

11/10/09
12:44:14

spam-exception.txt

1

Here's a spam that got through my spam classifier. Class discussion on why...

```
etern-Path: <micheall01@mail2world.com>
X-Original-To: dwallach@cs.rice.edu
Received: from mail.cs.rice.edu (localhost.localdomain [127.0.0.1])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 0FCB52C2B02
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 08:48:01 -0600 (CST)
X-Virus-Scanned: by amavis-2.4.0 at mail.cs.rice.edu
X-Spam-Flag: YES
X-Spam-Score: 9.406
X-Spam-Level: *****
X-Spam-Status: Yes, score=9.406 tagged_above=-999 required=6.31
    tests=[DNS_FROM_RFC_POST=1.44, DNS_FROM_RFC_WHOIS=0.879,
    FORGED_MUA_OUTLOOK=3.36, MAILTO_TO_SPAM_ADDR=0.276,
    RCVD_IN_BL_SPAMCOP_NET=1.332, RCVD_IN_SORBS_WEB=1.236,
    UNDISC_RECIPS=0.883]
Received: from mail.cs.rice.edu ([127.0.0.1])
    by mail.cs.rice.edu (mail.cs.rice.edu [127.0.0.1]) (amavisd-new, port 10024)
    with LMTP id iirS1tWvJI1k for <dwallach@cs.rice.edu>;
    Tue, 10 Nov 2009 08:47:51 -0600 (CST)
X-policyd-weight: DYN_PBL_SPAMHAUS=SKIP(0) SBL_XBL_SPAMHAUS=SKIP(-1.5) NOT_IN_SPAMCOP=-1.5 NOT_IN_BL_NJABL=-1.5 CL_IP_EQ_H
ELO_IP=-2 (check from: .mail2world. - helo: .mxo5.broadbandsupport. - helo-domain: .broadbandsupport.) FROM/MX_MATCHES_NO
T_HELO(DOMAIN)=1 IN_PM_RFCI=3.2; rate: -2.3
Received: from mxo5.broadbandsupport.net (mxo5.broadbandsupport.net [209.55.3.85])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 057342C2D70
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 08:47:47 -0600 (CST)
Authentication-Results: mxo5.broadbandsupport.net smtp.user=watterson@cebridge.net; auth=pass (LOGIN)
Received: from [210.51.12.185] ([210.51.12.185:1065] helo=User)
    by mxo5.broadbandsupport.net (ecelerity 2.1.1.23 r(18304)) with ESMTTPA
    id 93/19-23695-99C79FA4 for <dwiddle@brookes.ac.uk>; Tue, 10 Nov 2009 09:47:44 -0500
Message-ID: <93.19.23695.99C79FA4@mxo5.broadbandsupport.net>
Reply-To: <micheal01@mail2world.com>
From: "Micheal Patel" <micheall01@mail2world.com>
Subject: EARN NO LESS THAN 200.00 British Pounds.
Date: Tue, 10 Nov 2009 14:47:40 -0000
MIME-Version: 1.0
Content-Type: text/plain;
    charset="Windows-1251"
Content-Transfer-Encoding: 7bit
X-Priority: 3
X-MSMail-Priority: Normal
X-Mailer: Microsoft Outlook Express 6.00.2600.0000
X-MimeOLE: Produced By Microsoft MimeOLE V6.00.2600.0000
To: undisclosed-recipients;
```

X-DSPAM-Result: Innocent
X-DSPAM-Processed: Tue Nov 10 08:48:01 2009
X-DSPAM-Confidence: 0.5389
X-DSPAM-Probability: 0.0000
X-DSPAM-Signature: 4af97d2134392064717839
X-DSPAM-Factors: 27,
Subject*Pounds, 0.99990,
Earn+up, 0.00010,
the+stated, 0.00010,
a+Mystery, 0.99990,
me+explain, 0.00010,
No+experience, 0.99990,
Authentication-Results*smtp, 0.00010,
Mystery+Shoppers, 0.99990,
Subject*British+Pounds, 0.99990,
Authentication-Results*auth+pass, 0.00020,
hotels+for, 0.00020,
Authentication-Results*smtp+user, 0.00020,
Micheal, 0.00020,
Great+Pay, 0.99980,
unbiased, 0.00020,
Authentication-Results*user, 0.00020,
Authentication-Results*auth, 0.00020,
restaurant+and, 0.00020,
Authentication-Results*pass+LOGIN, 0.00020,
Schedules, 0.00020,
Stay+at, 0.00020,
like+being, 0.00020,
mall, 0.00020,
standpoint+This, 0.00020,
Authentication-Results*LOGIN, 0.00020,
Businesses, 0.00020,
free+travel, 0.99980

Need extra INCOME!
EARN NO LESS THAN 200.00 British Pounds

Become our [MYSTERY SHOPPER]:
Earn [NO LESS THAN 200.00 British Pounds] Per Venture:
It is Very Easy and Very Simple:
No Application fees:

Here's your chance to get paid for shopping and dining out.

