(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년03월14일
(11) 등록번호 10-1021395
(24) 등록일자 2011년03월03일

(51) Int. Cl.
G06Q 50/00 (2006.01)
(21) 출원번호 10-2005-7013142
(22) 출원일자(국제출원일자) 2004년02월25일
심사청구일자 2008년12월31일
(85) 번역문제출일자 2005년07월15일
(65) 공개번호 10-2006-0006769
(43) 공개일자 2006년01월19일
(86) 국제출원번호 PCT/US2004/005501
(87) 국제공개번호 WO 2004/079514
국제공개일자 2004년09월16일
(30) 우선권주장
10/378,463 2003년03월03일 미국(US)
(56) 선행기술조사문헌
US20020199095 A1
US6161130 A
전체 청구항 수 : 총 81 항

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
라운쓰웨이트, 로버트 엘.
미국 98024 워싱턴주 폴 시티 287번 애비뉴 사우스이스트4148
헤커맨, 데이비드 이.
미국 98008 워싱턴주 벨레뷰 더블유. 레이크 삼마미쉬레인 노스이스트 648
(뒷면에 계속)
(74) 대리인
주성민, 이중희, 백만기

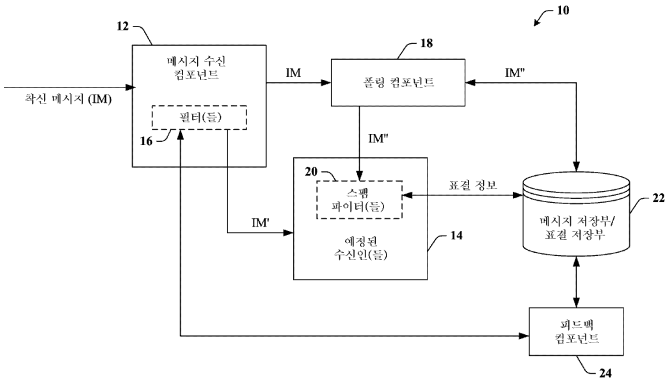
심사관 : 최석규

(54) 아이템을 분류하는 것을 용이하게 하는 시스템, 방법, 및 컴퓨터 판독가능 저장 매체

(57) 요 약

본 발명은 서버 및/또는 클라이언트 기반의 아키텍처에서 스팸 방지와 관련하여 아이템을 용이하게 분류하는 피드백 루프 시스템 및 방법을 제공한다. 본 발명은 스팸 필터에 적용된 기계 학습 방법을 사용하고, 특히 트레이닝 데이터 세트를 생성하기 위해 합법적 및/또는 정크/스팸 메일의 예들이 얻어지도록 착신 이메일 메시지를 임의로 샘플링한다. 스팸 파이터로 식별된 사용자는 그들의 착신 이메일 메시지의 선택이 개별적으로 합법적 메일인지 정크 메일인지의 여부를 표결하라는 요청을 받는다. 데이터베이스는 기계 학습 시스템에 대한 트레이닝 데이터를 생성하기 위해 각 메일에 대한 특성 및 사용자 정보와 같은 표결 트랜잭션, 메시지 특성 및 내용 요약, 및 메시지에 대한 폴링 결과를 저장한다. 기계 학습 시스템은 합법적 메일 및 스팸 필터를 인식하여 이들을 구별하도록 트레이닝되는 향상된 스팸 필터(들)을 용이하게 작성한다.

대 표 도



(72) 발명자

메어, 존 디.

미국 98103 시애틀 위트맨 애비뉴 엔. 넘버 3 3624

호웰, 나단 디.

미국 98133 워싱턴주 시애틀 엔. 105번 스트리트 아파트.에이 939

루퍼스버그, 미카 씨.

미국 98122 워싱턴주 시애틀 넘버 209 파인 스트리트 417

슬라우손, 딘 에이.

미국 98053 워싱턴주 레드몬드 더블유. 아메스 레이크드라이브 노스이스트 3343

굿맨, 조슈아 티.

미국 98052 레드몬드 노스이스트 38번 스트리트 17424

특허청구의 범위

청구항 1

컴퓨터-판독가능 저장 매체에 컴퓨터-실행가능 컴포넌트들이 구체화(embody)되어 있는, 스팸(spam) 방지와 관련하여 아이템을 분류하는 것을 용이하게 하는 시스템으로서,

한 세트의 아이템을 수신하는 컴포넌트;

상기 아이템들의 예정된(intended) 수신인을 식별하고, 폴링(polling)될 아이템들의 서브셋(subset)을 태그(tag)하는 컴포넌트- 상기 아이템들의 서브셋은 공지된 스팸 파이팅(spam fighting) 사용자들인 수신인들의 서브셋에 대응하고, 상기 폴링될 아이템들의 서브셋은 상기 아이템들이 스팸 또는 비 스팸(not spam)으로 라벨링되기 전에 결정되어, 현재 이용되고 있는 스팸 필터에 의해 스팸으로 지정된 아이템들을 비롯하여 모든 아이템들이 폴링용으로 고려됨-;

상기 폴링된 아이템들의 스팸 파이팅 사용자의 분류에 관련된 정보를 수신하고, 스팸 필터의 트레이닝(training) 및 스팸 리스트를 채우는 것(populating)과 관련하여 정보를 이용하는 피드백 컴포넌트- 상기 피드백 컴포넌트는 상기 스팸 필터를 트레이닝하기 위해 기계 학습 기술을 이용함-; 및

폴링을 위해 태그된 아이템을 폴링 아이템으로 식별하도록 수정하는 컴포넌트- 수정된 상기 아이템은 표결 명령어(voting instructions), 및 상기 사용자에 의한 상기 아이템의 분류를 용이하게 하는 적어도 2개의 각각의 아이템 클래스에 대응하는 적어도 2개의 표결 버튼 및 링크 중 어느 하나를 포함하고, 상기 표결 버튼들은 각각의 링크들에 대응하여서 상기 표결 버튼들 중 임의의 하나가 상기 사용자로부터의 입력에 의해 선택되었을 때, 선택된 상기 표결 버튼에 관련된 정보, 각각의 상기 사용자, 및 해당하는 사용자에 할당된 상기 아이템의 고유 ID가 저장을 위해 데이터베이스에 전송됨;

를 포함하는 것을 특징으로 하는 시스템.

청구항 2

제1항에 있어서, 상기 아이템들은 전자 메일(이메일) 및 메시지 중 적어도 하나를 포함하는 것을 특징으로 하는 시스템.

청구항 3

제1항에 있어서, 상기 한 세트의 아이템들을 수신하는 컴포넌트는 이메일 서버, 메시지 서버 및 클라이언트 이메일 소프트웨어 중 어느 하나인 것을 특징으로 하는 시스템.

청구항 4

제1항에 있어서, 상기 폴링될 아이템들의 서브셋은 수신된 모든 아이템들을 포함하는 것을 특징으로 하는 시스템.

