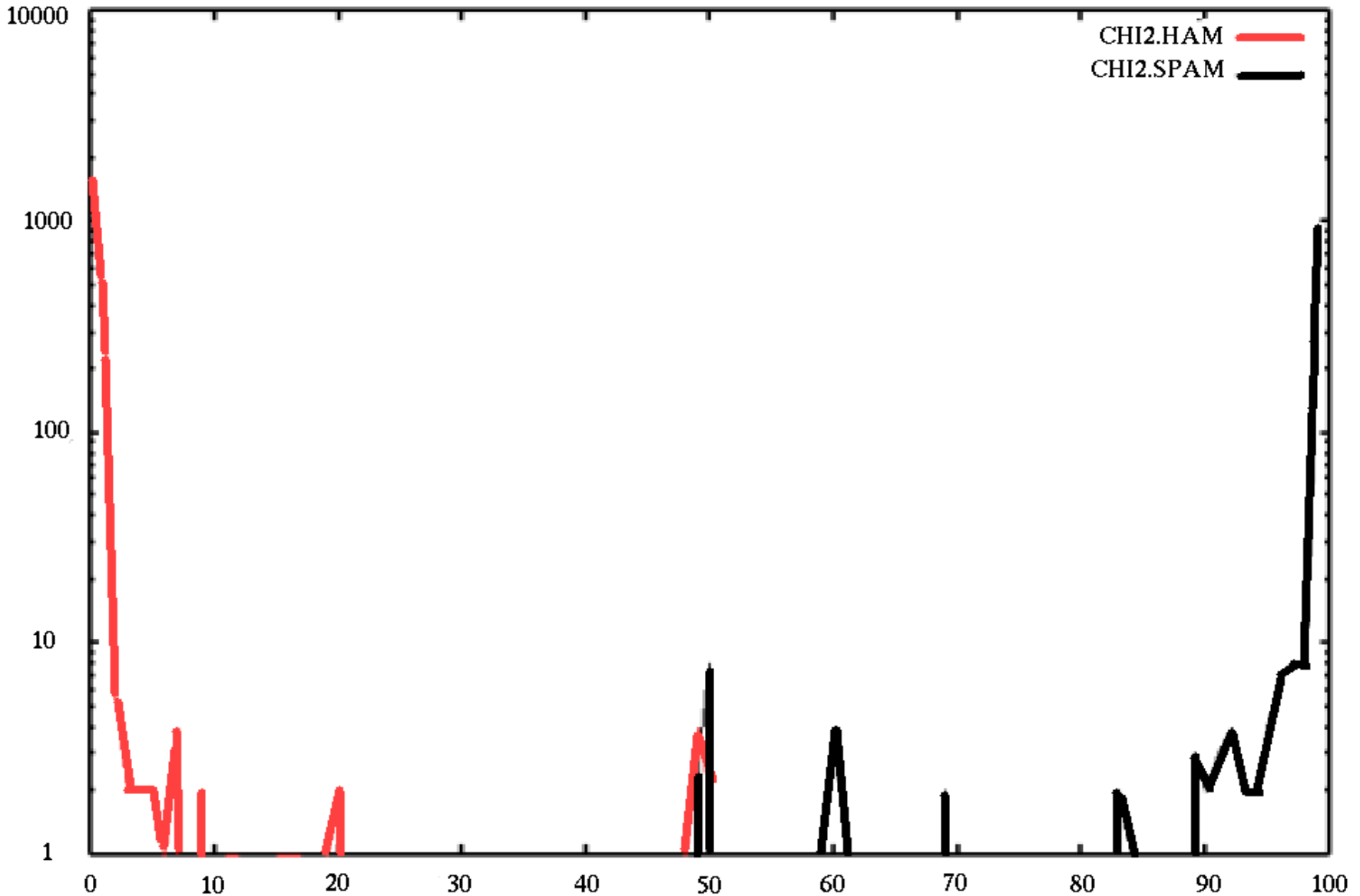


„Nehéz pontosan meghatározni
a határértéket . . .”



