Erster Anti-Spam-Kongress „trifft ins Schwarze“

http://www.heise.de/newsticker/data/hob-22.05.03-000/

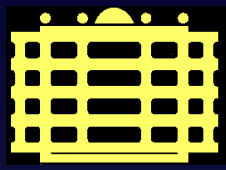
Man habe mit der Veranstaltung „voll ins Schwarze getroffen“, freute sich Sven Karge vom Verband der deutschen Internetwirtschaft eco.

. . .

In ihren Vorträgen signalisierten Vertreter von E-Mail-Providern Bereitschaft, im Kampf gegen unerwünschte Mails künftig kooperieren zu wollen. Malte Pollmann von Lycos Europe schlug beispielsweise vor, dass die Anbieter ihre Blacklists untereinander austauschen könnten, um die Effizienz der Filter zu erhöhen.

Es wurde aber auch klar, dass die konkurrierenden Provider wirksame Spam-Filter mittlerweile als Vermarktungsargument für ihre Dienste entdeckt haben. . . .

Malte Pollmann von Lycos berichtete, dass etwa 60 Prozent aller eingehenden E-Mails auf den Lycos-Servern unerwünscht seien. . . .



TECHNISCHE UNIVERSITÄT CHEMNITZ

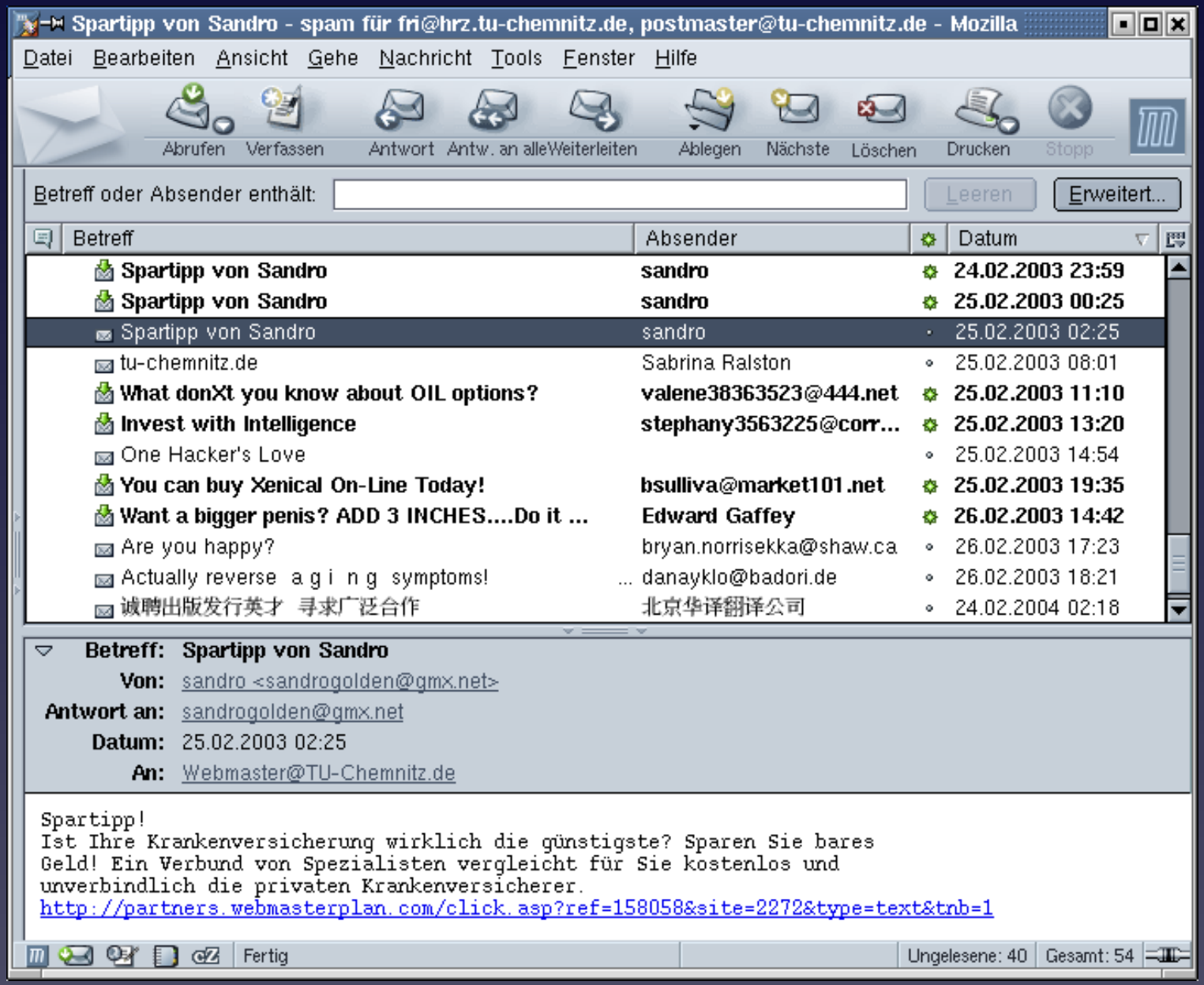
Universitätsrechenzentrum & Fakultät für Informatik

E-Mail für Dich – Lust oder Frust?
Wie Spam das Medium der Zukunft gefährdet.

Frank Richter & Ralph Sontag

Unix-Stammtisch
27. Mai 2003

Das Problem


Spartipp von Sandro - spam für fri@hrz.tu-chemnitz.de, postmaster@tu-chemnitz.de - Mozilla
Datei Bearbeiten Ansicht Gehe Nachricht Tools Fenster Hilfe
Abrufen Verfassen Antwort Antw. an alle Weiterleiten Ablegen Nächste Löschen Drucken Stopp
Betreff oder Absender enthält:
Leeren
Erweitert...
Betreff
Absender
Datum
Spartipp von Sandro sandro 24.02.2003 23:59
Spartipp von Sandro sandro 25.02.2003 00:25
Spartipp von Sandro sandro 25.02.2003 02:25
tu-chemnitz.de Sabrina Ralston 25.02.2003 08:01
What donXt you know about OIL options? valene38363523@444.net 25.02.2003 11:10
Invest with Intelligence stephany3563225@corr... 25.02.2003 13:20
One Hacker's Love 25.02.2003 14:54
You can buy Xenical On-Line Today! bsulliva@market101.net 25.02.2003 19:35
Want a bigger penis? ADD 3 INCHES....Do it ... Edward Gaffey 26.02.2003 14:42
Are you happy? bryan.norrisekka@shaw.ca 26.02.2003 17:23
Actually reverse a g i n g symptoms! ... danayklo@badori.de 26.02.2003 18:21
诚聘出版发行英才 寻求广泛合作 北京华译翻译公司 24.02.2004 02:18
Betreff: Spartipp von Sandro
Von: sandro <sandrogolden@gmx.net>
Antwort an: sandrogolden@gmx.net
Datum: 25.02.2003 02:25
An: Webmaster@TU-Chemnitz.de
Spartipp!
Ist Ihre Krankenversicherung wirklich die günstigste? Sparen Sie bares
Geld! Ein Verbund von Spezialisten vergleicht für Sie kostenlos und
unverbindlich die privaten Krankenversicherer.
http://partners.webmasterplan.com/click.asp?ref=158058&site=2272&type=text&tnb=1
Fertig
Ungelesene: 40
Gesamt: 54