청구항 5

제1항에 있어서, 상기 수신인들의 서브셋은 모든 수신인을 포함하는 것을 특징으로 하는 시스템.

청구항 6

제1항에 있어서, 상기 수신인들의 서브셋은 임의로 선택되는 것을 특징으로 하는 시스템.

청구항 7

제1항에 있어서, 상기 수신인들의 서브셋은 시스템의 유료 사용자를 포함하는 것을 특징으로 하는 시스템.

청구항 8

제1항에 있어서, 적어도, 통상적으로 필터링될 수 있는 메시지들의 서브셋은 폴링을 위해 고려되는 것을 특징으로 하는 시스템.

청구항 9

제1항에 있어서, 상기 폴링을 위해 태그된 아이템들의 서브셋은,

사용자 당 선택된 다수의 아이템;

기간 당 사용자 당 선택된 다수의 아이템; 및

공지된 사용자에 대응하는 아이템을 태그하는 확률

중 적어도 하나로 제한되는 것을 특징으로 하는 시스템.

청구항 10

제1항에 있어서, 상기 태그된 아이템들에는 태그된 아이템 및 그 태그된 아이템의 내용 중 어느 하나에 대응하는 고유 ID가 각각 할당되는 것을 특징으로 하는 시스템.

청구항 11

제1항에 있어서, 상기 수정된 아이템은,

수정된 "프롬(from)" 어드레스;

수정된 서브젝트 라인(subject line);

폴링 아이콘; 및

폴링 아이템으로서 식별하기 위한 폴링 컬러

중에서 적어도 하나를 포함하는 것을 특징으로 하는 시스템.

청구항 12

제1항에 있어서, 상기 수정된 아이템은 태그된 아이템을 어태치먼트(attachment)로서 포함하는 것을 특징으로 하는 시스템.

청구항 13

제1항에 있어서, 상기 수정된 아이템은 태그된 아이템의 요약(summary)을 포함하고, 상기 요약은 주제, 날짜, 메시지의 텍스트, 및 텍스트의 처음 몇 줄 중에서 적어도 하나를 포함하는 것을 특징으로 하는 시스템.

청구항 14

제1항에 있어서, 상기 적어도 2개의 표결 버튼은 "합법적(legitimate) 메일"을 나타내는 제1 표결 버튼, 및 "스팸"을 나타내는 제2 표결 버튼을 포함하는 것을 특징으로 하는 시스템.

청구항 15

제1항에 있어서, 상기 표결 버튼들은 아이템의 텍스트를 수정함으로써 구현되는 것을 특징으로 하는 시스템.

청구항 16

제1항에 있어서, 상기 표결 버튼들은 클라이언트 이메일 소프트웨어의 사용자 인터페이스를 수정함으로써 구현되는 것을 특징으로 하는 시스템.

청구항 17

제1항에 있어서, 사용자의 특성에 관련된 정보 및 데이터, 태그된 아이템과 연관된 아이템 내용 및 특성, 사용자 분류 및 표결 통계 데이터, 사용자 당 폴링 및 기간 당 사용자 당 폴링의 빈도 분석 데이터, 스팸 리스트, 합법적 메일 리스트, 및 블랙 홀 리스트를 저장하는 중앙 데이터베이스를 더 포함하는 것을 특징으로 하는 시스템.

청구항 18

제1항에 있어서, 현존하는 필터에 의해 스팸으로서 표시되는 폴링을 위해 태그된 아이템들은 사용자의 인박스(inbox)에 전달되고, 폴링을 위해 고려되는 것을 특징으로 하는 시스템.

청구항 19

제1항에 있어서, 폴링을 위해 태그된 아이템들은,

검출된 바이러스가 제거되고, 아이템이 폴링되는 것; 및

감염된 아이템이 폐기되는 것

중에서 하나가 발생하도록 바이러스가 스캐닝되는 것을 특징으로 하는 시스템.

청구항 20

제1항에 있어서, 상기 시스템은 각 회사로부터의 피드백이 각 회사와 동작적으로 인터페이스된 중앙 데이터베이스에 보내지도록 하나 이상의 스팸-파이팅 회사 에 제공되고, 상기 피드백의 일부는 사적기밀(privacy)의 이유로 제거되는 것을 특징으로 하는 시스템.

청구항 21

제20항에 있어서, 상기 회사 피드백은,

합법적 아이템을 배제한 스팸 전용 아이템들; 및

스팸 아이템들, 및 합법적 아이템들의 송신자명, 도메인 명칭 및 IP 어드레스

중에서 하나를 포함하는 것을 특징으로 하는 시스템.

청구항 22

제1항에 있어서, 사용자 신뢰도(reliability)를 테스트하는 사용자 분류 확인 컴포넌트를 더 포함하는 것을 특징으로 하는 시스템.

청구항 23

제22항에 있어서, 상기 사용자 분류 확인 컴포넌트는 상호-확인(cross-validation) 기술 및 공지된 결과 테스트 메시지 기술 중 적어도 하나인 것을 특징으로 하는 시스템.

청구항 24

제22항에 있어서, 상기 사용자 분류 확인 컴포넌트는 하나 이상의 의심스러운 사용자에게 적용될 수 있는 것을 특징으로 하는 시스템.

청구항 25

제1항에 있어서, 상기 피드백 컴포넌트는 사용자 피드백, 허니팟(honeypot) 피드백, 및 선택사양인 수신된 아이템의 사용자 수신인 피드백에 관련된 정보를 수신하는 것을 특징으로 하는 시스템.

청구항 26

스팸 방지와 관련하여 메시지들을 분류하는 것을 용이하게 하는, 컴퓨터 실행가능 컴포넌트들을 포함하는 시스템에 의해 실행되는 방법에 있어서,

한 세트의 메시지들을 수신하는 단계;

상기 메시지들의 예정된 수신인을 식별하는 단계;

폴링될 메시지들의 서브셋을 태그하는 단계- 상기 메시지들의 서브셋은 공지된 스팸 파이팅 사용자들인 수신인들의 서브셋에 대응하고, 상기 폴링될 메시지들의 서브셋은, 상기 메시지들이 스팸 또는 비 스팸으로 라벨링되기 전에 결정되어, 현재 이용되고 있는 스팸 필터에 의해 스팸으로 지정된 메시지들을 비롯하여 모든 메시지들이 폴링용으로 고려됨-;

상기 폴링 메시지의 사용자의 분류에 관련된 정보를 수신하는 단계;

스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 정보를 이용하는 단계 -상기 스팸 필터의 트레이닝은 기계 학습 기술을 통해 이용됨-; 및

폴링을 위해 태그된 메시지를 폴링 메시지로 식별하도록 수정하는 단계- 수정된 상기 메시지는 표결 명령어(voting instructions), 및 상기 사용자에 의한 상기 메시지의 분류를 용이하게 하는 적어도 2개의 각각의 메시지 클래스에 대응하는 적어도 2개의 표결 버튼 및 링크 중 어느 하나를 포함하고, 상기 표결 버튼들은 각각의 링크들에 대응하여서 상기 표결 버튼들 중 임의의 하나가 상기 사용자로부터의 입력에 의해 선택되었을 때, 선택된 상기 표결 버튼에 관련된 정보, 각각의 상기 사용자, 및 해당하는 사용자에 할당된 상기 메시지의 고유 ID가 저장을 위해 데이터베이스에 전송됨;

를 포함하는 것을 특징으로 하는 방법.