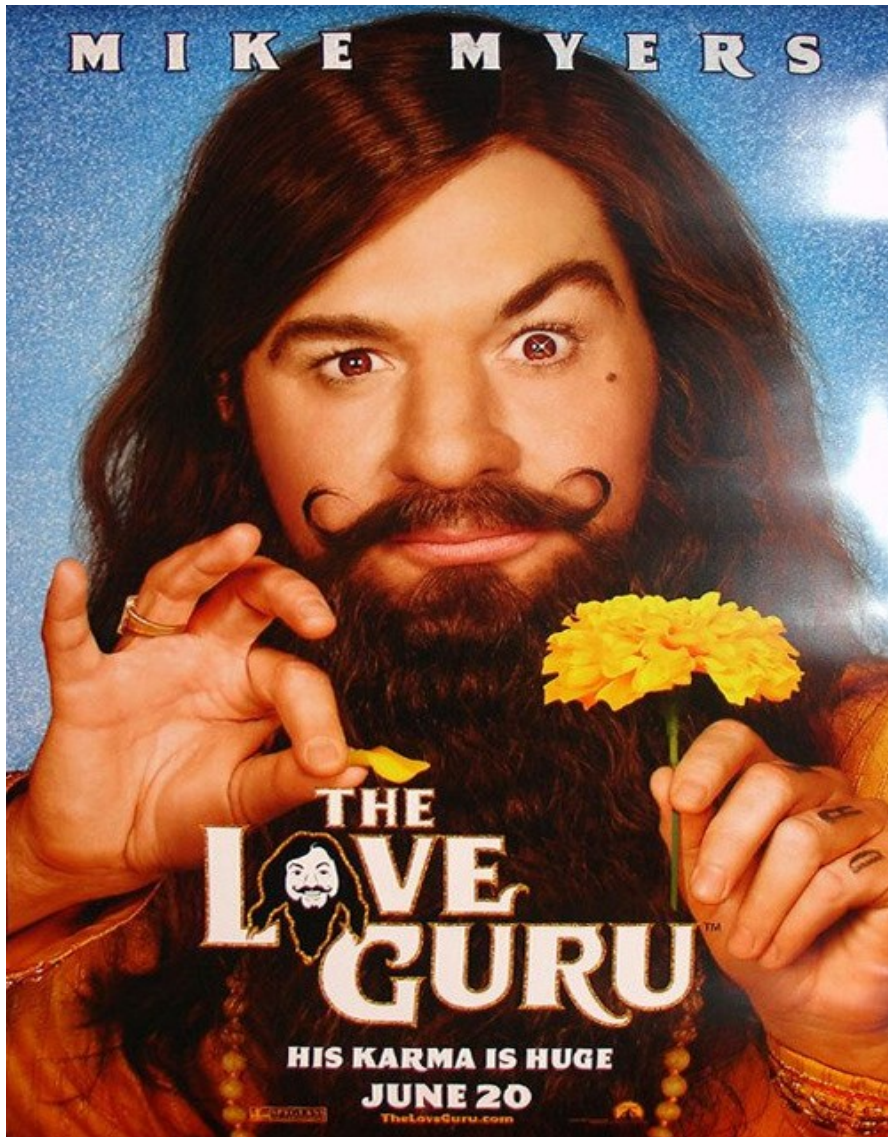
“...Ha túl magas, akkor kevés spamet ismer fel, ha túl alacsony, akkor sok jó levelet is elveszítünk.”

Valószínűség eloszlás



Forrás: http://spambayes.sourceforge.net/images/chi2_graph.png

A szakértő megmondja



„Egy embernél jól működhet, de 10000-nél már nem”

„... később komoly csalódás éri a téves biztonságérzetű felhasználót”

„A spammerek legjobb trükkjeit nem ismeri fel”

„A Bayes-alapú
technológiák
megbízhatósága
kétséges, két-
ségbe vonható”



Mikro/piko spam

Return-Path: <vpwwcomrigvri@yahoo.com>

X-Original-To: sj@xxxx.hu

Received: from av-engine (localhost [127.0.0.1])
by xx.xxxx.hu (Postfix) with SMTP id E04CC17013
for <sj@xxxx.hu>; Mon, 21 Jul 2008 04:38:58

Received: from 194.xx.xx.xx (unknown [58.248.77.145])
by xx.xxxx.hu (Postfix) with SMTP id ABD1617012
for <sj@xxxx.hu>; Mon, 21 Jul 2008 04:38:57