Spam = unerwünschte E-Mail / Ham = erwünschte Mail

- junk Mail – UBE – UCE
- sehr subjektiv - Spam != Werbung

Spam = unerwünschte E-Mail / Ham = erwünschte Mail

- junk Mail – UBE – UCE
- sehr subjektiv - Spam != Werbung

Wo ist das Problem?

Mißbrauch von Ressourcen:

Bandbreite, Plattenplatz, CPU – verloren für eigentliche Anwendungen

Zeit: Empfänger, Administratoren

Spam = unerwünschte E-Mail / Ham = erwünschte Mail

- junk Mail – UBE – UCE
- sehr subjektiv - Spam != Werbung

Wo ist das Problem?

Mißbrauch von Ressourcen:

Bandbreite, Plattenplatz, CPU – verloren für eigentliche Anwendungen

Zeit: Empfänger, Administratoren

Kosten beim Empfänger:

Studie der Europäischen Kommission 01/2001:

„Junk-E-Mails kosten die Internetnutzer weltweit jährlich 10 Mrd. €“

⇨ Theft of Service, Dienstqualität leidet, Kostenerhöhung

Verunsicherung der Nutzer und Admins, Nachahmungen

zerstört den Geist des Internet: Mailing-Listen geschlossen, Diensteinschränkungen

⇨ ernste Gefahr für die Nutzbarkeit des E-Mail-Dienstes

Modernes elektronisches Direktmarketing?
extrem niedrige Kosten beim Sender, Kosten beim Empfänger
deswegen meist ungezielte Massensendungen ⇨ offenbar erfolgreich

Wieviel Spam?

1997	AT& T, Microsoft	5 – 15 %
2003	TU Chemnitz	> 30 %
2003	Prognose Sommer	> 50 %

Geschichte

1975	RFC 706 „On the Junk Mail Problem“, Jon Postel, 1975
80er	Kettenbriefe
4/1994	„Greencard Lottery“ Canter & Siegel posten in 6000 Newsgroups
1995	Spam King – Spam als „Service“ gefälschter Absender, temporäre ISP-Accounts, via offene Relays
1996	„Spamford“ Wallace – cyberpromo.com
12/1997	CERT Summary „Relaying of Spam Email through Victim Sites“ – eine der häufigsten Formen von Computer-Attacken
bis heute	viele Nachahmer, Spam-Provider, Programme Adressensammler: News, WWW, IRC, Mailing Listen, erraten

Inhalte

1. Anbieten von Spam-Techniken
2. Pornographische Websites, Telefon-Sex
3. Dialer und anderer Betrug
4. Pyramiden-Schemas: „Make money fast …“
5. Dubiose Produkte (Haarwuchsmittel, Schlankheitspillen …)
6. Politische, religiöse Fanatiker

Spam – Beispiel

From: UmaBg92d7@excite.com
DATE: 02 Mar 01 9:41:00 PM
TO: members2796327936_2786237sgsyu378352_287@mail.com
SUBJECT: Targeted Bulk Emailing Service

Targeted Direct Emailing Service

Introducing One Of The Most Effective Direct Email Marketing Firms On The Internet:
Reach Millions Of Targeted Prospects and Fast!

Visit: http://3506561034/tr_ad
Call Toll Free: 1-877-529-7358

To be removed from this list send an email to remyou2@mail.com with your email address in the subject.

Was tun?

Ignorieren . . . Fluchen . . . Hilfe suchen . . . Hoffen auf Besserung . . . Resignieren

Selber handeln:

- Abmelden: `remove@...` oder „Hier Klicken“ – ACHTUNG FALLE!
- Beschweren beim Absender: meist sinnlos, da gefälscht
- Filter des Mailprogrammes:
  `if (Subject enthält „sex“) then delete`
  ⇨ die Grenzen sind schnell erreicht

Filter-Werkzeuge mit Regelsätzen, z.B.
SpamAssassin (SA): `http://spamassassin.org/`

SpamAssassin (SA)

- Untersucht Header auf typische Spam-Markierungen und Fehler
- Spam-Sprache im Inhalt: `H O T!!!`, `Klicken Sie hier`, `Buy now`
- Heuristiken: HTML, Encodings ...
- Errechnet „Spam-Score“: Zahl, „Je größer, desto Spam.“
- Markiert Mails als Spam (Header, Body)

Probleme:

- Ständige Pflege nötig, eigene Regeln schwierig
- Offenbar prüfen Spammer ihre Botschaften gegen SpamAssassin

Nutzersicht

Zentrale Filter? . . . Wenig Entscheidungsspielraum

Private Filter? . . . Regelverständnis? Regelmäßige Update? Abhängigkeit?

Einzige Chance: Das Ziel des Spammers direkt angreifen!

Spammer wollen etwas „vermarkten“:

- WWW-Seite
- Telefonnummer
- Mailadresse
- Produkt
- . . .

Nutzersicht

Zentrale Filter? . . . Wenig Entscheidungsspielraum

Private Filter? . . . Regelverständnis? Regelmäßige Update? Abhängigkeit?

Einzige Chance: Das Ziel des Spammers direkt angreifen!

Spammer wollen etwas „vermarkten“:

- WWW-Seite
- Telefonnummer
- Mailadresse
- Produkt
- . . .

Nutzersicht

Zentrale Filter? . . . Wenig Entscheidungsspielraum

Private Filter? . . . Regelverständnis? Regelmäßige Update? Abhängigkeit?

Einzige Chance: Das Ziel des Spammers direkt angreifen!

Spammer wollen etwas „vermarkten“:

- WWW-Seite
- Telefonnummer
- Mailadresse
- Produkt
- . . .

Nutzersicht

Zentrale Filter? . . . Wenig Entscheidungsspielraum

Private Filter? . . . Regelverständnis? Regelmäßige Update? Abhängigkeit?

Einzige Chance: Das Ziel des Spammers direkt angreifen!

