(19) Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 60 2005 001 046 T2** 2008.01.03

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 675 330 B1**
(21) Deutsches Aktenzeichen: **60 2005 001 046.7**
(96) Europäisches Aktenzeichen: **05 257 705.3**
(96) Europäischer Anmeldetag: **15.12.2005**
(97) Erstveröffentlichung durch das EPA: **28.06.2006**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **02.05.2007**
(47) Veröffentlichungstag im Patentblatt: **03.01.2008**

(51) Int Cl.[8]: ***H04L 12/58*** (2006.01)

(30) Unionspriorität:
18270 **21.12.2004** **US**

(84) Benannte Vertragsstaaten:
DE, FR, GB

(73) Patentinhaber:
Lucent Technologies Inc., Murray Hill, N.J., US

(74) Vertreter:
derzeit kein Vertreter bestellt

(72) Erfinder:
Cai, Yigang, Naperville, Illinois 60565, US; Qutub, Shehryar S., Hoffman Estates, Illinois 60194, US; Sharma, Alok, Lisle, Illinois 60532, US

(54) Bezeichnung: **Erkennung von unerwünschten Nachrichten (SPAM) auf Basis des Nachrichteninhalts**

Beschreibung

Technisches Gebiet

[0001] Die vorliegende Erfindung betrifft Verfahren zum Erkennen von Spam-Nachrichten auf der Basis des Inhalts der Nachricht.

Allgemeiner Stand der Technik

[0002] Mit dem Aufkommen des Internet ist es leicht geworden, mit geringen oder keinen Kosten für den Absender Nachrichten zu einer großen Anzahl von Zielen zu senden. Die Nachrichten umfassen die Kurznachrichten des SMS-Dienstes. Diese Nachrichten enthalten unerbetene und unerwünschte Nachrichten (Spam), die für den Empfänger der Nachricht, der die Nachricht löschen und bestimmen muß, ob sie von irgendwelcher Bedeutung ist, ärgerlich sind. Ferner sind sie für den Träger des zum Senden der Nachricht verwendeten Telekommunikationsnetzes ärgerlich, nicht nur weil sie mit Bezug auf zornige Kunden, die mit Spam überflutet werden, ein Kundenbeziehungsproblem darstellen, sondern auch weil diese Nachrichten, für die gewöhnlich wenig oder kein Umsatz entsteht, Netzbetriebsmittel benutzen. Eine Veranschaulichung der Ernsthaftigkeit dieses Problems wird durch die beiden folgenden Statistiken gegeben. In China wurden 2003 zwei Trillionen SMS-Nachrichten (Kurznachrichtendienst) über das chinesische Telekommunikationsnetz gesendet; von diesen Nachrichten waren schätzungsweise drei Viertel Spam-Nachrichten. Die zweite Statistik ist, daß in den Vereinigten Staaten schätzungsweise 85–90% der E-Mail-Nachrichten Spam sind.

[0003] Es wurden mehrere Anordnungen vorgeschlagen, um die Anzahl der abgelieferten Spam-Nachrichten zu verringern, und viele wurden implementiert. Es wurden viele Anordnungen vorgeschlagen, um Nachrichten vor ihrer Ablieferung zu analysieren. Gemäß einer Anordnung wird, wenn der anrufende Teilnehmer nicht zu einer vom Anrufer spezifizierten vorgewählten Gruppe gehört, die Nachricht blockiert. Spam-Nachrichten können auch abgefangen werden, indem man es einem angerufenen Teilnehmer erlaubt, zu spezifizieren, daß keine Nachrichten abgeliefert werden sollen, die für mehr als N Ziele bestimmt sind.

[0004] Ein angerufener Teilnehmer kann sich weigern, seine Telefonnummer oder E-Mail-Adresse zu publizieren. Zusätzlich zu den offensichtlichen Nachteilen, es Anrufern nicht zu erlauben, die Telefonnummer oder E-Mail-Adresse des angerufenen Teilnehmers nachzuschlagen, sind solche Anordnungen wahrscheinlich ineffektiv. Ein schlauer Hacker kann eine nicht aufgelistete E-Mail-Adresse aus dem IP-Netz erkennen, indem er zum Beispiel Nachrichtenkopfteile in einem Router überwacht. Eine nicht aufgelistete angerufene Nummer lädt den Anrufer dazu ein, Nachrichten zu allen zehntausend Telefonnummern eines Vermittlungsstellencodes zu senden; wie bereits erwähnt, ist dies mit derzeitigen Anordnungen zum Senden von Nachrichten zu mehreren Zielen sehr leicht.

