deutsches forschungsnetz

Forum Mail:
Richtig Spam trainieren im DFN-Mailsupport

81. DFN-Betriebstagung | 08.10.2024

Andrea Wardzichowski

Was zuletzt geschah (1)

- ~Juni 2022: Nutzer berichten, dass die Spamerkennung durch den Bayes-Filter schlechter wird
- Trainiert wird zwar durch Admins, aber dennoch scheint es Unklarheiten zu geben
- Neuaufbau der Bayes-Datenbank mit ca. „10.000 Spam-Mails, 10.000 Ham-Mails“ schlägt im ersten Versuch fehl
- Genutzt wurde das Trainingsmaterial der letzten drei Jahre

Was zuletzt geschah (2)

- Problem: Zuviel Spam, zuwenig Ham haben zur Fehlfunktion beigetragen
- Aber auch: Fehlerhaftes Training

Einordnung: Welche Checks gibt es?

- https://www.mailsupport.dfn.de/dokumentation/checks
  - Dynamische Blacklists (fail2ban)
  - Verhalten (postscreen) und äußere Form (Größe, Header, DNS..)
  - RBL, Black-/Whitelists
  - SPF, DKIM, DMARC, ARC
  - Inhalt: URLs, Bayes-Filter
- Jeder Check vergibt Punkte, die Summe ergibt den Spam-Score.

Begriffe (1)

- Spam

  “billiges Dosenfleisch“

  unerwünschte Massen- und Werbe-E-Mails

  Nicht: bestellte (!) Newsletter

  => erkannter Spam sollte in einen junk- oder Werbeordner gelegt werden

- Ham

  “der gute Schinken“

  echte Mails von echten Menschen

Begriffe (2)

- Bayessche Statistik

  statistisches Verfahren zur Bewertung von Worten „eher Spam“

  „eher Ham“

- Spamtraining

  wirkt nur auf den Bayes-Check (und den Bayes-Score)

  nicht über die bereits erreichte Spam-Schwelle hinaus

Training im Mailsupport

- Einrichtung meldet sich bei uns und gibt einliefernde Mailserver für das Training bekannt (IP-Whitelist für Training)
- Einrichtung erhält Trainings-Mailadressen
  - spam-XXXXXXXXX@sYYY.mx.srv.dfn.de
- Nutzer schicken die Mails zum Trainieren an die Admins der Einrichtung
- Der Admin prüft
- ...und sendet die Mail ggf. weiter ins Training.

Training – wie richtig? Spam (1)

- Nicht erkannte Spam-Mails
  - Müssen durch Admins gesichtet werden!
  - Als Attachment (!) an die Spam-Trainingsadresse senden
  - Mehrere Attachments pro Mail möglich

Training – wie richtig? Spam (2)

- Nicht trainieren: Mails der Art
  - „Will ich nicht haben“ (besonders schlecht mit Einrichtungs-Footern!
    Ggf. von Maillinglisten unsubscriben.)
  - Newsletter (bitte abbestellen)
  - Phishing-Mails
  - Bereits erkannter Spam (bringt nichts, der Score wird nicht höher)

Training – wie richtig? Ham

- Fälschlich als Spam erkannte Mails
  - Müssen durch Admins gesichtet werden!
  - Als Attachment (!) an die Ham-Trainingsadresse senden
  - Mehrere Attachments pro Mail möglich
  - Gern ab und zu einen Schwung normaler Admin-Mails (z.B. Anfragen an uns) an die Ham-Adresse senden
- Eine ausgeglichene Anzahl an Spam- und Ham-Mails im Training ist wichtig, damit die Datenbank nicht erneut in ein Ungleichgewicht kommt

..und was passiert dann?

- **in beiden Fällen:** Markierungen im Subject wie [EXTERN] und diverse Header werden vor dem Training herausgefiltert
- Die Mails werden automatisiert und ohne Aktionen von Mailsupport-Mitarbeiter:innen verarbeitet
- Ausnahme: Newsletter-Training,
  an die Hotline eingesandte Phishing-Beispiele
- Aktuell sichten wir (aus gegebenem Anlaß) aber das Trainingsmaterial für den Neu-Aufbau der Bayes-Datenbank

Haben Sie noch Fragen?

- Kontakt
  - https://www.mailsupport.dfn.de/
  - hotline@mailsupport.dfn.de
  - Telefon: 0049 711 633 14 217