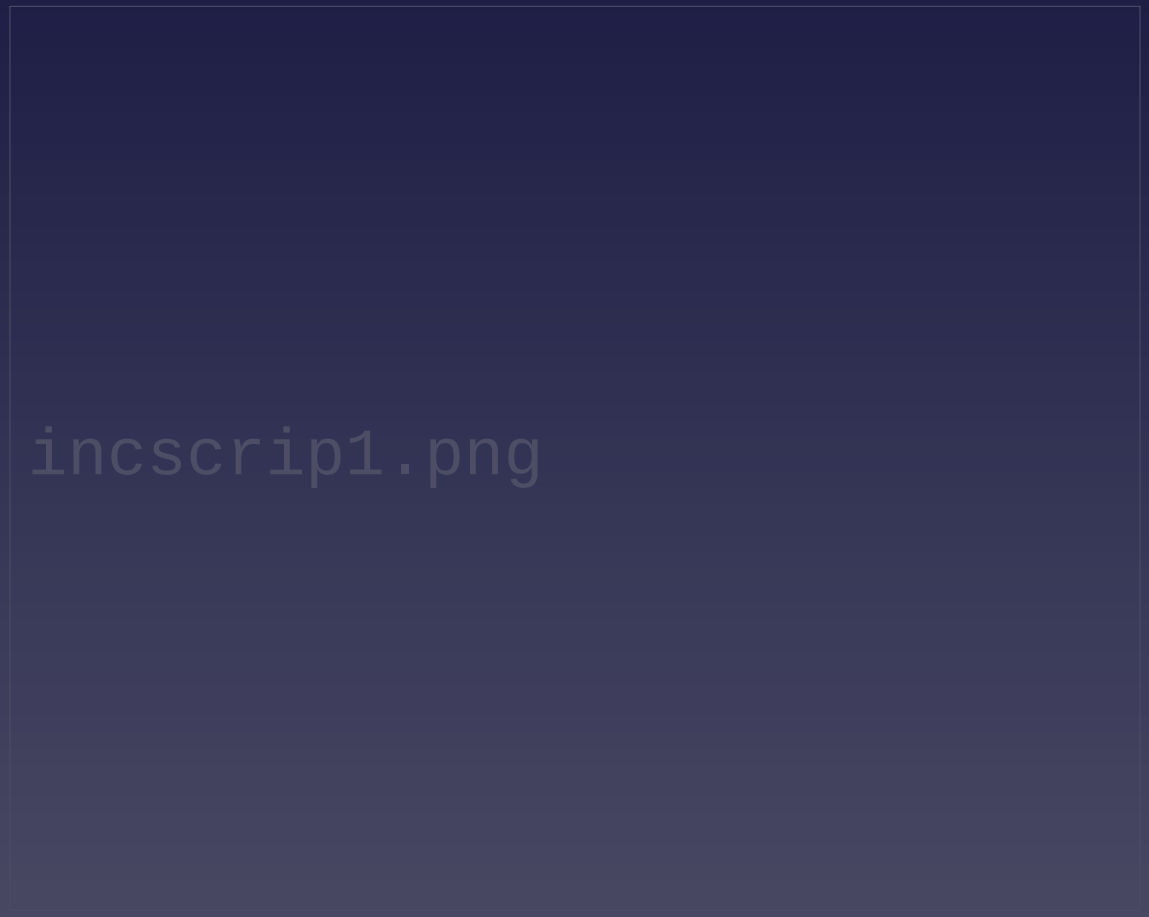
Spammer wollen etwas „vermarkten“:

- WWW-Seite
- Telefonnummer
- Mailadresse
- Produkt
- . . .

Irgendwann wird nur die „Nutzlast“ ein Kriterium liefern.

Wir erkennen Spam sofort. Kann das ein Programm lernen?

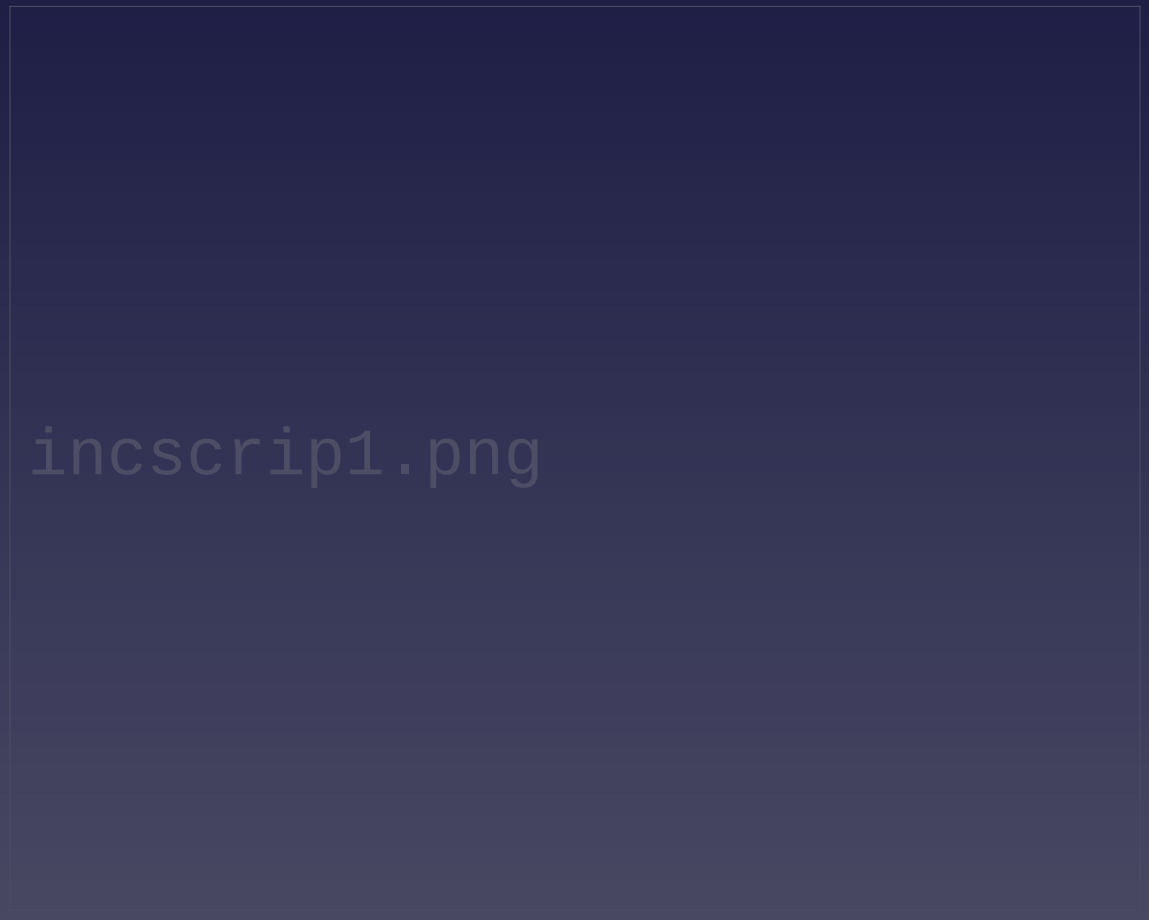
Wir gehen 250 Jahre zurück:



Irgendwann wird nur die „Nutzlast“ ein Kriterium liefern.