[0005] Zu den schwerer zu fassenden Spam-Nachrichten gehören widerwärtige Nachrichten für pornographische Zwecke oder zum Übermitteln unerwünschter Werbung zu den Empfängern. Häufig können solche Nachrichten nur durch Untersuchung des Inhalts der Nachricht abgefangen werden, da die Absender möglicherweise viele harmlose Nachrichten von derselben Quelle senden. Ein Hauptproblem der Spam-Erkennung besteht darin, Spam auf der Basis des Inhalts der Nachricht zu erkennen.

[0006] Die Patentschrift WO 00/26795 (JUSTSYSTEM Pittsburgh Research Center, M. Kantrowitz et al.; 11.05.2000) beschreibt Anordnungen, die aus einem Begriffslexikon jedem Begriff eines Dokuments ein Gewicht zuweisen, das Gewicht mit der Anzahl des Auftretens multiplizieren und durch die Gesamtzahl der Wörter oder die Anzahl einzigartiger Wörter dividieren, um eine Bewertung abzuleiten, die die Möglichkeit darstellt, daß ein Dokument eine Junk-Nachricht repräsentieren kann.

[0007] Die Patentschrift WO 2004/061698 (Activestate Corporation, I. Dougherty et al.; 22.07.2004) beschreibt Anordnungen zum Erkennen von Spam-Nachrichten auf der Basis von Spam-Merkmalen. Die Merkmale dienen zum Ableiten von Klassifikationsinformationen zur Analyse von Nachrichten, um zu bestimmen, ob eine Nachricht eine Spam-Nachricht ist.

Kurzfassung der Erfindung

[0008] Gemäß der Erfindung der Anmelder wird das obige Problem gelöst und ein Fortschritt gegenüber dem Stand der Technik erzielt, wobei verdächtige Nachrichten auf die Anwesenheit bestimmter Eigenschaften, wie zum Beispiel Schlüsselwörter, und auf die Häufigkeit solcher Eigenschaften analysiert werden; jeder Eigenschaft wird ein entsprechender Spam-Index gegeben, eine Größe, die praktisch statisch ist und vordefiniert und provisioniert wird und vorteilhafterweise hängt ein Gewichtungsfaktor, der sich dynamisch ändert, von dem Verkehrsvolumen und Nachrichten-Inhaltstypen ab. Nachrichten werden auf jede Eigenschaft hin untersucht, deren Häufigkeit der Verwendung eine Schwelle übersteigt; auf vorbestimmte Kombinationen von Eigenschaften, deren kombinierte Verwendung eine Schwelle übersteigt; und auf alle Eigenschaften, deren kombinierte Verwendung eine Schwelle übersteigt. Gemäß einem Merkmal der Erfindung der Anmelder kann der Gewichtungsfaktor jeder Eigenschaft dynamisch einge-

stellt werden, um eine Anpassung an die Ergebnisse einer Untersuchung verdächtiger Nachricht durch einen menschlichen Analysierer zu erreichen. Vorteilhafterweise kann der Detektionsprozeß durch die Verwendung eines menschlichen Analysierers lernen.

[0009] Kurze Beschreibung der Zeichnung(en)

[0010] Fig. 1 zeigt die Funktionsweise der Erfindung der Anmelder; und

[0011] Fig. 2 ist ein Flußdiagramm der Erfindung der Anmelder.

Ausführliche Beschreibung

[0012] Fig. 1 zeigt die Funktionsweise der Erfindung der Anmelder. Eine Quelle **1** möchte eine Nachricht zu einem Ziel **2** senden. Die Nachricht wird zu einem Netz **3** gesendet, das erkennt, daß es sich hierbei um eine Spam-Nachricht handelt, aber um eine solche, die zur Bestimmung Nachrichteninhaltsanalyse erfordert. Das Netz **3** leitet die Nachricht zu einem Nachrichtenanalysator **10**. Wenn der Nachrichtenanalysator schließt, daß dies keine Spam-Nachricht ist, wird die Nachricht über das Netz **4** zu dem Ziel **2** gesendet.

[0013] Der Nachrichtenanalysator **10** enthält Tabellendaten **14** von Eigenschaften, einen Härteindex für jede Eigenschaft, einen Gewichtungsfaktor für jeden Härteindex und eine Härteindexschwelle für die Eigenschaft.