청구항 27

제26항에 있어서, 적어도, 통상적으로 필터링될 수 있는 메시지들의 서브셋은 이메일 서버에 의해 그리고 피드백 루프 시스템 내로 수신되는 것을 특징으로 하는 방법.

청구항 28

제26항에 있어서, 폴링을 위해 선택된 메시지들이 개별 사용자의 선호도(preference)에 따라 지정되도록 모든 착신 메시지는 클라이언트 이메일 소프트웨어에 의해 다루어지는 것을 특징으로 하는 방법.

청구항 29

제26항에 있어서, 수신된 모든 메시지는 데이터의 바이어스를 경감시키기 위해 폴링용으로 고려되는 것을 특징으로 하는 방법.

청구항 30

제26항에 있어서, 상기 폴링될 메시지들의 서브셋은 모든 메시지를 포함하는 것을 특징으로 하는 방법.

청구항 31

제26항에 있어서, 상기 수신인의 서브셋은 모든 수신인을 포함하는 것을 특징으로 하는 방법.

청구항 32

제26항에 있어서, 공지된 스팸 파이팅 사용자들인 상기 수신인들의 서브셋은,

새로운 스팸 필터의 트레이닝을 용이하게 하기 위해 메시지들에 관한 피드백을 제공하기 위한, 상기 수신자들 중 제1 수신자로부터의 입력을 수신하는 단계;

메시지들에 관한 피드백을 제공하기 위해 상기 수신자들 중 제2 수신자에게 통지를 제공하고, 상기 제2 수신자로부터 탈퇴하기 위한 응답 입력을 제공받지 않는 단계;

참여 메시지 서버에 의해 제공되는 이메일 및 메시지 서비스에 대해 지불하기 위한, 상기 수신자들 중 제3 수신자로부터의 입력을 수신하는 단계; 및

참여 메시지 서버에 이메일 어카운트를 오픈하기 위한, 상기 수신자들 중 제4 수신자로부터의 입력을 수신하는 단계

중 적어도 하나에 의하여 결정되는 것을 특징으로 하는 방법.

청구항 33

제26항에 있어서, 메시지 폴링에 참여하도록 선택된 사용자들의 서브셋은 임의(random)로 선택되는 것을 특징으로 하는 방법.

청구항 34

제26항에 있어서, 메시지 폴링에 참여하도록 선택된 사용자들의 서브셋은 모든 유료 사용자로부터 선택됨으로써, 일부 스패머(spammer)들이 스팸 필터 트레이닝을 파괴시키는데 더 많은 비용이 들게 하는 것을 특징으로 하는 방법.

청구항 35

제26항에 있어서, 상기 폴링을 위해 태그된 메시지들의 서브셋은 임의로 선택되는 것을 특징으로 하는 방법.

청구항 36

제26항에 있어서, 상기 폴링용으로 태그된 메시지들의 서브셋은 하나 이상의 폴링 제한(polling limits)에 의해 제한되는 것을 특징으로 하는 방법.

청구항 37

제36항에 있어서, 하나 이상의 폴링 제한은 데이터의 바이어스를 경감시키기 위해 사용자 당 한계 및 기간 당 사용자 당 제한을 포함하는 것을 특징으로 하는 방법.

청구항 38

제26항에 있어서, 태그된 메시지들을 수정하여 이를 폴링 메시지로서 표시하고 식별하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 39

제38항에 있어서, 상기 태그된 메시지들을 수정하는 단계는,

상기 태그된 메시지를 폴링 메시지용의 분리된 폴더로 이동시키는 단계;

상기 태그된 메시지의 "프롬(from)" 어드레스를 수정하는 단계;

상기 태그된 메시지의 서브젝트 라인을 수정하는 단계;

상기 태그된 메시지를 폴링 메시지로서 식별하기 위해 상기 태그된 메시지에 관한 폴링 아이콘을 사용하는 단계; 및

상기 태그된 메시지를 폴링 메시지로서 식별하기 위해 고유 컬러를 사용하는 단계

중에서 적어도 하나를 실행하는 단계를 포함하는 것을 특징으로 하는 방법.

청구항 40

제26항에 있어서, 상기 폴링 메시지는 원래 수신된 메시지의 어태치먼트, 및 표결하는 방법에 관해 사용자에게 지시하는 한 세트의 명령어를 포함하는 것을 특징으로 하는 방법.

청구항 41

제40항에 있어서, 메시지를 스팸 및 비 스팸으로 분류하는 것을 용이하게 하기 위해 적어도 2개의 표결 버튼을 더 포함하는 것을 특징으로 하는 방법.

청구항 42

제41항에 있어서, 장래 폴링에서 탈퇴(out)하도록 선택(opt)하기 위한 제3 표결 버튼을 더 포함하는 것을 특징으로 하는 방법.

청구항 43

제41항에 있어서, 상기 표결 버튼들은 폴링 메시지를 각 사용자에게 보내기 전에 메시지의 텍스트를 수정함으로써 폴링 메시지에 결합되는 것을 특징으로 하는 방법.

청구항 44

제41항에 있어서, 상기 표결 버튼들은 클라이언트 이메일 소프트웨어의 사용자 인터페이스를 수정함으로써 구현

되는 것을 특징으로 하는 방법.

청구항 45

제41항에 있어서, 상기 표결 버튼들은 폴링 메시지에 결합되는 것을 특징으로 하는 방법.

청구항 46

제40항에 있어서, 메시지의 요약을 더 포함하고, 상기 요약은 서브젝트 라인, 메시지 송신자, 메시지가 보내진 날짜, 메시지가 수신된 날짜, 메시지에서의 텍스트의 시작 부분 중 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 47

제26항에 있어서, 상기 태그된 메시지들이 폴링용으로 다운로드되기 전에 상기 태그된 메시지들을 바이러스에 대해 스캐닝하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 48

제47항에 있어서, 임의의 감염된 메시지들로부터 바이러스를 제거하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 49

제47항에 있어서, 바이러스가 감염된 태그된 메시지들은 폐기되는 것을 특징으로 하는 방법.

청구항 50

제26항에 있어서, 각 사용자들이 원래 형태의 제1 메시지 카피 및 폴링하기 위한 형태의 제2 메시지 카피를 수신하도록 원래 수신된 각각의 태그된 메시지의 카피를 작성하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 51

제26항에 있어서, 상기 태그된 메시지들은 상기 태그된 메시지와 상기 태그된 메시지의 내용 중에서 적어도 하나에 대응하는 고유 ID가 개별적으로 할당되는 것을 특징으로 하는 방법.