Your job will be to evaluate and comment on customer service in a wide Variety of shops, Stores, restaurant and services i

n your area.

Mystery shoppers are needed throughout the UK. You'll be paid to shop and dine out-plus, you can also get free meals, free merchandise, free services, free entertainment, free travel and more. Great Pay. Fun Work. Flexible Schedules.

No experience required.

If you can shop-you are qualified!
Let me explain to you what it entails
What is a mystery shopper?

A mystery shopper is like being 007 at the mall. Mystery shoppers must complete their assignments and go un-detected.

Mystery Shoppers receive assignments consisting of a variety of business types such as: restaurants, food chains, automobile dealerships, retail stores, and more.

Businesses benefit from the services of a mystery shopper because they report their unbiased review from a customer's standpoint.

This enables the business to identify problems that could result in a unhappy customer and loss of sales.

As you can see, mystery shoppers have a great responsibility and are paid accordingly.

As a Mystery Shopper you can:

Earn up to 200 British Pounds per assignment have a flexible schedule and hours pick and choose what assignments you want and bring your kids with you on assignments.

Work part time or full time.
Get free merchandise
Get paid for dining out
Stay at hotels for free
Get paid for going to the movies
And much more...

What you need to do is to get back to me with the stated details below:

Your Full Name:
Residential Address:
City:
Post code:
Age:
Phone Number(s):
Banking Institution:

11/10/09
12:44:14

spam-exception.txt

4

We are looking forward to read back from you.

Micheal Patel.

Mystery Shopper Inc.
micheal01@mail2world.com

11/10/09
12:45:45

spam-exception2.txt

1

Here's another exception to the rules. Let's analyze.

```
Return-Path: <>
X-Original-To: dwallach@cs.rice.edu
Received: from mail.cs.rice.edu (localhost.localdomain [127.0.0.1])
    by mail.cs.rice.edu (Postfix) with ESMTTP id A12062C2D70
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 09:56:37 -0600 (CST)
X-Virus-Scanned: by amavis-2.4.0 at mail.cs.rice.edu
X-Spam-Score: -
X-Spam-Level:
X-Spam-Status: No, score=x tagged_above=-999 required=6.31 WHITELISTED
    tests=[]
Received: from mail.cs.rice.edu ([127.0.0.1])
    by mail.cs.rice.edu (mail.cs.rice.edu [127.0.0.1]) (amavisd-new, port 10024)
    with LMTP id yx3zY5E3HAPy for <dwallach@cs.rice.edu>;
    Tue, 10 Nov 2009 09:56:30 -0600 (CST)
Received: from mh2.mail.rice.edu (mh2.mail.rice.edu [128.42.201.21])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 04BC62C2B02
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 09:56:30 -0600 (CST)
Received: from mh2.mail.rice.edu (localhost.localdomain [127.0.0.1])
    by mh2.mail.rice.edu (Postfix) with ESMTTP id D7CF8CC004
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 09:56:29 -0600 (CST)
X-Virus-Scanned: by amavis-2.6.0 at mh2.mail.rice.edu
Received: from mh2.mail.rice.edu ([127.0.0.1])
    by mh2.mail.rice.edu (mh2.mail.rice.edu [127.0.0.1]) (amavisd-new, port 10024)
    with ESMTTP id GR8I99daXgNs for <dwallach@cs.rice.edu>;
    Tue, 10 Nov 2009 09:56:29 -0600 (CST)
X-Greylist: from auto-whitelisted by SQLgrey-1.7.6
Received: from 60-241-219-81.static.tpgi.com.au (60-241-219-81.static.tpgi.com.au [60.241.219.81])
    by mh2.mail.rice.edu (Postfix) with SMTP id B07CFCC001
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 09:56:28 -0600 (CST)
Received: from ghr ([40.133.76.63])
    by 60-241-219-81.static.tpgi.com.au (8.13.2/8.13.2) with SMTP id 200911110155011825;
    Wed, 11 Nov 2009 01:55:00 +1000
Message-ID: <20091111015350.8030605@areva.com>
Date: Wed, 11 Nov 2009 01:53:50 +1000
From: "Graham Lawson" <tim.parker@areva.com>
User-Agent: Thunderbird 2.0.0.9 (Windows/20071031)
MIME-Version: 1.0
To: dwallach@cs.rice.edu
Subject: Enjoy the prices and save with us
Content-Type: text/plain; charset=ISO-8859-1; format=flowed
Content-Transfer-Encoding: 7bit
X-DSPAM-Result: Innocent
```