Wir erkennen Spam sofort. Kann das ein Programm lernen?

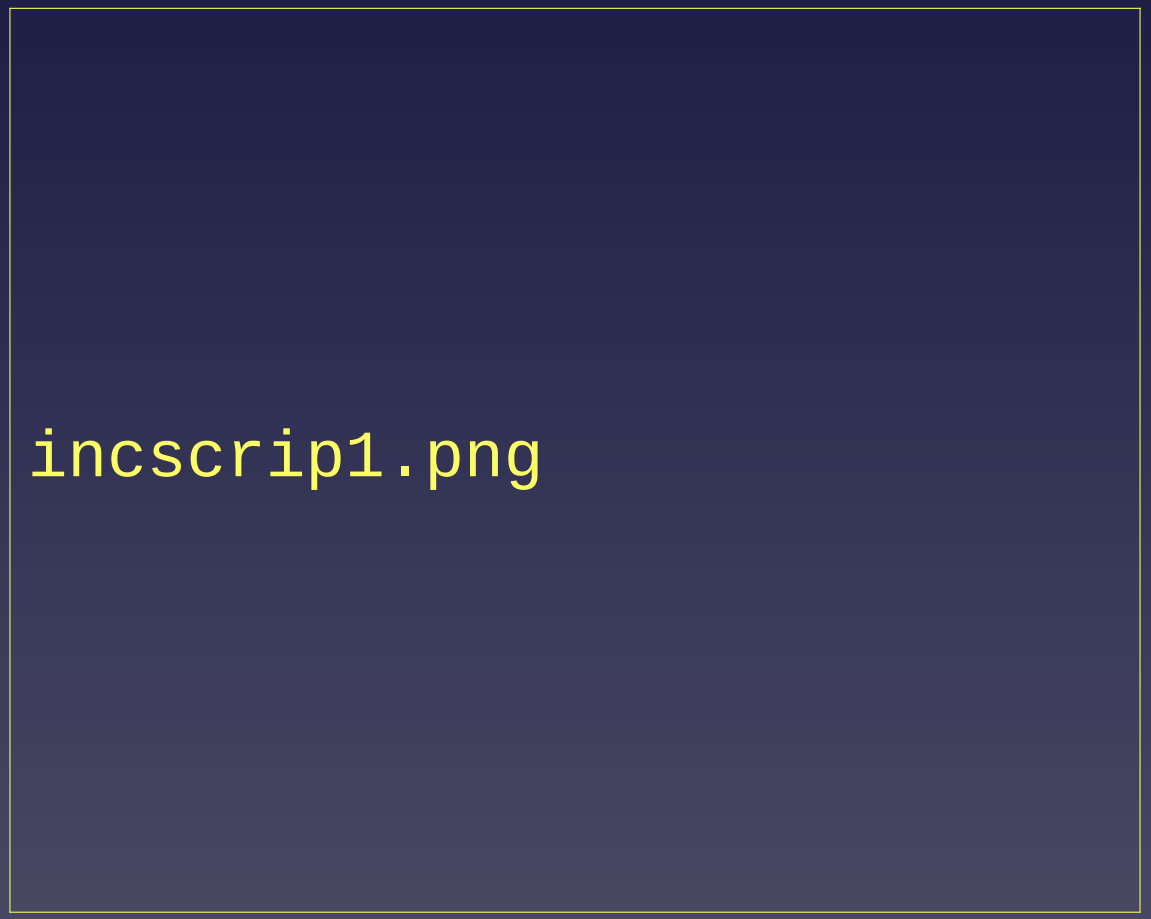
Wir gehen 250 Jahre zurück:



Irgendwann wird nur die „Nutzlast“ ein Kriterium liefern.

Wir erkennen Spam sofort. Kann das ein Programm lernen?

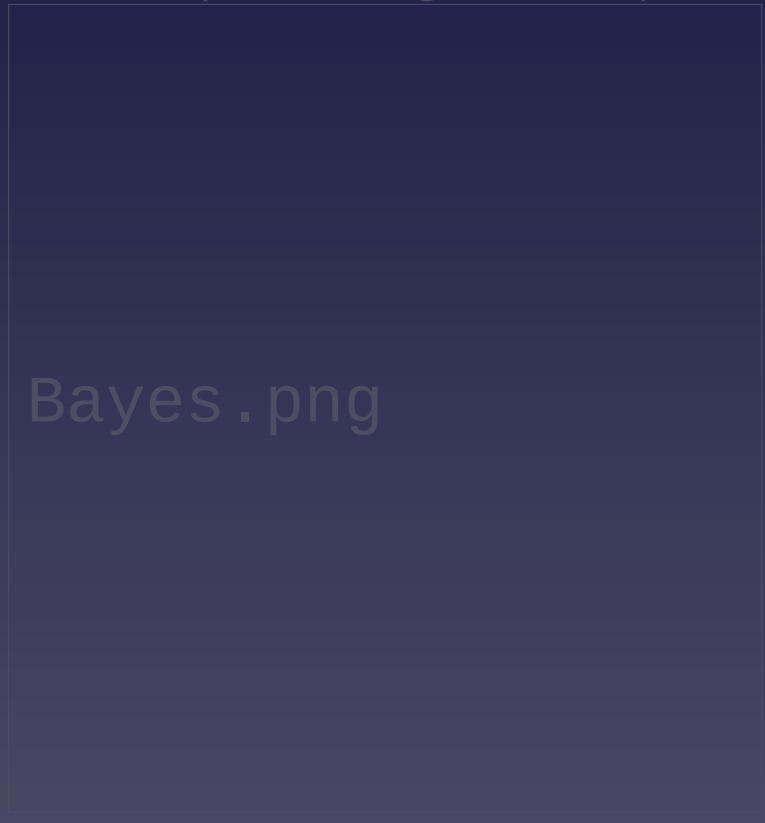
Wir gehen 250 Jahre zurück:



Kopf oder Zahl?

Klassische Wahrscheinlichkeitstheorie: $P = \frac{1}{2}$
Und nach acht Mal Zahl? ⇨ *Bayes*: Intuitiveres Herangehen.
Thomas Bayes: ✻ 1702 (London), ✝ 17. 4. 1761 (Tunbridge Wells)

In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.
Richard Price über Bayes' „Essay towards solving a problem in the doctrine of chances"



Kopf oder Zahl?