청구항 52

제51항에 있어서, 상기 태그된 메시지들 및 그 할당된 ID는 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 데이터베이스에 저장되는 것을 특징으로 하는 방법.

청구항 53

제26항에 있어서, 피드백 컴포넌트는 폴링 메시지의 사용자 분류에 관련된 정보를 수신하고, 상기 피드백 컴포넌트는 중앙 데이터베이스를 포함하는 것을 특징으로 하는 방법.

청구항 54

제53항에 있어서, 상기 데이터베이스는 기계 학습 기술을 통해 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 정보를 제공하는 것을 특징으로 하는 방법.

청구항 55

제53항에 있어서, 메일 서버 및 클라이언트 이메일 소프트웨어에 의해 생성된 데이터가 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 저장하기 위해 중앙 데이터베이스로 복귀되도록, 사용자를 식별하고 및 폴링을 위해 메시지를 태그하는 것은 하나 이상의 메일 서버 및 하나 이상의 클라이언트 이메일 소프트웨어에 제공되는 것을 특징으로 하는 방법.

청구항 56

제55항에 있어서, 데이터의 일부분만이 스팸 필터의 트레이닝을 용이하게 하기 위해 중앙 데이터베이스로 보내

지도록, 정보는 메일 서버 및 클라이언트 이메일 소프트웨어에 의해 중앙 데이터베이스로 보내지는 소정의 데이터로부터 사적기밀의 이유로 제거되는 것을 특징으로 하는 방법.

청구항 57

제56항에 있어서, 중앙 데이터베이스로 보내진 데이터의 일부분은,

스팸 메시지에 관련된 정보;

합법적 메시지에 포함된 도메인 명칭; 및

합법적 메시지에 포함된 IP 어드레스

중에서 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 58

제55항에 있어서, 메일 서버에 의해 생성된 데이터 및 클라이언트 이메일 소프트웨어에 의해 생성된 데이터는 각각 폴링 결과 및 폴링 메시지에 대응하는 통계 데이터에 통합됨으로써, 데이터를 중앙 데이터베이스로 전송하는데 요구된 대역폭을 경감시키는 것을 특징으로 하는 방법.

청구항 59

제58항에 있어서, 상기 메시지는 능동적 학습 기술, 즉 새롭거나 갱신된 필터를 학습하기 위해 이들의 평가치에 기초하여 메시지를 선택하는 기술을 사용하여 선택되는 것을 특징으로 하는 방법.

청구항 60

제26항에 있어서, 상기 스팸 필터는 폴링 데이터의 바이어스, 및 폴링 메시지의 잘못된 분류(misclassification)를 경감시키기 위해 스팸 및 비 스팸으로 분류된 메시지를 사용하여 트레이닝되는 것을 특징으로 하는 방법.

청구항 61

제26항에 있어서, 트레이닝된 스팸 필터를 하나 이상의 서버에 분배시키는 단계를 더 포함하고, 상기 분배는 이메일 메시지, 및 다운로딩을 위한 웹사이트 상의 포스팅(posting) 중의 적어도 하나에 의해 자동적으로, 요청에 의해, 또는 자동적으로 및 요청에 의해 발생하는 것을 특징으로 하는 방법.

청구항 62

제26항에 있어서, 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것은 사용자 분류 피드백에 기초한 데이터, 및 선택사양인 하나 이상의 추가 소스에 의해 생성된 데이터를 사용하여 기계 학습 기술에 의해 실행되고, 상기 하나 이상의 소스는 허니팟, 수신인 비사용자 분류 피드백, 및 능동적 학습 기술을 포함하는 것을 특징으로 하는 방법.

청구항 63

제62항에 있어서, 상기 하나 이상의 소스에 의해 생성된 데이터는 데이터의 바이어스되지 않은 샘플링을 용이하게 얻기 위해, 소스에 의해 생성된 데이터의 유형에 관하여 그리고 사용자 분류 데이터에 관련하여 비례적으로 재웨이팅(re-weighting)되는 것을 특징으로 하는 방법.

청구항 64

제62항에 있어서, 허니팟은 합법적인 메시지를 보내고 있는 사람이 알려지도록 제한적인 방식으로 개시된 이메일 어드레스에 대응함으로써, 스패머의 즉시 식별, 사용자 가입자 정보를 스패머에게 분배하는 의심스러운 상인의 검증, 및 사용자 분류를 기다리지 않고 스팸 메시지의 즉시 분류를 용이하게 하는 것을 특징으로 하는 방법.

청구항 65

제64항에 있어서, 허니팟에 의해 생성된 정보는 사용자 분류 피드백을 포함하는 다수의 다른 소스에 관련하여 사용시에 적어도 부분적으로 다수의 허니팟에 의존하여 선택적으로 다운 웨이팅(down weighting)되는 것을 특징

으로 하는 방법.

청구항 66

제64항에 있어서, 허니팟에 의해 생성된 데이터는 중앙 데이터베이스에 실시간 통합되고, 사용자 분류 및 폴링 메시지에 관련된 정보는 또한 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 나중에 사용하기 위해 저장되는 것을 특징으로 하는 방법.

청구항 67

제26항에 있어서,

착신 메시지 각각의 하나 이상의 포지티브 특징(positive feature)들에 대해 착신 메시지를 모니터하는 단계;

수신된 포지티브 특징의 빈도를 판정하는 단계;

수신된 하나 이상의 포지티브 특징이 적어도 부분적으로 이력 데이터에 기초한 임계 빈도를 초과하는 지를 판정하는 단계; 및

의심스러운 메시지가 스팸인지 판정하기 위해 추가적인 분류 데이터가 이용가능할 때까지, 임계 빈도를 초과하는 하나 이상의 포지티브 특징에 대응하는 의심스러운 메시지를 차단하는 단계

를 더 포함하는 것을 특징으로 하는 방법.

청구항 68

제67항에 있어서, 사용된 특징은 송신자의 IP 어드레스 및 도메인 중의 적어도 하나를 포함하는 송신자에 대한 정보인 것을 특징으로 하는 방법.

청구항 69

제67항에 있어서, 의심스러운 메시지를 차단하는 단계는,

의심스러운 메시지를 스팸으로서 임시로 라벨링(provisionally labeling)하고, 이들을 스팸 폴더로 이동시키는 동작;

추가 분류 데이터가 이용가능할 때까지 사용자(들)로의 의심스러운 메시지의 전달을 지연하는 동작; 및

의심스러운 메시지를 사용자(들)에게 보이지 않는 폴더 내에 저장하는 동작

중의 적어도 하나에 의해 실행되는 것을 특징으로 하는 방법.