X-DSPAM-Processed: Tue Nov 10 09:56:37 2009
X-DSPAM-Confidence: 0.5577
X-DSPAM-Probability: 0.2181
X-DSPAM-Signature: 4af98d3562471958816358
X-DSPAM-Factors: 27,
 Subject*with+us, 0.99990,
 Subject*prices+and, 0.99990,
 Subject*Enjoy+the, 0.99990,
 User-Agent*0+0, 0.04323,
 User-Agent*Thunderbird+2, 0.04347,
 User-Agent*2+0, 0.04356,
 User-Agent*2, 0.05091,
 Content-Type*1+format, 0.05183,
 Subject*prices, 0.94448,
 X-Spam-Status*31+WHITELISTED, 0.05748,
 User-Agent*0, 0.05749,
 User-Agent*0, 0.05749,
 X-Spam-Status*x, 0.05911,
 X-Spam-Status*score+x, 0.05924,
 X-Spam-Status*x+tagged, 0.05924,
 Content-Type*flowed, 0.06063,
 Content-Type*format, 0.06125,
 Content-Type*format+flowed, 0.06126,
 Content-Type*ISO+8859, 0.06375,
 User-Agent*Thunderbird, 0.06762,
 X-Spam-Status*WHITELISTED, 0.07498,
 Content-Type*ISO, 0.07792,
 Content-Type*charset+ISO, 0.07793,
 Subject*Enjoy, 0.91897,
 User-Agent*Windows, 0.08147,
 X-Spam-Status*No+score, 0.09988,
 *Tue+Nov, 0.10026

Become the best man of her bedroom. <http://zukd.steadbits.com/>

11/10/09
13:25:29

spam-ham.txt

1

This example shows classification working well. I've only exchanged a very few emails with this person, but look at the low coefficients.

```
Return-Path: <BrendaWhitney@kivahome.com>
X-Original-To: dwallach@cs.rice.edu
Received: from mail.cs.rice.edu (localhost.localdomain [127.0.0.1])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 106EF2C2D42
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 12:12:58 -0600 (CST)
X-Virus-Scanned: by amavis-2.4.0 at mail.cs.rice.edu
X-Spam-Score: 0.253
X-Spam-Level:
X-Spam-Status: No, score=0.253 tagged_above=-999 required=6.31
    tests=[AWL=0.252, HTML_MESSAGE=0.001]
Received: from mail.cs.rice.edu ([127.0.0.1])
    by mail.cs.rice.edu (mail.cs.rice.edu [127.0.0.1]) (amavisd-new, port 10024)
    with LMTP id LSy9UTFBQJbs for <dwallach@cs.rice.edu>;
    Tue, 10 Nov 2009 12:12:49 -0600 (CST)
X-policyd-weight: DYN_PBL_SPAMHAUS=SKIP(0) SBL_XBL_SPAMHAUS=SKIP(-1.5) NOT_IN_SPAMCOP=-1.5 NOT_IN_BL_NJABL=-1.5 CL_IP_EQ_H
ELO_IP=-2 (check from: .kivahome. - helo: .owa.kivahome. - helo-domain: .kivahome.) FROM/MX_MATCHES_HELO(DOMAIN)=-2; rate
: -8.5
Received: from owa.kivahome.com (smtp.kivahome.com [66.60.233.229])
    (using TLSv1 with cipher RC4-MD5 (128/128 bits))
    (No client certificate requested)
    by mail.cs.rice.edu (Postfix) with ESMTTP id 1D4282C2C55
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 12:12:45 -0600 (CST)
Received: from kkb-hou-exc2.Kiva.Local ([172.16.1.8]) by
    kkb-hou-exc2.Kiva.Local ([172.16.1.8]) with mapi; Tue, 10 Nov 2009 12:12:45
    -0600
From: Brenda Whitney <BrendaWhitney@kivahome.com>
To: 'Dan Wallach' <dwallach@cs.rice.edu>
Date: Tue, 10 Nov 2009 12:12:44 -0600
Subject: Still working on it..
Thread-Topic: Still working on it..
Thread-Index: AcpiMWbcimztKix3TQi8znCIV6SA2g==
Message-ID: <A028C002DE27AB4A8B2CD65603480B2563D1A0E59C@kkb-hou-exc2.Kiva.Local>
Accept-Language: en-US
Content-Language: en-US
X-MS-Has-Attach:
X-MS-TNEF-Correlator:
acceptlanguage: en-US
Content-Type: multipart/alternative;
    boundary="_000_A028C002DE27AB4A8B2CD65603480B2563D1A0E59Ckkbhhouexc2Kiv_"
MIME-Version: 1.0
X-DSPAM-Result: Whitelisted
X-DSPAM-Processed: Tue Nov 10 12:12:58 2009
```