[0014] Eine Spam-Eigenschaft ist ein Wort, eine Phrase, ein Satz, ein Bild oder ein Videosegment, das ein möglicher Indikator einer Spam-Nachricht ist. Das Wort „Madam" ist ein Beispiel. Für jede in der Nachricht auftretende Eigenschaft wird ein Produkt der Anzahl, wie oft die Eigenschaft auftritt, des Härteindex und des Gewichtungsfaktors berechnet, um ein Härteniveau abzuleiten. Die Härteniveaus dienen zur Bestimmung, ob die Nachricht als Spam-Nachricht zu behandeln ist.

[0015] Der Härteindex und die Härteschwelle werden relativ konstant gehalten, aber der Gewichtungsfaktor kann als Reaktion auf Nachrichten von einem Spam-Dienstbüro **15** geändert werden, als Reaktion auf die Erkennung in dem Büro spezieller Problembereiche (um den Gewichtungsfaktor zu vergrößern) oder von Bereichen, in denen sehr wenig Spam-Aktivität bestand (um den Gewichtungsfaktor zu reduzieren).

[0016] Der Nachrichtenanalysator nimmt den Inhalt der Nachricht und sucht nach vorgespeicherten Eigenschaften, wie zum Beispiel den Wörtern „Madam" und „Lovers". Für jede vorgespeicherte Eigenschaft besteht ein Gewichtungsfaktor, um anzuzeigen, wie stark diese Eigenschaft gewichtet werden soll, um zu einem Härteniveau zu kommen. Nachrichten, deren Härteniveau eine vorbestimmte Schwelle übersteigt, werden blockiert und können für weitere menschliche Analyse gespeichert werden.

[0017] Fig. 2 ist ein Flußdiagramm der Funktionsweise der Spam-Prüfung der Anmelder. Eine ankommende Nachricht wird empfangen und zur Spam-Analyse gepuffert (Aktionsblock **201**). Die Spam-Tabellendaten werden erhalten, um den Spam-Härteindex für Eigenschaften der Nachricht zu berechnen (Aktionsblock **203**). Die Spam-Analyse kehrt für Nachrichteneigenschaften der Nachricht zu dem Spam-Härteindex zurück (Aktionsblock **205**). Dienstlogik füllt eine Analyse-Tabellenkalkulation mit dem Härteindex für jede Eigenschaft und erhält das verteilte Spam-Härteindexprofilmuster (Aktionsblock **207**). Die Prüfung **209** prüft, ob der Härteindex einer einzelnen die Schwelle für diese Eigenschaft übersteigt. Wenn etwaige die Grenze übersteigen (nachfolgend zu beschreibender Aktionsblock **221**), erfolgt der Eintritt. Andernfalls erfolgt der Eintritt in die Prüfung **211**, um zu prüfen, ob etwaige Muster des Härteindex eine Schwelle übersteigen. Wenn etwaige die Schwelle für das Muster übersteigen, erfolgt der Eintritt in den Aktionsblock **221**. Andernfalls wird unter Verwendung aller Eigenschaften oder aller Eigenschaften, deren Härteindex eine Schwelle übersteigt, ein aggregierter Spam-Härteindex berechnet (Aktionsblock **213**). Wenn dieser aggregierte Index eine Oberschwelle übersteigt (Prüfung **215**), ist die Nachricht schwarz. Wenn er kleiner als eine Unterschwelle ist (Prüfung **216**), ist die Nachricht weiß. Für andere Nachrichten bestimmt man mit der Prüfung **217**, ob die Nachricht einer menschlichen Analyse unterzogen werden soll. Wenn nicht, wird die Nachricht zu seinem Ziel weitergeleitet (Aktionsblock **223**). Wenn sie für menschliche Analyse ausgewählt wurde, wird die Nachricht zu einem Dienstbüro gesendet (Aktionsblock **218**). Das Ergebnis der menschlichen Untersuchung (Prüfung **219**) bestimmt entweder ein zufriedenstellendes Ergebnis, und die Nachricht wird weitergeleitet (Aktionsblock **223**), oder ein unzufriedenstellendes Ergebnis, und die Nachricht wird als Spam behandelt und den Funktionen des Aktionsblocks **221** unterzogen.

[0018] Aktionsblock **221** speichert gegebenenfalls die Spam-Nachricht, speichert eine aktualisierte Spamfilter- und regeldienstdatenbank, die durch die menschliche Untersuchung abgeleitet wurde, und aktualisiert den Spam-Härtegewichtsfaktor und die Indexobergrenze, und fügt gegebenenfalls neue verteilte Spam-Muster hinzu.