청구항 70

제26항에 있어서, 스팸 필터의 최적화를 용이하게 하기 위해 스팸 필터의 잘못된 포지티브 및 캐치 비율(catch rate)을 판정하는 단계를 더 포함하고, 상기 잘못된 포지티브 및 캐치 비율을 판정하는 단계는,

제1 세트의 폴링 결과를 포함하는 트레이닝 데이터 세트를 사용하여 스팸 필터를 트레이닝하는 단계;

제2 세트의 폴링 결과를 생성하기 위해 사용자 피드백을 사용하여 제2 세트의 폴링 메시지를 분류하는 단계;

트레이닝된 스팸 필터를 통해 제2 세트의 폴링 메시지를 통하게 하는 단계; 및

필터의 잘못된 포지티브 및 캐치 비율을 판정하기 위해 제2 세트의 폴링 결과를 트레이닝된 스팸 필터 결과에 비교함으로써, 최적의 필터 성능에 따라 필터 파라미터를 평가하고 조정하는 단계

를 포함하는 것을 특징으로 하는 방법.

청구항 71

제70항에 있어서, 각 스팸 필터의 잘못된 포지티브 및 캐치 비율이 스팸 필터링을 위한 최적 파라미터를 판정하기 위해 적어도 하나의 다른 스팸 필터에 비교되도록, 하나 이상의 스팸 필터가 설정되고, 각각 상이한 파라미터를 갖고, 동일한 트레이닝 데이터 세트 상에서 트레이닝되는 것을 특징으로 하는 방법.

청구항 72

제26항에 있어서, 착신 메시지의 추가 세트를 사용하여 향상된 스팸 필터를 만드는 단계를 더 포함하고, 상기 착신 메시지의 서브셋은 향상된 스팸 필터의 트레이닝과 관련하여 새로운 정보를 생성하기 위해 폴링되며, 이전에 획득한 정보는 해당 정보가 얼마만큼 오래전에 획득한 것인지에 적어도 부분적으로 기초하여 재웨이팅되는 것을 특징으로 하는 방법.

청구항 73

제26항에 있어서, 합법적인 송신자 리스트를 만들기 위해 정보를 사용하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 74

제73항에 있어서, 합법적인 송신자 리스트는 정당한 것으로 분류된 메시지의 퍼센트에 따라 정당한 메일의 소스로서 실질적으로 분류되는 IP 어드레스, 도메인 명칭 및 URL 중에서 어느 하나를 포함하는 것을 특징으로 하는 방법.

청구항 75

제26항에 있어서, 상기 스팸 리스트는 받아들여질 수 있는 메일이 없는 어드레스의 블랙홀 리스트를 생성하는데 사용되는 것을 특징으로 하는 방법.

청구항 76

제26항에 있어서, 스패머의 어카운트를 용이하게 종결시키기 위해 정보를 사용하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 77

제76항에 있어서, ISP를 사용하고 있는 스패머를 식별하여 ISP에 스패밍(spamming)을 자동으로 통지하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 78

제76항에 있어서, 스팸의 송신을 담당하는 도메인을 식별하여, 도메인의 이메일 제공자 및 도메인의 ISP 중 적어도 하나에 스패밍을 자동으로 통지하는 단계를 더 포함하는 것을 특징으로 하는 방법.

청구항 79

제26항에 있어서, 스팸 필터 및 스팸 리스트 중의 적어도 하나를 메일 서버, 이메일 서버 및 클라이언트 이메일 소프트웨어 중 어느 하나에 분배시키는 단계를 더 포함하고, 분배는,

스팸 필터 및 스팸 리스트가 다운로딩에 이용가능하다는 것을 통지하는 웹사이트 상에 통지를 게시하는(posting) 단계;

스팸 필터 및 스팸 리스트를 메일 서버, 이메일 서버 및 클라이언트 이메일 소프트웨어에 자동적으로 보내는 단계; 및

스팸 필터 및 스팸 리스트를 메일 서버, 이메일 서버 및 클라이언트 이메일 소프트웨어에 수동적으로 보내는 단계

중에서 적어도 하나를 포함하는 것을 특징으로 하는 방법.

청구항 80

하나 이상의 프로세서에 의해 실행될 때, 스팸 방지와 관련하여 메시지들을 분류하는 것을 용이하게 하기 위한 컴퓨터 컴포넌트들이 저장되어 있는 컴퓨터-판독가능 저장 매체로서, 상기 컴퓨터 컴포넌트들은,

한 세트의 메시지를 수신하는 컴포넌트;

메시지의 예정된 수신인을 식별하고, 폴링될 메시지의 서브셋을 태그하는 컴포넌트- 상기 메시지의 서브셋은 공지된 스팸 파이팅 사용자들인 수신인의 서브셋에 대응하고, 상기 폴링될 메시지들의 서브셋은, 상기 메시지들이 스팸 또는 비 스팸으로 라벨링되기 전에 결정되어, 현재 이용되고 있는 스팸 필터에 의해 스팸으로 지정된 메시지들을 비롯하여 모든 메시지들이 폴링용으로 고려됨;

사용자에 대한 폴링 메시지로서 식별하기 위해 태그된 메시지를 수정하는 메시지 수정 컴포넌트;

폴링된 메시지의 사용자 분류에 관련된 정보를 수신하고, 스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 정보를 사용하는 피드백 컴포넌트- 상기 피드백 컴포넌트는 상기 스팸 필터를 트레이닝하기 위한 기계 학습 기술을 이용함-; 및

폴링을 위해 태그된 메시지를 폴링 메시지로 식별하도록 수정하는 컴포넌트- 수정된 상기 메시지는 표결 명령어(voting instructions), 및 상기 사용자에 의한 상기 메시지의 분류를 용이하게 하는 적어도 2개의 각각의 메시지 클래스에 대응하는 적어도 2개의 표결 버튼 및 링크 중 어느 하나를 포함하고, 상기 표결 버튼들은 각각의 링크들에 대응하여서 상기 표결 버튼들 중 임의의 하나가 상기 사용자로부터의 입력에 의해 선택되었을 때, 선택된 상기 표결 버튼에 관련된 정보, 각각의 상기 사용자, 및 해당하는 사용자에 할당된 상기 메시지의 고유 ID가 저장을 위해 데이터베이스에 전송됨-;

를 포함하는 컴퓨터 판독가능 저장 매체.