X-DSPAM-Confidence: 0.9997
X-DSPAM-Probability: 0.0000
X-DSPAM-Signature: 4af9ad2a127891364714457
X-DSPAM-Factors: 27,
 Whitney+Builder, 0.00010,
 Message-ID*Kiva+Local, 0.00010,
 From*Whitney+BrendaWhitney, 0.00010,
 cursive, 0.00010,
 Message-ID*Kiva, 0.00010,
 Message-ID*kkb, 0.00010,
 direct+font, 0.00010,
 Message-ID*Local, 0.00010,
 X-policyd-weight*owa+kivahome, 0.00010,
 713+266, 0.00010,
 713+266, 0.00010,
 Received*smtp.kivahome.com, 0.00010,
 X-policyd-weight*from+kivahome, 0.00010,
 Return-Path*kivahome+com, 0.00010,
 X-policyd-weight*kivahome+helo, 0.00010,
 X-policyd-weight*kivahome+helo, 0.00010,
 X-policyd-weight*helo+owa, 0.00010,
 Message-ID*exc2+Kiva, 0.00010,
 4328, 0.00010,
 4328, 0.00010,
 Message-ID*hou, 0.00010,
 From*Whitney, 0.00010,
 Received*kkb-hou-exc2.Kiva.Local, 0.00010,
 Received*kkb-hou-exc2.Kiva.Local, 0.00010,
 From*kivahome+com, 0.00010,
 *kivahome+com, 0.00010,
 cursive+size, 0.00010

--_000_A028C002DE27AB4A8B2CD65603480B2563D1A0E59Ckkbhouexc2Kiv_
Content-Type: text/plain; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

Brenda Whitney
Builder Sales
713-219-4328 (direct)
713-266-4646 (fax)
Kiva Kitchen & Bath
7071 Southwest Frwy

Houston, Tx. 77074

```
--_000_A028C002DE27AB4A8B2CD65603480B2563D1A0E59Ckkbhouexc2Kiv_
Content-Type: text/html; charset="us-ascii"
Content-Transfer-Encoding: quoted-printable

<html>
<head>
<meta http-equiv=3D"Content-Type" content=3D"text/html; charset=3Dus-ascii"=
>
<meta name=3D"Generator" content=3D"Microsoft Exchange Server">
<!-- converted from rtf -->
<style><!-- .EmailQuote { margin-left: 1pt; padding-left: 4pt; border-left:=
#800000 2px solid; } --></style>
</head>
<body>
<font face=3D"Arial, sans-serif" size=3D"3">
<div>&nbsp;</div>
<div>&nbsp;</div>
<div><font face=3D"Mistral, cursive" size=3D"1" color=3D"#000080">&nbsp;</f=
ont></div>
<div><font size=3D"2" color=3D"#000080">Brenda Whitney</font></div>
<div><font size=3D"2" color=3D"#000080">Builder Sales</font></div>
<div><font size=3D"1" color=3D"#008080"><b>713-219-4328 (direct)</b></font>=
</div>
<div><font size=3D"1" color=3D"#008080"><b>713-266-4646 (fax)</b></font></d=
iv>
<div><font face=3D"Lucida Console, monospace" color=3D"#008080"><b>Kiva Kit=
chen & Bath</b></font></div>
<div><font face=3D"Lucida Console, monospace" color=3D"#008080"><b>7071 Sou=
thwest Frwy</b></font></div>
<div><font face=3D"Lucida Console, monospace" color=3D"#008080"><b>Houston,=
Tx. 77074</b></font></div>
<div><font face=3D"Lucida Console, monospace" color=3D"#008000">&nbsp;</fon=
t></div>
<div><font face=3D"Lucida Console, monospace" color=3D"#008000">&nbsp;</fon=
t></div>
<div><font face=3D"Lucida Console, monospace" color=3D"#008080">&nbsp;</fon=
t></div>
```