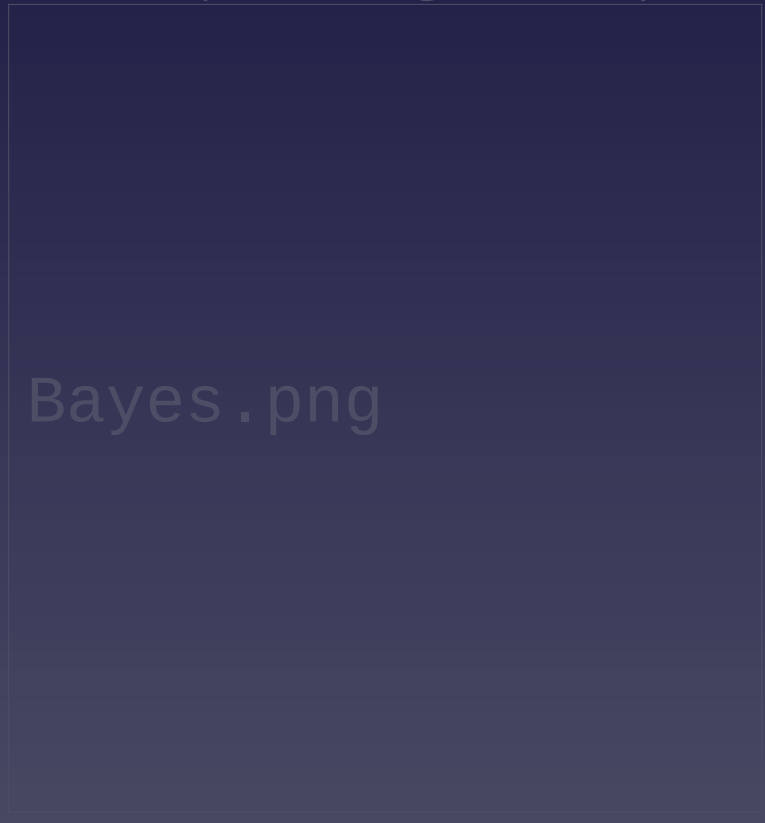
Klassische Wahrscheinlichkeitstheorie: $P = \frac{1}{2}$

Und nach acht Mal Zahl? ⇨ *Bayes*: Intuitiveres Herangehen.

Thomas Bayes: ✻ 1702 (London), ✝ 17. 4. 1761 (Tunbridge Wells)

In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

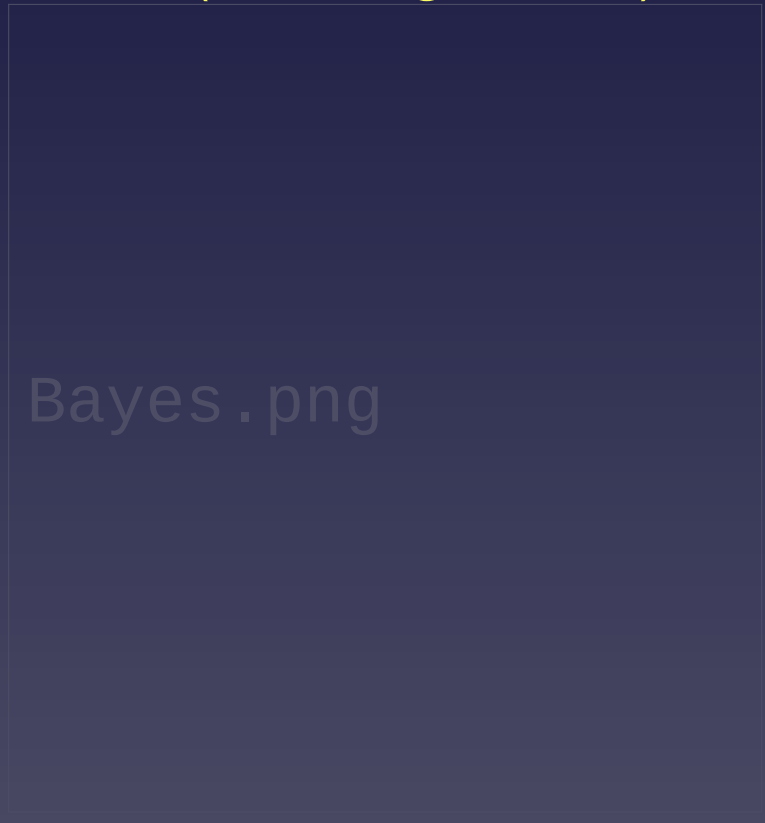
Richard Price über Bayes' „Essay towards solving a problem in the doctrine of chances"



Kopf oder Zahl?

Klassische Wahrscheinlichkeitstheorie: $P = \frac{1}{2}$
Und nach acht Mal Zahl? ⇨ *Bayes*: Intuitiveres Herangehen.
Thomas Bayes: ✻ 1702 (London), ✝ 17. 4. 1761 (Tunbridge Wells)

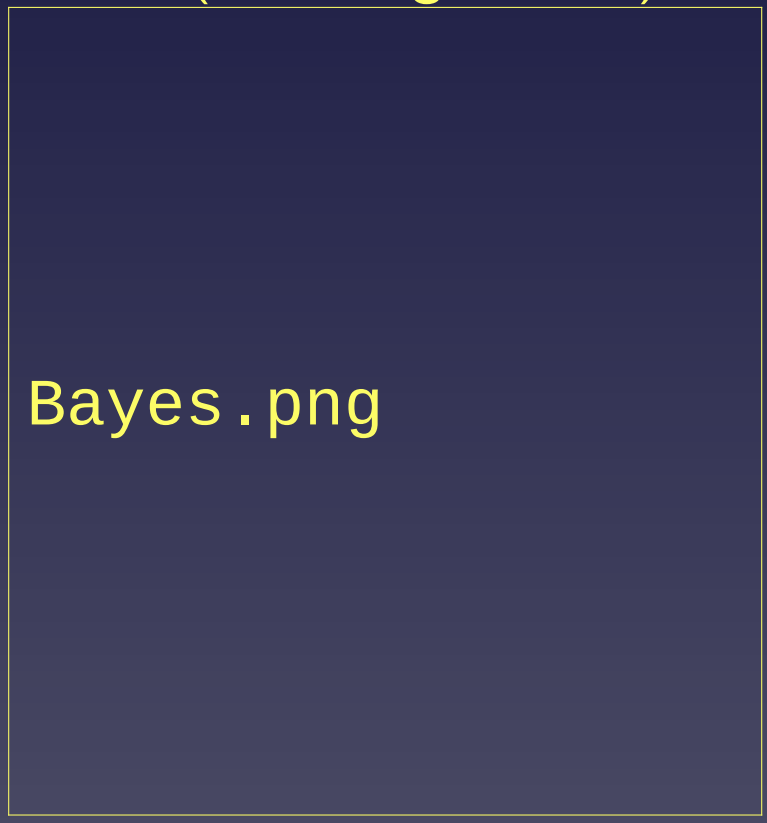
In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.
Richard Price über Bayes' „Essay towards solving a problem in the doctrine of chances"



Kopf oder Zahl?