Received: from 22.185.62.188 by Sun, 20 Jul 2008 21:35:59

Message-ID: <C[20

Date: Mon, 21 Jul 2008 04:38:57 +0200 (CEST)

From: vpwwcomrigvri@yahoo.com

To: undisclosed-recipients::;

FROM*yahoo.com	0.9507	1
NO_SUBJECT*	0.9999	1
HEADER*undisclosed-recipients	0.8764	1

level84: **0.9995** in 68 [ms]

Received: from **mx2.xxxx.hu** (mx2.xxxx.hu [**195.xx.xx.xx**])

Received: from [200.158.63.250
(200-158-63-250.dsl.telesp.net.br [200.158.63.250])

Received: from [200.158.63.250] by
mail2.newfashionproducts.com; Wed, 23 Jul
Hora oficial do Brasil

From: <nuyicjfry@bmount.com>

To: <sj@xxx.hu>

Subject: Vulcan!

Date: Wed, 23 Jul Hora oficial do Brasil

Content-Type: text/plain;
format=flowed;
charset="**Windows-1252**";
reply-type=original

Content-Transfer-Encoding: 7bit

For bad boys only

<http://ingenuitycopy.com/?said=r17>

HEADER*windows-1252 0.9285 1
HEADER*mx2.xxxx.hu 0.9974 1
HEADER*195.xx.xx.xx 0.9972 1

level90: **0.9999** in 30 [ms]

ingenuitycopy.com.multi.surbl.org has address 127.0.0.86
ingenuitycopy.com.multi.uribl.com has address 127.0.0.2

SURBL0*ingenuitycopy.com 0.9999 1

Received: from **mx2.xxxx.hu** (mx2.xxxx.hu[**195.xx.xx.xx**])
by xx.xxx.hu (Postfix) with SMTP id C3F7917018
for <sj@xxxx.hu>; Tue, 15 Jul 2008 10:57:23 +0200

Received: from av-engine (localhost [127.0.0.1])
by mx2.xxxx.hu (Postfix) with SMTP id C1C3642A0C6
for <sj@xxxx.hu>; Tue, 15 Jul 2008 10:57:23 +0200

Received: from host1-111-dynamic.17-87-
r.retail.telecomitalia.it (host1-111-dynamic.17-87-
r.retail.telecomitalia.it [87.17.111.1])
by mx2.xxxx.hu (Postfix) with SMTP id BC2D1429EA9
for <sj@xxxx.hu>; Tue, 15 Jul 2008 10:57:22 +0200

Message-ID: <487C6670.6010903@acts.hu>

Date: Tue, 15 Jul 2008 10:57:20 +0100

From: "Anne" <m@acts.hu>

To: "Vickie" <sj@xxxx.hu>

Subject: Style **casino**

Win, win **with** us - **our** casino **<http://casdream.net/>**

HEADER*mx2 .xxxx .hu	0.9973	1
HEADER*195 .xx .xx .xx	0.9971	1
FROM*acts .hu	0.8960	1
Subject*casino	0.9955	1
with+our	0.9181	1

level65: **1.0000** in 41 [ms]

Received: from 65.204.40.211 (unknown [65.204.40.211])
by xx.xxx.hu (Postfix) with SMTP id83F6C17018

Message-ID: <000601c8e6a4\$053bc22c\$c1d44282@bjoawxy>

From: "kalle sandgorg" <art@bonne-route.com>

To: <xxx@xxxx.hu>

Subject: We **caught you naked** in the **shower** xxxx

MIME-Version: 1.0

Content-Type: text/plain;
charset="iso-8859-1"

Content-Transfer-Encoding: 7bit

X-Priority: 3

X-MSMail-Priority: Normal

X-Mailer: Microsoft Outlook Express 6.00.2900.3138

X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2900.3198

WATCH: **<http://autodromomouras.com/view.exe>**

Subject*caught	0.9680	1
Subject*you	0.9843	1
Subject*shower	0.9747	1
Subject*naked	0.9791	1

level72: 0.9999 in 27 [ms]

Kreatív írás

Received: from abqr197.neoplus.adsl.tpnet.pl
(abqr197.neoplus.adsl.tpnet.pl [83.8.85.197])
Received: from 141.156.31.140 (HELO merlin.cssiinc.com)
by xx.xxxx.hu with ESMTP ({nChar[8-12]}nChar[4-6])
From: "Louis Hernandez" <Louis@cssiinc.com>
To: "Melvin King" <xx@xxxx.hu>
Subject: Get a rod of colossal measurements!