청구항 81

스팸 방지와 관련하여 메시지들을 분류하는 것을 용이하게 하는, 컴퓨터-판독가능 저장 매체를 포함하는 시스템에 있어서,

한 세트의 메시지를 수신하는 컴퓨터-실행가능 수단;

메시지의 예정된 수신인을 식별하는 컴퓨터-실행가능 수단;

폴링될 메시지의 서브셋을 태그하는 컴퓨터-실행가능 수단- 상기 메시지의 서브셋은 공지된 스팸 파이팅 사용자들인 수신인들의 서브셋에 대응하고, 상기 폴링될 메시지들의 서브셋은, 상기 메시지들이 스팸 또는 비 스팸으로 라벨링되기 전에, 현재 이용되고 있는 스팸 필터에 의해 스팸으로 지정되는 메시지들을 비롯하여 모든 메시지들이 폴링용으로 고려되는 것으로 판정됨;

폴링 메시지의 사용자 분류에 관련된 정보를 수신하는 컴퓨터-실행가능 수단;

스팸 필터의 트레이닝 및 스팸 리스트를 채우는 것과 관련하여 정보를 사용하는 컴퓨터-실행가능 수단 -상기 스팸 필터의 트레이닝은 기계 학습 기술을 통해 이용됨-; 및

폴링을 위해 태그된 메시지를 폴링 메시지로 식별하도록 수정하는 컴포넌트- 수정된 상기 메시지는 표결 명령어(voting instructions), 및 상기 사용자에 의한 상기 메시지의 분류를 용이하게 하는 적어도 2개의 각각의 메시지 클래스에 대응하는 적어도 2개의 표결 버튼 및 링크 중 어느 하나를 포함하고, 상기 표결 버튼들은 각각의 링크들에 대응하여서 상기 표결 버튼들 중 임의의 하나가 상기 사용자로부터의 입력에 의해 선택되었을 때, 선택된 상기 표결 버튼에 관련된 정보, 각각의 상기 사용자, 및 해당하는 사용자에 할당된 상기 메시지의 고유 ID가 저장을 위해 데이터베이스에 전송됨-;

을 포함하는 시스템.

청구항 82

삭제

청구항 83

삭제

청구항 84

삭제

청구항 85

삭제

청구항 86

삭제

청구항 87

삭제

청구항 88

삭제

청구항 89

삭제

청구항 90

삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

명 세 서

기 술 분 야

[0001] 본 발명은 합법적인 정보(예를 들어, 정당한 메일)와 원하지 않는 정보(예를 들어, 정크 메일)를 식별하는 시스템 및 방법에 관한 것으로, 더욱 구체적으로 스팸 방지를 위해 전자 메일 서신을 분류하는 것에 관한 것이다.

배 경 기 술

[0002] 인터넷과 같은 글로벌 통신 네트워크의 출현은 거대한 수의 고객들에게 도달할 수 있는 상업적인 기회를 제공했다. 전자 메시징, 특히 전자 메일("이메일")은 원하지 않는 광고 및 선전용 자료를 네트워크 사용자에게 유포시키는 수단으로서 점점 더 널리 퍼지고 있다.

[0003] 컨설팅 및 마켓 리서치 회사인 Radicati Group, Inc.는 2002년 8월 현재, 20억의 정크 이메일 메시지가 매일 보내지고 있으며, 이 수는 2년마다 3배가 될 것으로 예측하고 있다. 개인 및 엔티티(예를 들어, 판매 대리점 및 정부 기관)는 정크 메시지에 의해 점점 더 불편해지고, 자주 불쾌해지게 되었다. 이와 같이 정크 이메일은 현재 또는 곧, 신뢰할 수 있는 컴퓨팅에 가장 큰 위협이 될 것이다.

[0004] 정크 이메일을 방해하는데 사용된 핵심적인 기술은 필터링 시스템/방법론의 사용이다. 한가지 증명된 필터링 기술은 기계 학습 방법에 기초하고 있으며-기계 학습 필터는 메시지가 정크일 확률을 착신 메시지에 부여한다.

이러한 방법에서, 2가지 종류의 예시적인 메시지(예를 들어, 정크 메시지와 비정크 메시지)로부터 전형적으로 특징이 추출되고, 학습 필터는 2가지 종류 사이에서 개연적으로 판별하기 위해 적용된다. 다수의 메시지 특징이 내용(예를 들어, 제목 및/또는 메시지 본문 안의 단어 및 구)에 관련되기 때문에, 이러한 유형의 필터는 통상 "내용 기반의 필터"라고 칭해진다.

[0005] 몇몇의 정크/스팸 필터들이 적응될 수 있는데, 이것은 여러나라 말을 하는 사용자들 및 희귀한 언어를 쓰는 사용자들이 그들의 특정 요구에 적응할 수 있는 필터를 필요로 한다는 점에서 중요하다. 더우기, 모든 사용자들이 정크/스팸인 것 및 정크/스팸이 아닌 것에 대하여 의견을 같이하는 것은 아니다. 따라서, 암시적으로(예를 들어, 사용자 행위를 유지함으로써) 트레이닝될 수 있는 필터를 사용함으로써, 각각의 필터는 사용자의 특정 메시지 식별 요구에 부합하도록 동적으로 맞추어질 수 있다.

[0006] 필터링 적응을 위한 한가지 방법은 사용자에게 메시지를 정크 및 비정크로 명명하도록 요청하는 것이다. 불행하게도, 이러한 수동적인 강도높은 트레이닝 기술은 이러한 트레이닝을 적절하게 실행하는데 요구된 시간량은 말할 것도 없고 이러한 트레이닝과 연관된 복잡성으로 인해 많은 사용자들에게 바람직하지 않다. 또한, 이러한 수동적인 트레이닝 기술은 개인 사용자들에 의해 자주 무효로 된다. 예를 들어, 무료 메일링 리스트로의 가입은 사용자들에 의해 종종 잊혀지므로, 정크 메일로서 잘못 명명된다. 그 결과, 합법적인 메일이 사용자의 메일 박스로부터 무기한으로 차단된다. 또 다른 적응 필터 트레이닝 방법은 암시적인 트레이닝 큐(cue)를 사용하는 것이다. 예를 들어, 사용자(들)이 메시지에 응답하거나 메시지를 보내면, 이 방법은 메시지가 비정크인 것으로 추정한다. 그러나, 이런 종류의 메시지 큐만을 사용하는 것은 통계적 바이어스를 트레이닝 프로세스에 도입하게 되어, 각각 더 낮은 정확도의 필터들을 초래하게 된다.

[0007] 또 다른 방법은 트레이닝을 위해 모든 사용자(들) 이메일을 사용하는 것인데, 초기 레이블은 현존하는 필터에 의해 지정되고, 사용자(들)은 언젠가 명시된 큐(예를 들어, "사용자-정정" 방법)-예를 들어, "정크로서 삭제" 및 "비정크"와 같은 옵션을 선택하는 것- 및/또는 암시적인 큐로 그 지정을 무효로 한다. 이러한 방법은 이전에 설명된 기술보다는 양호하지만, 후술되는 본 발명에 비해 여전히 불충분하다.

[0008] <발명의 요약>

[0009] 다음은 본 발명의 몇몇 실시양상에 관한 기본적인 이해를 제공하기 위해 본 발명에 대한 간략한 개요를 나타낸다. 이 개요는 본 발명의 광범위한 개략은 아니다. 이것은 본 발명의 핵심적인/중요한 요소를 확인하거나 본 발명의 범위를 나타내고자 하는 것이 아니다. 이것의 고유 목적은 후술되는 더욱 상세한 설명에 대한 서론으로서 본 발명의 몇몇 개념을 간략한 형태로 나타내고자 하는 것이다.