Klassische Wahrscheinlichkeitstheorie: $P = \frac{1}{2}$
Und nach acht Mal Zahl? ⇨ *Bayes*: Intuitiveres Herangehen.
Thomas Bayes: ✻ 1702 (London), ✞ 17. 4. 1761 (Tunbridge Wells)

In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.
Richard Price über Bayes' „Essay towards solving a problem in the doctrine of chances"



Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \qquad \text{(I)}$$

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \qquad (I)$$

$P()$ Wahrscheinlichkeit

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \quad (I)$$

$P()$	Wahrscheinlichkeit	
H	Hypothese	Das ist eine gute Mail!

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \quad (1)$$

$P()$	Wahrscheinlichkeit	
H	Hypothese	Das ist eine gute Mail!
D	Daten, Ereignis	Die Mail enthält „Money“

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \quad (1)$$

$P()$	Wahrscheinlichkeit	
H	Hypothese	Das ist eine gute Mail!
D	Daten, Ereignis	Die Mail enthält „Money“
$P(H\|D)$	Bedingung	Wahrscheinlichkeit, daß die Mail gut ist, wenn sie „Money“ enthält.

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \quad (1)$$

$P()$	Wahrscheinlichkeit	
H	Hypothese	Das ist eine gute Mail!
D	Daten, Ereignis	Die Mail enthält „Money“
$P(H\|D)$	Bedingung	Wahrscheinlichkeit, daß die Mail gut ist, wenn sie „Money“ enthält.

Rechte Seite der Formel: Häufigkeiten als sinnvolle Annahmen für die Wahrscheinlichkeit

Der Satz von Bayes

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D)} \quad (I)$$

$P()$	Wahrscheinlichkeit	
H	Hypothese	Das ist eine gute Mail!
D	Daten, Ereignis	Die Mail enthält „Money“
$P(H\|D)$	Bedingung	Wahrscheinlichkeit, daß die Mail gut ist, wenn sie „Money“ enthält.

Rechte Seite der Formel: Häufigkeiten als sinnvolle Annahmen für die Wahrscheinlichkeit

. . . und praktisch?

Anmerkungen

- Unbekannte Wörter: Hier irrt Graham!

Anmerkungen

- Unbekannte Wörter: Hier irrt Graham!
- Auswerten von mehreren Werten $p_1, \ldots p_n$

Anmerkungen

- Unbekannte Wörter: Hier irrt Graham!
- Auswerten von mehreren Werten $p_1, \ldots p_n$

  Annahme: Unabhängigkeit (ein zweiter mathematischer Fehler :-))

Anmerkungen

- Unbekannte Wörter: Hier irrt Graham!
- Auswerten von mehreren Werten $p_1, \ldots p_n$

  Annahme: Unabhängigkeit (ein zweiter mathematischer Fehler :-))

  Beschränkung auf n (noch ein Fehler ...)

$$P = \frac{\prod_{i=1}^{n} p_i}{\prod_{i=1}^{n} (1 - p_i) + \prod_{i=1}^{n} p_i} \tag{2}$$

- Eindeutige Wörter ⇨ Wahrscheinlichkeit = 1.0 ⇨ Gefahr von Fehlern

Besser: Interpretation der Folder als Stichprobe in klassischer Manier ⇨ Maximalwert von 0.05 bzw. 0.95

Besser: Interpretation der Folder als Stichprobe in klassischer Manier ⇨ Maximalwert von 0.05 bzw. 0.95

Noch besser: Abhängig von Anzahl des Vorkommens Maximum festlegen!

Jedenfalls erhalten wir keine Wahrscheinlichkeit, sondern auch nur eine Maßzahl.

Der Nutzer muß konsistent entscheiden!

Wer heute als Spam wegwirft, was er gestern als Ham verdaute, trainiert einen Kalender aber keinen Filter.

`info.png`

Algorithmus: Training

1. E-Mail einordnen (Nutzer!)

Algorithmus: Training

1. E-Mail einordnen (Nutzer!)
2. E-Mail zerlegen

   Wie? (Domains trennen? Wörter trennen?)

Algorithmus: Training

1. E-Mail einordnen (Nutzer!)
2. E-Mail zerlegen

   Wie? (Domains trennen? Wörter trennen?)
3. Tabellen für Spam und Ham aufbauen, Tokens zählen

Algorithmus: Sortieren

1. E-Mail zerlegen

Algorithmus: Sortieren

1. E-Mail zerlegen

2. Wörter in Tabellen suchen

Algorithmus: Sortieren

1. E-Mail zerlegen
2. Wörter in Tabellen suchen
3. Vorschlag: n Extremwerte wählen. Graham: $n = 15$
   (Evtl. Wichtung nötig ⇨ häufige Wörter!)

Algorithmus: Sortieren

1. E-Mail zerlegen
2. Wörter in Tabellen suchen
3. Vorschlag: n Extremwerte wählen. Graham: $n = 15$
   (Evtl. Wichtung nötig ⇨ häufige Wörter!)
4. Wahrscheinlichkeit nach (2) berechnen

Algorithmus: Sortieren

1. E-Mail zerlegen
2. Wörter in Tabellen suchen
3. Vorschlag: n Extremwerte wählen. Graham: $n = 15$
   (Evtl. Wichtung nötig ⇨ häufige Wörter!)
4. Wahrscheinlichkeit nach (2) berechnen
5. Mail einsortieren (Graham: $P > 0.9 \Rightarrow$ Spam)

Algorithmus: Sortieren

1. E-Mail zerlegen
2. Wörter in Tabellen suchen
3. Vorschlag: n Extremwerte wählen. Graham: $n = 15$
   (Evtl. Wichtung nötig ⇨ häufige Wörter!)
4. Wahrscheinlichkeit nach (2) berechnen
5. Mail einsortieren (Graham: $P > 0.9 \Rightarrow$ Spam)
6. Filter justieren, evtl. Nutzerkorrektur beachten

Intermezzo

- Nicht nur Trennung Spam/Ham möglich
- Folder sollten veralten – Änderungen der Spammer auffangen!
- Vermeidung Fehler 1. Art? (Graham: Verdopplung der Wortzahl im Ham-Folder ⇨ ???)
- White-Lists unnötig, sparen aber Rechenzeit
- Mitlernendes System: „Hallo Ralph“ **war** mal eindeutig Ham
- Risiko: Wortkarge Freunde mit wechselnden Adressen:

> Hallo, bin unterwegs – schau mal auf http:// . . . – Gruß! R.