11/10/09
13:25:29

spam-ham.txt

4

```
<div><font face=3D"Verdana, sans-serif" size=3D"1" color=3D"#008080">&nbsp;=  
</font></div>  
<div><font face=3D"Verdana, sans-serif" size=3D"1">&nbsp;</font></div>  
<div><font size=3D"2">&nbsp;</font></div>  
</font>  
</body>  
</html>
```

```
--_000_A028C002DE27AB4A8B2CD65603480B2563D1A0E59Ckkbhouexc2Kiv_--
```

11/10/09
13:28:18

spam-ham2.txt

1

This example is somebody with whom I've never corresponded. This is the first email from them to me, in response to a message I sent to them.

```
Return-Path: <academic@soyatec.com>
X-Original-To: dwallach@cs.rice.edu
Received: from mail.cs.rice.edu (localhost.localdomain [127.0.0.1])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 8841A2C2D4A
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 10:35:29 -0600 (CST)
X-Virus-Scanned: by amavis-2.4.0 at mail.cs.rice.edu
X-Spam-Score: 2.017
X-Spam-Level: **
X-Spam-Status: No, score=2.017 tagged_above=-999 required=6.31
    tests=[FROM_BLANK_NAME=1.467, NO_REAL_NAME=0.55]
Received: from mail.cs.rice.edu ([127.0.0.1])
    by mail.cs.rice.edu (mail.cs.rice.edu [127.0.0.1]) (amavisd-new, port 10024)
    with LMTP id hrrE9OxJvgax for <dwallach@cs.rice.edu>;
    Tue, 10 Nov 2009 10:35:17 -0600 (CST)
X-policyd-weight: DYN_PBL_SPAMHAUS=SKIP(0) SBL_XBL_SPAMHAUS=ERR(-1.5) NOT_IN_SPAMCOP=-1.5 NOT_IN_BL_NJABL=-1.5 CL_IP_EQ_HE
LO_IP=-2 (check from: .soyatec. - helo: .echo.dreamhost. - helo-domain: .dreamhost.) FROM/MX_MATCHES_HELO(DOMAIN)=-2; rat
e: -8.5
Received: from echo.dreamhost.com (echo.dreamhost.com [64.111.116.119])
    by mail.cs.rice.edu (Postfix) with ESMTTP id 060802C2D74
    for <dwallach@cs.rice.edu>; Tue, 10 Nov 2009 10:35:07 -0600 (CST)
Received: by echo.dreamhost.com (Postfix, from userid 0)
    id D491459152; Tue, 10 Nov 2009 08:35:02 -0800 (PST)
From: "" <academic@soyatec.com>
To: dwallach@cs.rice.edu
Precedence: bulk
Subject: Re: Academic license request
Content-Type: text/plain;
    charset="iso-8859-1"
Content-Transfer-Encoding: 8bit
Message-Id: <20091110163502.D491459152@echo.dreamhost.com>
Date: Tue, 10 Nov 2009 08:35:02 -0800 (PST)
X-DSPAM-Result: Innocent
X-DSPAM-Processed: Tue Nov 10 10:35:29 2009
X-DSPAM-Confidence: 0.9741
X-DSPAM-Probability: 0.0000
X-DSPAM-Signature: 4af9965183374691112619
X-DSPAM-Factors: 27,
    Incomplete, 0.00010,
    request+on, 0.00010,
    Subject*license, 0.00010,
    X-policyd-weight*dreamhost, 0.00020,
    X-policyd-weight*dreamhost, 0.00020,
```

X-policyd-weight*domain+dreamhost, 0.00020,
X-policyd-weight*dreamhost+FROM, 0.00020,
an+auto, 0.00020,
X-policyd-weight*dreamhost+helo, 0.00020,
indicated+on, 0.00571,
Precedence*bulk, 0.01287,
be+ignored, 0.01442,
Date*PST, 0.01653,
the+Academic, 0.01715,
Date*0800+PST, 0.01782,
X-Spam-Score*017, 0.03287,
X-Spam-Status*017+tagged, 0.03323,
X-Spam-Status*017, 0.03934,
provided+the, 0.04881,
Subject*Re, 0.05299,
Please+make, 0.06028,
email+We, 0.06255,
Subject*Academic, 0.07413,
X-policyd-weight*rate+8, 0.07939,
on+this, 0.07961,
requests, 0.07973,
X-policyd-weight*8+5, 0.08151

This is an auto-response email. Please don't reply this email.

We have received your request on the Academic license. Please make sure you have provided the necessary information indicated on this page: <http://www.soyatec.com/euml2/order/academic.php>

Incomplete requests will be ignored. We'll process your request shortly.