[0019] Die obige Beschreibung beschreibt eine bevorzugte Ausführungsform der Erfindung der Anmelder. Durchschnittsfachleuten werden andere Ausfüh-

rungsformen einfallen, ohne von dem Schutzumfang der Erfindung abzuweichen. Die Erfindung wird nur durch die angefügten Ansprüche beschränkt.

Patentansprüche

1. Verfahren zum Erkennen von unerwünschten Spam-Nachrichten in einem Telekommunikationsnetz, mit den folgenden Schritten:
Speichern (**14**) eines Index, der für jede Eigenschaft einer potentiellen Nachricht vordefiniert ist;
Erkennen und Speichern (**10**, **11**, **201**) einer Nachricht, von der verdächtigt wird, daß sie Spam ist;
Ableiten von Eigenschaften der gespeicherten Spam-Nachricht (**10**, **203**);
Berechnen des Produkts der Anzahl, wie oft jede Eigenschaft auftritt, und ihres Index (**203**);
Bilden eines verteilten Spam-Profils aus den Produkten (**207**); und
Bestimmen (**213**, **215**), ob das verteilte Spam-Profil den Kriterien für die Klassifizierung einer Nachricht als eine Spam-Nachricht genügt;
dadurch gekennzeichnet, daß
der Schritt des Speicherns (**10**, **11**, **201**) ferner den Schritt des Speicherns eines dynamisch einstellbaren Gewichtungsfaktors und einer Grenze für jede Eigenschaft einer potentiellen Nachricht umfaßt; und
der Schritt des Berechnens (**203**) den Schritt des Berechnens des Produkts der Anzahl, wie oft jede Eigenschaft auftritt, des dynamisch einstellbaren Gewichtungsfaktors und des vordefinierten Index umfaßt.

2. Verfahren nach Anspruch 1, wobei, wenn irgendein Produkt seine Obergrenze für die Eigenschaft dieses Produkts (**209**) überschreitet, die assoziierte Nachricht als eine Spam-Nachricht deklariert wird (**221**).

3. Verfahren nach Anspruch 1, ferner mit den folgenden Schritten:
für mehrere Muster von Eigenschaften wird eine Obergrenze für jedes Muster gespeichert (**211**); und
wenn die Obergrenze für irgendein Muster überschritten wird, Deklarieren einer Nachricht als Spam-Nachricht (**221**).

4. Verfahren nach Anspruch 1, wobei, wenn die Summe aller Produkte für die Nachricht eine vorbestimmte obere Schwelle (**213**) überschreitet, die Nachricht als eine Spam-Nachricht behandelt wird (**221**).

5. Verfahren nach Anspruch 1, wobei der Gewichtungsfaktor oder die Obergrenze einer Eigenschaft als Reaktion auf eine Nachricht von einem Servicebüro geändert werden können.

6. Vorrichtung zum Erkennen von unerwünschten Spam-Nachrichten in einem Telekommunikationsnetz, umfassend:
ein Mittel zum Speichern (**14**) eines Index, der für jede Eigenschaft einer potentiellen Nachricht vordefiniert ist;
ein Mittel zum Speichern (**10**, **11**, **201**) einer Nachricht, von der verdächtigt wird, daß sie Spam ist;
ein Mittel zum Ableiten (**10**, **203**) von Eigenschaften der gespeicherten Spam-Nachricht;
ein Mittel zum Berechnen (**10**, **203**) des Produkts der Anzahl, wie oft jede Eigenschaft auftritt, und ihres vordefinierten Index;
ein Mittel (**10**, **207**) zum Bilden eines verteilten Spam-Profils aus den Produkten; und
ein Mittel (**10**, **213**, **215**) zum Bestimmen, ob das verteilte Spam-Profil den Kriterien für die Klassifizierung einer Nachricht als eine Spam-Nachricht genügt;
dadurch gekennzeichnet, daß
das Mittel zum Speichern (**14**) ferner ein Mittel zum Speichern eines dynamisch einstellbaren Gewichtungsfaktors und einer Grenze für jede Eigenschaft einer potentiellen Nachricht umfaßt; und
das Mittel zum Berechnen (**10**, **201**) ein Mittel zum Berechnen des Produkts der Anzahl, wie oft jede Eigenschaft auftritt, des dynamisch einstellbaren Gewichtungsfaktors und des vordefinierten Index umfaßt.