-----=_NextPart_6255_18D9_01C8ECCA.BFE87290
Content-Type: text/plain; charset="windows-1250"
Content-Transfer-Encoding: quoted-printable

I thought I could not do anything with this **small dimension until I tried this** remedy! Make it get bigger, go for it! <http://stillgen.com/>

once, I **didnt want** to hit it again.would tell him, **You dont** look Jewish. Nothing made him feel happierones.

FROM*louis	0.9314	1
FROM*hernandez	0.9771	1
HEADER*nchar	0.9946	1
windows-1250+meta	0.9051	1
Subject*get	0.9595	1
Subject*rod	0.9040	1
small+dimension	0.9040	1
made+him	0.9926	1
tried+this	0.9047	1
it!+font	0.9952	1
didnt+want	0.9942	1
until+tried	0.9952	1
you+dont	0.9958	1
mshtml+6.00.3790.1830	0.9952	1
6.00.3790.1830+name	0.9952	1

level93: **1.0000** in 33 [ms]

Adobe + „eltört” URL

Received: from (ipbf.tokyo.ocn.ne.jp [118.6.117.90])
From: "boote kirby" <1kazman@jaring.my>
To: <sj@xxxx.hu>
Subject: to sj

Get the **great discounts** on popular software today !

Windows XP Pro With SP2 - \$59.95

Adobe Acrobat Pro 8 - \$69.95

Office 2003 Pro - **\$59.95**

Adobe Photoshop CS2 - \$79.95

Microsoft Office 2004 for MAC **\$79.95**

Adobe Acrobat 7 Professional for MAC \$59.95

Adobe **Creative Suite 2 Premium for MAC** \$229.95

- Visit **our site:** [www.rahwosoft\[DOT\]com](http://www.rahwosoft[DOT]com)
(**copy this** link and then **replace** "[DOT]" to ".")

great+discounts 0.9952 1
adobe+acrobat 0.9952 1
adobe+creative 0.9954 1
suite+premium 0.9952 1
creative+suite 0.9952 1
microsoft+office 0.9570 1
\$59.95+adobe 0.9952 1
\$79.95+adobe 0.9957 1
NO_SUBJECT* 0.9999 1
adobe+photoshop 0.9958 1
acrobat+professional 0.9932 1
photoshop+cs2 0.9952 1
office+pro 0.9948 1
premium+for 0.9952 1
...
our+site 0.9027 1
com+copy 0.9631 1
copy+this 0.9680 1
replace+dot 0.9631 1

level66: 1.0000 in 41 [ms]

Image spam

Received: from x.x.x.x.pldt.net (unknown [210.213.113.173])
From: =?koi8-r?B?8i7nLiDz1NLVxdc=?= <whitney@apr.com>
To: =?koi8-r?B?4c3exc7Dxdc=?= <xxx@xxxx.hu>
Subject: =?koi8-r?B?NyDU28suIOFxMXU08vPQ0sxc7JxSDP1CDv0sLJ?=-

<META http-equiv=Content-Type content="text/html;
charset=koi8-r">

<DIV align=center><**img**
=src="cid:000401c8\$4b27d4bb@mohjueg" align=middle>

=C6=D2=C1=CE=DB=C9=DA=CE=CF=C7=CF=20=**D0=C1=CB=C5=D4=C1 sp;**
«ORBY&**raquo;**; :
C9=DE=C5=CE=CE=CF=C5 =D0=D2=C5=C4=CC=CF=D6=C5=CE=C9=C5 =