Software: Popfile

- `http://popfile.sourceforge.net/` Perl-Skript
- agiert als POP-Proxy: Popserver: *localhost*; Nutzer: *org-server:nutzer*
- Sortieren in beliebig viele gleichwertige „Buckets“ möglich
- Markierung durch Headerzeile oder im Subject

Software: Popfile

- http://popfile.sourceforge.net/ Perl-Skript
- agiert als POP-Proxy: Popserver: *localhost*; Nutzer: *org-server:nutzer*
- Sortieren in beliebig viele gleichwertige „Buckets“ möglich
- Markierung durch Headerzeile oder im Subject

Weitere Software:

- Bogofilter: http://bogofilter.sourceforge.net
- SpamBayes: http://spambayes.sourceforge.net
- CRM114: http://crm114.sourceforge.net
- Mozilla ab 1.3: http://www.mozilla.org/mailnews/spam.html
- SpamAssassin ab 2.50

Ausblick

Warum erst jetzt?

Ausblick

Warum erst jetzt? Zu viele falsche Positive, zu wenig Erfolg!?

Patrick Pantel and Dekang Lin. „SpamCop – A Spam Classification & Organization Program.“

Ausblick

Warum erst jetzt? Zu viele falsche Positive, zu wenig Erfolg!?

Patrick Pantel and Dekang Lin. „SpamCop – A Spam Classification & Organization Program.“

Ursachen:

- zu wenig Training
- nur Body ausgewertet
- Versuch, Wortstämme zu nutzen
- Auswertung aller Tokens, nicht nur der n signifikantesten ⇨ Fehler bei längerem Spam, leicht auszuhebeln
- keine reine Textanalyse – Email-Strukturen beachten
- Fehler nicht zufällig: Spammer sind aktiv!

Problem: HTML

```
<STRONG>D</  >o</  >n’</  >t be f</  >o</  >ole</  >d b</  >y</  >
</  >pr</  >od</  >uct</  >s </  >cl</  >ai</  >ming to b</  >e</  >
```

Ignorieren?

Problem: HTML

```
<STRONG>D</  >o</  >n'</  >t be f</  >o</  >ole</  >d b</  >y</  >
</  >pr</  >od</  >uct</  >s </  >cl</  >ai</  >ming to b</  >e</  >
```

Ignorieren? ⇨ Verzicht auf viele wertvolle Hinweise

Parsen?

Problem: HTML

```
<STRONG>D</  >o</  >n'</  >t be f</  >o</  >ole</  >d b</  >y</  >
</  >pr</  >od</  >uct</  >s </  >cl</  >ai</  >ming to b</  >e</  >
```

Ignorieren? ⇨ Verzicht auf viele wertvolle Hinweise

Parsen? ⇨ Filter wird zum HTML-Erkenner

Vielleicht nur bestimmte Tags verwerten: *img, bgcolor, font* . . .

Und weiter?

- Analyse von Wortpaaren ⇨ bessere Kontextsensitivität
- Zerlegung von Wörtern (xxxporn ⇨ xxx + porn)
- Unscharfe Erkennung: *M0ney*
- Tokentrennung tunen: Punkte und Kommas zwischen Ziffen nicht als Trenner auffassen: IP-Adressen, Preise . . .
- Headerzeilen extra behandeln: *Subject*Money*

Und weiter?

- Analyse von Wortpaaren ⇨ bessere Kontextsensitivität
- Zerlegung von Wörtern (xxxporn ⇨ xxx + porn)
- Unscharfe Erkennung: *M0ney*
- Tokentrennung tunen: Punkte und Kommas zwischen Ziffen nicht als Trenner auffassen: IP-Adressen, Preise ...
- Headerzeilen extra behandeln: *Subject*Money*

Subject*FREE	0.9999
free!!	0.9999
To*free	0.9998
Subject*free	0.9782
free!	0.9199
Free	0.9198
Url*free	0.9091
FREE	0.8747
From*free	0.7636
free	0.6546

Und weiter?

- Analyse von Wortpaaren ⇨ bessere Kontextsensitivität
- Zerlegung von Wörtern (xxxporn ⇨ xxx + porn)
- Unscharfe Erkennung: *M0ney*
- Tokentrennung tunen: Punkte und Kommas zwischen Ziffen nicht als Trenner auffassen: IP-Adressen, Preise …
- Headerzeilen extra behandeln: *Subject*Money*

Subject*FREE	0.9999
free!!	0.9999
To*free	0.9998
Subject*free	0.9782
free!	0.9199
Free	0.9198
Url*free	0.9091
FREE	0.8747
From*free	0.7636
free	0.6546

Degeneration: *Free!!!!!* ⇨ *Free!!* … *Free* ⇨ *free*

SMTP - Simple Mail Transfer Protocol (RFC 2821)

```
210.15.231.24 >>> tim.hrz.tu-chemnitz.de:25
<<< 220 tim.hrz.tu-chemnitz.de ESMTP Exim ...
>>> HELO dmpserver.dmp.com.au
<<< 250 tim.hrz.tu-chemnitz.de Hello [210.15.231.24]
>>> MAIL FROM:<UmaBg92d7@excite.com>
<<< 250 <UmaBg92d7@excite.com> is syntactically correct
>>> RCPT TO:<frank.richter@hrz.tu-chemnitz.de>
<<< 250 Ok
>>> DATA
<<< 354 Enter message, ending with "." on a line by itself
>>> ...
>>> DATE: 02 Mar 01 9:41:00 PM
>>> Reply-to: 2378tg_gogi_98902huh@mail.com
>>> Message-ID: <25zdDw9det07g2>
>>> TO: members2797327936_2786237sgsyu378352_287@mail.com
>>> SUBJECT: Targeted Bulk Emailing Service
>>> ...
>>> .
<<< 250 OK
>>> QUIT
<<< 221 Closing connection
```

DNSBL – DNS-basierte Blocklisten

„Schwarze Listen“ von Spam-freundlichen Servern, open relays

. . . DNS-basiert

Mailserver fragt beim Empfang einer Mail, ob der Server auf der Sperrliste steht.