[0010] 본 발명은 스팸 방지와 관련하여 아이템 분류를 용이하게 하는 피드백 루프 시스템 및 방법을 제공한다. 본 발명은 스팸 필터에 적용된 기계-학습 방법을 사용하고, 특히, 트레이닝 데이터 세트를 생성하기 위해 합법적 메일 및 정크/스팸 메일의 예들이 얻어지도록 착신 이메일 메시지를 임의로 샘플링한다. 미리 선택된 개인들은 스팸 파이터의 역할을 하고, 샘플들의 각각의 복제(선택적으로 약간 변경될 수 있음)를 분류하는데 참여한다.

[0011] 일반적으로, 폴링(polling)을 위해 선택된 메시지는 폴링 메시지로서 출연하기 위해 여러가지 양상으로 변경된다. 본 발명의 고유 양상은 어떤 사용자들(스팸 파일터들)이 동일한 메시지(예를 들어, 메시지 내용면에서)를 두번- 한번은 폴링 메시지의 형태로, 또 한번은 원래의 형태로- 수신할 수 있도록, 폴링을 위해 선택된 착신 메시지의 카피가 이루어지는 것이다. 본 발명의 다른 고유 양상은, 현존하는 필터에 의해 레이블되어 있는 것들을 포함하여, 모든 메시지가 폴링을 위해 고려되는 것이다. 스팸-레이블된 메시지는 폴링을 위해 고려되고, 선택되는 경우, 현존하는 필터의 명세(예를 들어, 정크 폴더로의 이동, 삭제 …)에 따른 스팸으로서 처리되지 않는다.

[0012] 종래의 스팸 필터와 달리, 정당한 메일과 스팸 간의 구별을 학습하도록 본 발명의 피드백 기술에 따라 스팸 필터를 트레이닝함으로써 더욱 정확한 스팸 필터가 작성될 수 있고, 이에 의해 바이어스되고 부정확한 필터링이 경감된다. 피드백은 소정의 적절한 수의 사용자들의 착신 이메일에 관한 피드백을 얻기 위해 그 사용자들을 폴링함으로써 적어도 부분적으로 달성된다. 스팸 파이터로서 식별된 사용자들에게는 착신 메시지의 선택이 합법적 메일인지 정크 메일인 지를 표결하는 임무가 주어진다. 착신 이메일의 긍정적 및 부정적 분류는 사용자를 위해 준비된 정당한(예를 들어, 스팸이 아닌) 메일을 스팸으로 부적절하게 필터링해내는 것을 경감시키기 위해 요구된다. 각 메일 트랜잭션(transaction)과 연관된 소정의 다른 정보와 함께 각각의 분류가 데이터베이스로 이동되어 스팸 필터를 트레이닝하는 것을 용이하게 한다. 데이터베이스 및 관련된 구성요소들은 기계 학습 시스템을 위한 트레이닝 데이터의 세트를 생성하기 위해, 사용자 특성, 사용자 투표 정보 및 이력, 각각의 선택된

메시지에 할당된 고유 식별 번호와 같은 메시지 특성, 메시지 분류, 및 메시지 내용 요약, 또는 소정의 상술된 것과 관련된 통계적 데이터를 포함하는 선택된 메시지(들)에 대한 특성(또는 선택된 메일 트랜잭션)을 컴파일하여 저장할 수 있다. 기계 학습 시스템(예를 들어, 신경망, 서포트 벡터 기계(SVM), Bayesian Belief Network)은 합법적 메일 및 스팸 메일을 인식하고, 더 나아가 그들을 구별하도록 트레이닝된 향상된 스팸 필터의 작성을 용이하게 한다. 일단 새로운 스팸 필터가 본 발명에 따라 트레이닝되었으면, 그것은 서버 및 클라이언트 이메일 소프트웨어 프로그램을 메일링하도록 배포될 수 있다. 더우기, 새로운 스팸 필터는 개인전용 필터(들)의 성능을 향상시키기 위해 특정 사용자(들)에 관련하여 트레이닝될 수 있다. 새로운 트레이닝 데이터 세트가 만들어짐에 따라, 스팸 필터는 그 성능 및 정확도를 최적화하기 위해 기계 학습을 통해 그 이상의 트레이닝을 받을 수 있다. 메시지 분류를 통한 사용자 피드백은 또한 스팸 필터용 리스트 및 부모 제어를 생성하기 위해, 스팸 필터 성능을 테스트하기 위해, 그리고/또는 스팸 발신지를 확인하기 위해 사용될 수 있다.

[0013] 본 발명의 다른 양상은 상호-확인 기술을 통해 그리고/또는 공지된 결과 테스트 메시지에 의해 신뢰할 수 없는 사용자를 검출하는 방법을 제공한다. 상호-확인은 몇몇 사용자들의 폴링 결과가 배제되는 필터를 트레이닝하는 것과 관련된다. 즉, 필터는 사용자들의 서브셋으로부터의 폴링 결과를 사용하여 트레이닝된다. 평균적으로, 이러한 사용자들의 서브셋은 약간의 실수로 일반적으로 그들과 일치하지 않은 사람들을 검출하더라도 상당히 잘 해낼 수 있다. 배제된 사용자들로부터의 폴링 결과는 트레이닝된 필터의 것들과 비교된다. 이 비교는 기본적으로, 트레이닝 서브셋으로부터의 사용자들이 배제된 사용자들에게 속하는 메시지를 표결할 수 있는 방법을 결정한다. 배제된 사용자들의 표결과 필터 간의 일치가 낮으면, 그 사용자들로부터의 폴링 결과는 폐기되거나 또는 수동적 조사를 위해 표시될 수 있다. 이 기술은 매번 상이한 사용자들로부터의 데이터를 제외하고, 원하는 만큼 반복될 수 있다.

[0014] 개별 메시지에 관한 실수는 필터와 사용자 표결이 강하게 불일치하는 메시지와 같이 검출될 수 있다. 이들 메시지는 자동 제거 및/또는 수동 조사로 플래그될 수 있다. 상호-확인에 대한 대안으로서, 필터는 모든 또는 거의 모든 사용자에 대해 트레이닝될 수 있다. 필터와 불일치하는 사용자 표결 및/또는 메시지는 폐기될 수 있다. 상호-확인에 대한 다른 대안은 사용자가 결과가 알려져 있는 메시지(들)을 표결하라는 요청을 받는 공지된 결과 테스트 메시지에 관련된다. 사용자에 의한 메시지의 정확한 분류(예를 들어, 사용자 표결이 필터 동작에 부합)는 사용자의 신뢰성을 검증하고, 사용자의 분류를 트레이닝으로부터 제거할 것인지의 여부 및 사용자를 장래의 폴링으로부터 제거할 것인지의 여부를 판정한다.