7. Vorrichtung nach Anspruch 6, ferner mit einem Mittel zum Behandeln der assoziierten Nachricht als eine Spam-Nachricht (**10**, **211**), wenn irgendein Produkt seine Obergrenze für die Eigenschaft dieses Produkts (**10**, **209**) überschreitet.

8. Vorrichtung nach Anspruch 6, ferner umfassend: ein Mittel, das für mehrere Muster von Eigenschaften eine Obergrenze für jedes Muster speichert (**10**, **211**); und ein Mittel zum Behandeln einer Nachricht als eine Spam-Nachricht (**10**, **221**), wenn die Obergrenze für irgendein Muster überschritten wird.

9. Vorrichtung nach Anspruch 6, ferner mit einem Mittel zum Behandeln der assoziierten Nachricht als eine Spam-Nachricht (**10**, **211**), wenn die Summe aller Produkte für die Nachricht eine vorbestimmte obere Schwelle (**10**, **213**) überschreitet.

10. Vorrichtung nach Anspruch 6, ferner mit einem Mittel (**10**) zum Ändern des Gewichtungsfaktors oder der Obergrenze einer Eigenschaft als Reaktion auf eine Nachricht von einem Servicebüro (**15**).

Es folgen 2 Blatt Zeichnungen

Anhängende Zeichnungen

FIG. 1

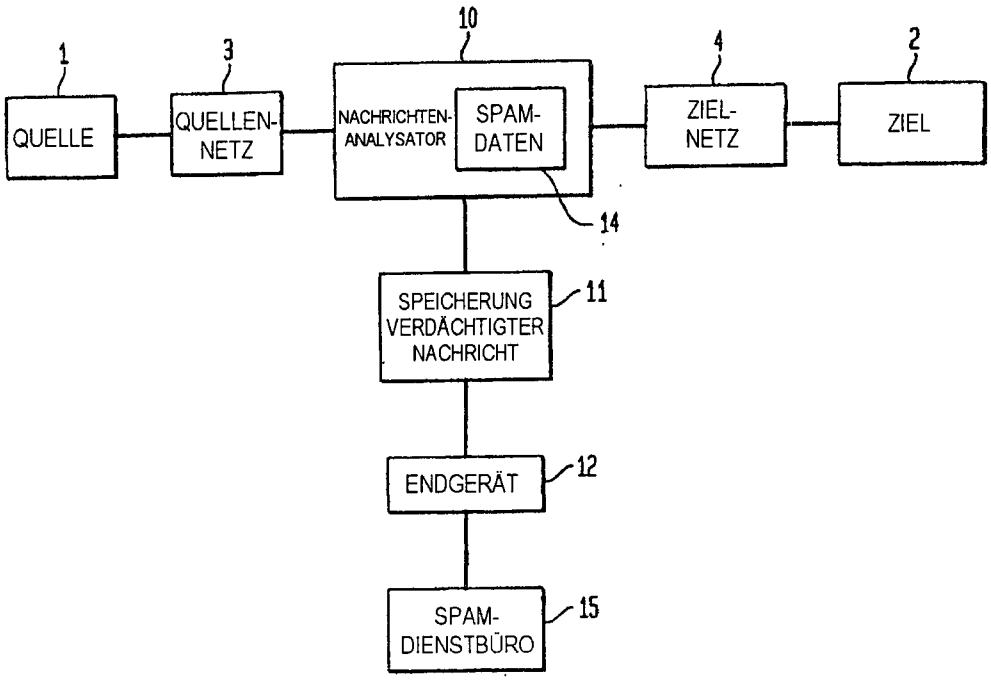
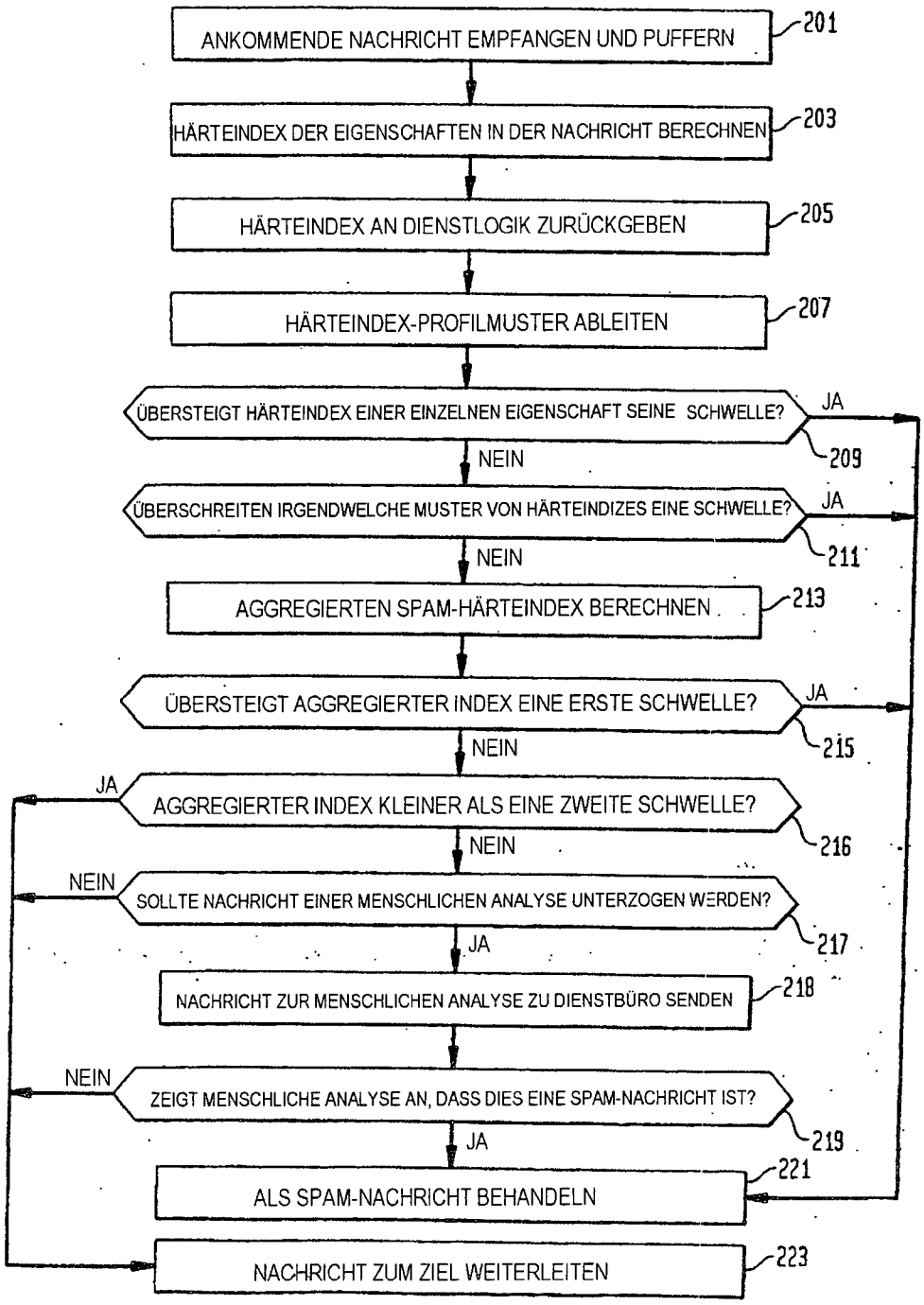