janko.png

```
210.15.231.24 >>> tim.hrz.tu-chemnitz.de:25

<<< 220 tim.hrz.tu-chemnitz.de ESMTP Exim ...
>>> HELO dmpserver.dmp.com.au
<<< 250 tim.hrz.tu-chemnitz.de Hello [210.15.231.24]

... DNS-Abfrage: 24.231.15.210.relays.ordb.org ...  -
... 24.231.15.210.rbl-plus.mail-abuse.org ...
127.0.0.4

>>> MAIL FROM:<buynowornever@gmx.de>
<<< 250 <buynowornever@gmx.de> is syntactically correct
>>> RCPT TO:<frank.richter@hrz.tu-chemnitz.de>

<<< 550 User unknown

>>> RCPT TO:<spamforscher@phil.tu-chemnitz.de>

<<< 250 Ok
```

Anbieter:

- MAPS = http://www.mail-abuse.org kostenpflichtig seit 8/2001
- relays.ordb.org Open Relay Datenbank
- mehr in DNS-based Spam Databases: http://www.declude.com/junkmail/support/ip4r.htm

Anbieter:

- MAPS = http://www.mail-abuse.org kostenpflichtig seit 8/2001
- relays.ordb.org Open Relay Datenbank
- mehr in DNS-based Spam Databases: http://www.declude.com/junkmail/support/ip4r.htm

Probleme:

- Boykott ganzer Server betrifft auch Unschuldige
- Mail-Forwarding durch Benutzer (GMX -> Web.de -> TU)
- Große Provider betroffen (kundenserver.de = Schlund & Partner)

Überprüfen/Sperren von Absende-Adressen

- Domain des Senders wird geprüft:
  - im DNS?, MX!= 127.0.0.1
  - Neue Blocklisten – RHSBL: `http://www.rfc-ignorant.org`
  - Absende-Domain verstößt gegen RFCs, z.B. Ablehnen von DSN oder `postmaster@domain`
- Manuelle Sperr-Liste für besonders schlimme Spam-Absender / Domains
- Verschärft: Mailserver baut Rück-Verbindung zu Mailserver der Absender-Domain auf:

```
210.15.231.24 >>> tim.hrz.tu-chemnitz.de:25

<<< 220 tim.hrz.tu-chemnitz.de ESMTP Exim ...
>>> HELO dmpserver.dmp.com.au
<<< 250 tim.hrz.tu-chemnitz.de Hello [210.15.231.24]
>>> MAIL FROM:<buynowornever@gmx.de>
 tim.hrz.tu-chemnitz.de >>> mx0.gmx.de:25
 ...
 >>> RCPT TO:<buynowornever@gmx.de>
 <<< 550 {mx001-rz3} <buynowornever@gmx.de>...
 User unknown or not available
<<< 550 <buynowornever@yahoo.com> is not accepted
```

Probleme:

- Aufwändig . . . umstritten
- Geht nicht immer: Nicht alle verraten, ob Nutzer existiert (yahoo.com, aol.com)

Maßnahmen als Administrator / ISP

Spam-Policy festlegen:

- Verhalten bei Spam durch eigene Nutzer
- Zeit einplanen / erkämpfen, informieren, `abuse@domain`
- Welche Spamschutz-Maßnahmen kommen wo zum Einsatz?
  „Unterdrücken von E-Mails durch den Provider/Administrator ohne Zustimmung des Empfängers ist strafbar (§ 206 StGB).“
  ⇨ zentral am Mail-Server nur, wenn nutzerspezifisches Anwenden der Filter machbar, Kontrollmöglichkeit durch Nutzer
- Markieren von E-Mail ist ok – Filtern dezentral beim Empfänger – Bereitstellen von Werkzeugen, Schulung der Anwender …
- CPU … Ressourcen am Mail-Server einplanen
- siehe: Maßnahmen an der TU Chemnitz
  `http://www.tu-chemnitz.de/urz/mail/filter/`

Kein Relaying!

d.h kein Durchreichen von E-Mails, die nichts mit meinen Nutzern zu tun haben.

IMC - Relay Survey[1]:

02/1998: 55%, 01/2002: < 1% der Mail-Server offen

Test auf offenes Relay: `http://www.abuse.net/relay.html`

Infos zu Software und Konfiguration von Mail-Servern:
`http://mail-abuse.org/tsi/ar-fix.html`

Router-Konfiguration: SMTP Port 25 – nur für Mail-Server offen

[1] http://www.imc.org/ube-relay.html

Prävention durch Nutzer

Vernünftige TO-Zeilen!

. . . oder bekommen Sie sonst auch komplette Kundenlisten?

⇨ Niemals „an alle“ schreiben!

Prävention durch Nutzer

Vernünftige TO-Zeilen!

. . . oder bekommen Sie sonst auch komplette Kundenlisten?

⇨ Niemals „an alle“ schreiben!

Virenschutz!

. . . es trifft auch Unbeteiligte!

⇨ Vorsicht mit Attachments!

Prävention durch Nutzer

Vernünftige TO-Zeilen!

. . . oder bekommen Sie sonst auch komplette Kundenlisten?

⇨ Niemals „an alle“ schreiben!

Virenschutz!

. . . es trifft auch Unbeteiligte!

⇨ Vorsicht mit Attachments!

Achtung: 1-Pixel-Falle!

Bilder laden (bei HTML-Mails) ausschalten

⇨ Nicht klicken! Am besten auch kein HTML senden!

„Please unsubscribe“ ignorieren

. . . *„spring aus dem Fenster“ auch!*

⇨ Die Adresse wird sonst noch wertvoller – für Spammer!

„Please unsubscribe“ ignorieren

. . . „spring aus dem Fenster“ auch!

⇨ Die Adresse wird sonst noch wertvoller – für Spammer!

Keine falschen Adressen! (z.B. nospam-max@gmx.de)

. . . die Spammer reparieren das automatisch

⇨ Kommunikation muß funktionieren!

- Die Mail belastet das sonst Netz mehrfach!
- Füllwörter eher in der Domain als im Nutzernamen!
- Man sollte auch die verstümmelte Adresse besitzen!
- Adressen im WWW als Bild oder `mailto:%20user@domain`
- evtl. Zweitadresse mit stärkerem Filter

Fazit

Das Problem wächst: Konferenzen, Urteile, Zeitungsartikel . . .

Arbeitsorte: Gesetze, Nutzer, Server

Wenn einer der Orte vernachlässigt wird, ist die Zukunft unklar.

Chancen: Unklar! *Was meinen Sie?*

Vielen Dank!

Kontakt:
Frank Richter, *fri@hrz.tu-chemnitz.de*
Ralph Sontag, *sontag@mathematik.tu-chemnitz.de*

Zum Nachlesen:
Paul Graham: „A Plan for Spam“ und „Better Bayesian Filtering“
`http://www.paulgraham.com/spam.html`