Content-Type: image/jpeg; name="orbi0.jpg"
Content-Transfer-Encoding: base64

Subject*koi8-r	0.9962	1
charset+koi8-r	0.9918	1
koi8-r+content-transfer-encoding	0.9915	1
koi8-r+meta	0.9993	1
div+bgcolor	0.9460	1
bgcolor+div	0.8803	1
raquo+font	0.0743	1
jjjjj+nbsp	0.9680	1
nbsp+jjj	0.9040	1
color+jjj	0.9040	1
jjj+jjjjj	0.9721	1
generator+style	0.8783	1
raquo+nbsp	0.9718	1
color+img	0.9761	1
strong+nbsp	0.9467	1
src+cid	0.8929	1
image+jpeg	0.8960	1
jpeg+content-transfer-encoding	0.9121	1
content-transfer-encoding+base64	0.8750	1
IMAGE*	0.9999	1
EMBED*	0.9999	1

level70: 1.0000 in 51 [ms]

Received: from mary (24-183-185-151.charter.com
[24.183.185.151])

Received: from [24.183.185.151] by **f.mx.mail.yahoo.com;**

From: "Carey Brunson"@xx.xxxx.hu

To: <xx@xxxx.hu>

Subject: Great variety of little helpers for your health.

<szöveges rész az elején>

-----=_NextPart_001_000F_01C8E6D2.031FDE80

Content-Type: text/html; charset="iso-8859-2"

Content-Transfer-Encoding: quoted-printable

<p>

CLICK HERE URL!!!

-----=_NextPart_000_000E_01C8E6D2.031FDE80

Content-Type: image/gif; name="10.gif"

Content-Transfer-Encoding: base64

HEADER*f.mx.mail.yahoo.com	0.9631	1
FROM*carey	0.9870	1
Subject*great	0.9964	1
Subject*little	0.9955	1
Subject*helpers	0.9631	1
Subject*your	0.8795	1
Subject*health.	0.9834	1
Subject*for+Subject*your	0.8975	1
center+bordercolor	0.9954	1
mshtml+6.00.2900.2670	0.9957	1
body+body	0.9961	1
body+table	0.9436	1
here+font	0.9149	1
border+href	0.8995	1
6.00.2900.2670+name	0.9957	1
content-type+meta	0.8783	1
head+title	0.9251	1
div+align	0.8783	1
size+click	0.9299	1
IMAGE*	0.9999	1

level97: **1.0000** in 33 [ms]

„A töltelékszavak lerontják a felismerés hatékonyságát ...”



“...Spamre nem jellemző szavakat tesznek bele, így fog átjutni a spamszűrőn. Ha tanítod, akkor mérgezi az adatbázist, sokkal rosszabbul fog működni.”

Received: from **dsldevice.lan** (x.x.x.x.orange.es[85.60.30.163])
Subject: Give freedom to the **desires**

<szöveges rész a lenti szósalátával>

Content-Type: text/html; charset="**Windows-1252**"
Content-Transfer-Encoding: quoted-printable

```
<META http-equiv=Content-Type content="text/html;
charset=Windows-1252">
<META content=3D"MSHTML 6.00.2720.1081" name=GENERATOR>
<STYLE></STYLE></HEAD>
<BODY bgColor=#ffffff><DIV align=center><FONT face=Comic Sans
MS size=3>A better way to give up smoking.</FONT></DIV><DIV
align=center><FONT face=Comic Sans size=3>Your new source of
great health.</FONT></DIV>

<DIV><FONT face=Comic Sans MS size=3></FONT>&nbsp;&nbsp;&nbsp;
</DIV><DIV align=center><FONT face=Comic Sans size=2><A
href="http://q.bay.livefilestore.com/y1py...9A/gqoet.htm">
Here!</A></FONT></DIV><BR><BR><BR><BR><BR>
```

exporatory euroaquilo equilivent esthesises entogenous
eurobridge ethermeter fBTERMINFO enterozoan februaryius

HEADER*dsldevice.lan	0.9944	1
FROM*charles	0.9845	1
Subject*give	0.9953	1
Subject*freedom	0.8816	1
Subject*desires	0.9902	1
windows-1252+meta	0.9981	1
charset+windows-1252	0.9799	1
windows-1252+content-transfer-encoding	0.9716	1
generator+style	0.8786	1
here!+font	0.9958	1
size+better	0.9952	1
size+your	0.8784	1
div+align	0.8776	1
quoted-printable+the	0.9769	1
bgcolor+div	0.8806	1
you+dreamt	0.9893	1
which+one	0.9963	1
health.+font	0.9808	1
URL*livefilestore.com	0.9771	1

level73: 1.0000 in 27 [ms]

Received: from x-x-x-x.aaa.com.br (unknown[200.138.161.196])
From: "Mckenzie Lavinder" <wwinflation@dialusformurder.com>
To: "xxx" <xxx@xxxx.hu>
Subject: Actually you do not try them?