FIG. 2


ANKOMMENDE NACHRICHT EMPFANGEN UND PUFFERN
201
HÄRTEINDEX DER EIGENSCHAFTEN IN DER NACHRICHT BERECHNEN
203
HÄRTEINDEX AN DIENSTLOGIK ZURÜCKGEBEN
205
HÄRTEINDEX-PROFILMUSTER ABLEITEN
207
ÜBERSTEIGT HÄRTEINDEX EINER EINZELNEN EIGENSCHAFT SEINE SCHWELLE?
JA
209
NEIN
ÜBERSCHREITEN IRGENDWELCHE MUSTER VON HÄRTEINDIZES EINE SCHWELLE?
JA
211
NEIN
AGGREGIERTEN SPAM-HÄRTEINDEX BERECHNEN
213
ÜBERSTEIGT AGGREGIERTER INDEX EINE ERSTE SCHWELLE?
JA
215
NEIN
JA
AGGREGIERTER INDEX KLEINER ALS EINE ZWEITE SCHWELLE?
216
NEIN
NEIN
SOLLTE NACHRICHT EINER MENSCHLICHEN ANALYSE UNTERZOGEN WERDEN?
217
JA
NACHRICHT ZUR MENSCHLICHEN ANALYSE ZU DIENSTBÜRO SENDEN
218
NEIN
ZEIGT MENSCHLICHE ANALYSE AN, DASS DIES EINE SPAM-NACHRICHT IST?
219
JA
ALS SPAM-NACHRICHT BEHANDELN
221
NACHRICHT ZUM ZIEL WEITERLEITEN
223