Content-Type: text/plain; **charset="iso-8859-15"**
<szöveges rész összefüggéstelen szavakból álló mondatokkal>

<IMG alt="As sunken" hspace=0
src="**http://cufbtg.bay.livefilestore.com/y1pBw...g/jskd.jpg**"
align=baseline border=0></DIV>

<DIV align=left>**But**
sold? Is by consequence indistinguishable. Is therefore It
viewer. For counteract. At to broaden. on by supplementary
seriously sentiment. avoid at twist. my what, dangerous the
pipes.</DIV>

<DIV align=left>**Which**
disappear. He is ignorant, repose. you timetable faculty.
That easy it welch appellant. by go prescribed broken
misguided. mistress the mindless. prophet social a hive. To
be violate master.</DIV>

Subject*actually	0.9863	1
Subject*you	0.9843	2
Subject*not	0.9956	2
Subject*try	0.9952	1
Subject*them	0.9924	1
charset+iso-8859-15	0.9808	1
align+baseline	0.9808	1
border+div	0.9416	1
URL*livefilestore.com	0.9771	1
hspace=0+src	0.9480	1
div+align	0.8776	1
bgcolor+div	0.8810	1
generator+style	0.8786	1
baseline+border	0.9838	1
size+but	0.9952	1
size+which	0.9528	1

level78: **1.0000** in 58 [ms]

Magyar nyelvű spam

Received:from **3lyw4** (x.x.x.x.vodafone.hu[**89.223.208.212**])
From: =?ISO-8859-1?Q?**Anikó?**= <balogh.aniko@vipmail.hu>
Subject: Nyerj LCD TV-t!
To: sj@xxxx.hu
Content-Type: text/plain; charset="ISO-8859-1"
Reply-To: balogh.aniko@vipmail.hu

Hello!

Szívesen néznéd Te is a **filmeket** egy nagyképernyős LCD TV-n, de nincs rá pénzed? Akkor nyerj egyet! Csak **egy sms-t** kell elküldened, és máris jó esélyed **van** rá, **hogyan** jelképes áron vásárolhasd meg álmaid nagyképernyős plazmatv-jét! Nézd meg ezt az oldalt, **megéri!**

<http://www.matutinyersz.eu/>

FROM*vipmail.hu	0.9877	1	
FROM*anikó	0.9631	1	
HEADER*3lyw4	0.9888	1	
egy+sms-t	0.9631	1	
van+hogy	0.0202	1	phrase: 0.8139

RBL checking: 212.208.223.89.zen.spamhaus.org
surbl check for matutinyersz.eu (0) took 0 ms

FROM*vipmail.hu	0.9877	2	
FROM*anikó	0.9631	2	
HEADER*3lyw4	0.9888	2	
egy+sms-t	0.9631	1	
filmeket	0.9718	1	
van+hogy	0.0202	1	
van	0.1005	1	
sms-t	0.9631	1	
hogy	0.1005	1	mix: 0.8703
megéri!	0.9631	1	caught by rbl
RBL0*89.223.208.212	0.9999	1	

level76: 0.9797 in 28 [ms]

Received: from mx2.xxxx.hu (mx2.xxxx.hu [195.xx.xx.xx])
Received: from 192.168.1.1(x.x.x.x.pool.tvnet.hu[85.238.82.84])
From: "Molnár Zsolt" <molnarzsolt2007@gmail.com>
To: <xx@xxxx.hu>
Subject: Címlista

Tisztelt Hölgyem/ Uram!

Megrendelhető nálam az alábbi email címlista:

- 200.000 **magyarországi vállalkozás email** címe, tetszés szerinti csoportosításban (összesítve, illetve számos kategória külön is).
- 400.000 **magyarországi magán** email cím.

Továbbá el tudom vállalni az email kampány lebonyolítását (kiküldés), valamint tudok ajánlani email címeket gyűjtő, valamint hírlevél kiküldő programot.

Valamennyi címlistát idén, 2008. januárjában állítottam össze, a hibaszázalék mindössze 5-10 százalék, vagyis aktuálisabb, mint szinte valamennyi más **hasonló adatbázis**.

Üdvözlettel, Molnár Zsolt

Subject*címlista	0.9564	1
uram!+megrendelhető	0.9564	1
megrendelhető+nálam	0.9564	1
részletes+tájékoztatót	0.9564	1
email+címlista	0.9564	1
valamennyi+címlistát	0.9564	1
tudok+ajánlani	0.9040	1
küldök+illetve	0.9564	1
email+címem	0.9564	1
vállalkozás+email	0.9564	1
magyarországi+magán	0.9564	1
bármilyen+felmerülő	0.9564	1
alábbi+email	0.9564	1
tájékoztatót+küldök	0.9564	1
illetve+válaszolok	0.9564	1
hasonló+adatbázis.	0.9564	1
válaszolok+bármilyen	0.9564	1
számos+kategória	0.9564	1
címlista+200.000	0.9564	1
magyarországi+vállalkozás	0.9564	1
Üdvözlettel+molnár	0.9564	1
molnár+zsolt	0.9564	1

Received: from **mobilier1** (dslXXXX.pool.t-online.hu [78.92.82.147])

From: Kontra =?ISO-8859-1?Q?György?= <flexen2@freemail.hu>

To: <xx@xxxx.hu>

Subject: =?ISO-8859-1?Q?Bemutatózás?=>

X-DCC--Metrics: mail02a.mail.t-online.hu 0; Body=1 Fuz1=1 Fuz2=1

<html><head><title>Tisztelt Hölgyem</title></head><body>

<p align="left">Tisztelt Hölgyem/Uram!
Kérem **engedje meg hogy, bemutatkozzunk** pár másodperc alatt.
 Tekintse meg

szórólapunkat, és ha felkeltettük érdeklődését tekintse meg web oldalunkat és keressen fel személyesen vagy telefonon. </p>

Tekintse meg aktuális akciónkat!</p><p align="left">Köszönjük a türelmét!

Üdvözlettel: Kontra György

Az Ön e-mail címe valamely nyilvános, mindenki által elérhető adatbázisból származik. Ez az **e-mail nem minősül spamnek, hanem a 2001. évi CVIII. Törvény** 14. §-ban előírt "hozzájárulás" **kérés** az informáláshoz. Amennyiben a továbbiakban nem kíván ilyen üzenetet kapni, kérjük, jelezze ezt egy válasz e-mailben. Ha levelünkkel zavartuk, elnézést kérünk.

HEADER*78.92.82.147	0.9631	1
HEADER*mobilier1	0.9631	1
FROM*Kontra =?ISO-8859-1?Q?Gy=F6rgy?= <flexen2@freemail.hu>	0.9631	1
Subject*bemutakozás	0.9631	1
microsoft+frontpage	0.9953	1
frontpage.editor.document+meta	0.9631	1
content+frontpage.editor.document	0.9952	1
content-language+content	0.9499	1
title+tisztelt	0.9631	1
keressen+fel	0.9631	1
engedje+meg	0.9822	1
hogy+bemutakozzunk	0.9631	1
oldalunkat+href	0.9631	1
windows-1250+title	0.9631	1
URL*mobilier.hu	0.9631	1
URL*freeweb.hu	0.9631	1
kérjük+jelezze	0.9040	1
alatt.+tekintse	0.9631	1
jelezze+ezt	0.9631	1
web+oldalunkat	0.9631	1
e-mail+nem	0.9631	1
spamnek+hanem	0.9631	1
hanem+2001.	0.9845	1
2001.+évi	0.9845	1
évi+cviii.	0.9845	1
cviii.+törvény	0.9747	1
törvény+14.	0.9747	1

Bayes mérgezés



Date: Tue, 15 Jul 2008 10:57:20 +0100
From: "Anne" <m@teszt.hu>
To: "Vickie" <sj@xxxx.hu>
X-Abcde: ajaja ahah aua akakaka akaka
Subject: Re: attack on spam filters

I agree with you!

**ajaj ajaju azaha
azazazz akak**

**<!-- jaja ajajaj ajah uzauaz u azuua
zuazauahaha -->**

Hogyan lehet legyőzni a statisztikai spamszűrőket?



Ha nem hasznárod . . .