[0015] 본 발명의 또 다른 실시양상은 착신 메일을 스팸으로 식별하고/하거나 특정 상인 이메일 어드레스 프로세싱을 트랙하기 위해 공지된 스팸 타겟(예를 들어, 허니팟(honeypots))의 작성을 제공한다. 공지된 스팸 타겟, 또는 허니팟은 합법적인 메일의 세트가 결정될 수 있는 이메일 어드레스이고, 다른 모든 메일은 스팸이 고려될 수 있다. 예를 들어, 이메일 어드레스는 사람들에 의해 발견될 수 없는 제한된 방식으로 웹사이트 상에 개시될 수 있다. 따라서, 이 어드레스에 보내진 소정의 메일은 스팸이 고려될 수 있다. 대안적으로, 이메일 어드레스는 합법적인 메일이 수신되기로 되어 있는 상인에게만 개시되어 있을 수 있다. 그러므로, 상인으로부터 수신된 메일은 합법적인 메일이지만, 수신된 다른 모든 메일은 안전하게 스팸이 고려될 수 있다. 허니팟 및/또는 다른 소스(예를 들어, 사용자들)로부터 얻어진 스팸 데이터는 피드백 루프 시스템 내로 통합될 수 있지만, 허니팟으로 인한 스팸 분류의 상당한 증가때문에, 그러한 데이터는 더욱 상세하게 후술되는 바와 같이, 바이어스된 폴링 결과를 얻는 것을 경감시키기 위해 다운 웨이팅(down weighting)되어야 한다.

[0016] 본 발명의 다른 실시양상은 피드백 루프 시스템에 의해 또는 필터에 의해 불확실한 것으로 간주되는 메시지를 차단하는 것을 제공한다. 그러한 메시지는 폐기되거나, 분류되는 대신에 소정의 적절한 기간 동안 보유된다. 이 기간은 미리 설정될 수 있거나, 또는 메시지는 예를 들어 동일한 IP 어드레스로부터의 또는 유사한 내용을 갖는, 메시지와 유사한 결정된 수의 폴 결과의 수신시까지 보유될 수 있다.

[0017] 상술한 것 및 관련된 목적을 달성하기 위해, 본 발명의 소정의 예시적인 실시양상은 다음의 상세한 설명 및 첨부된 도면과 관련하여 여기에서 설명된다. 그러나, 이들 실시양상은 본 발명의 원리가 사용될 수 있는 여러가지 방식 중의 일부만을 나타내는 것이고, 본 발명은 그러한 모든 실시양상 및 그 등가물을 포함하도록 이루어진 것이다. 본 발명의 다른 장점 및 신규한 특징은 도면과 관련하여 설명된 다음의 상세한 설명으로부터 명백해질 것이다.

발명의 상세한 설명

[0029] 본 발명은 이제 도면을 참조하여 설명되는데, 동일한 구성요소에는 동일한 참조부호가 사용된다. 다음 설명에 있어서, 다수의 특정 상세는 본 발명의 완벽한 이해를 제공하기 위해 설명된다. 그러나, 본 발명은 이들 특정

상세없이도 실시될 수 있다는 것은 명백하다. 다른 경우에, 널리 알려진 구조 및 장치는 본 발명의 설명을 용이하게 하기 위해 블록도 형태로 도시된다.

[0030] 이 출원에서 사용된 바와 같이, "컴포넌트" 및 "시스템"이라는 용어는 컴퓨터 관련된 엔티티, 및 하드웨어, 하드웨어와 소프트웨어의 조합, 소프트웨어, 또는 실행시의 소프트웨어를 칭하고자 사용된 것이다. 예를 들어, 컴포넌트는 프로세서에서 실행되는 프로세스, 프로세서, 오브젝트, 실행가능부, 실행 스레드, 프로그램, 및/또는 컴퓨터일 수 있는데, 이것에 제한되는 것은 아니다. 실례로서, 서버에서 실행되는 어플리케이션 및 서버는 컴포넌트일 수 있다. 하나 이상의 컴포넌트는 프로세스 및/또는 실행 스레드 내에 상주(populate)할 수 있고, 컴포넌트는 하나의 컴퓨터 상에 국한될 수 있고/있거나 2개 이상의 컴퓨터들 사이에서 분산될 수 있다.

[0031] 본 발명은 기계 학습형 스팸 필터링을 위한 트레이닝 데이터를 생성하는 것과 관련하여 다양한 추론 스킴 및/또는 기술을 포함할 수 있다. 여기에서 사용된 바와 같이, "추론(inference)"이라는 용어는 이벤트 및/또는 데이터를 통해 획득된 한 세트의 관찰결과로부터 시스템, 환경 및/또는 사용자에 대해 추론하거나, 또는 그 상태를 추론하는 프로세스를 일반적으로 칭하는 것이다. 추론은 특정 문맥 또는 동작을 식별하기 위해 사용될 수 있거나, 또는 예를 들어 상태들에 관한 확률 분포를 생성할 수 있다. 추론은 개연적으로 될 수 있고-즉, 데이터 및 이벤트의 고려에 기초하여 관심있는 상태에 관한 확률 분포의 계산결과일 수 있다. 이러한 추론은 이벤트가 시간적으로 근접하게 상관되었든 상관되지 않았든, 그리고 이벤트 및 데이터가 하나의 이벤트 및 데이터 소스로부터 나왔든 몇개의 이벤트 및 데이터 소스로부터 나왔든, 한 세트의 관찰된 이벤트 및/또는 저장된 이벤트 데이터로부터 새로운 이벤트 또는 동작의 구성을 초래한다.

[0032] 메시지라는 용어가 명세서 전반에서 광범위하게 사용되었지만, 이러한 용어는 그 자체로서 전자 메일에 제한되지 않지고, 소정의 적절한 통신 아키텍처를 통해 분산될 수 있는 소정 형태의 전자 메시징을 포함하도록 적절하게 적응될 수 있다. 예를 들어, 원하지 않는 텍스트가 사용자가 메시지를 교환할 때 통상의 채트 메시지 내로 전자적으로 산재될 수 있고/있거나 개시 메시지, 종료 메시지, 또는 상기 모든 메시지로서 삽입될 수 있기 때문에, 두명 이상의 사람들 사이의 회의를 용이하게 하는 회의 어플리케이션(예를 들어, 대화식 채트 프로그램 및 즉시 메시징 프로그램)은 또한 여기에 개시된 필터링 이점을 이용할 수 있다. 이러한 특정 어플리케이션에서, 필터는 바람직하지 않은 내용(예를 들어, 상업용, 선전용, 광고용)을 포착하여 정크로서 태그하기 위해 특정 메시지 내용(텍스트 및 이미지)을 자동으로 필터링하도록 트레이닝될 수 있다.

[0033] 본 발명에서, "수신인(recipient)"라는 용어는 착신 메시지 또는 아이템의 어드레스를 칭하는 것이다. "사용자"라는 용어는 여기에 개시된 피드백 루프 시스템 및 프로세스에 참여하기 위해 수동적으로 또는 능동적으로 선택된 수신인을 칭하는 것이다.

[0034] 이제 1A를 참조하면, 본 발명의 실시양상에 따른 피드백 트레이닝 시스템(10)의 일반적인 블록도가 도시되어 있다. 메시지 수신 컴포넌트(12)는 착신 메시지(IM으로 표시됨)를 수신하여 예정된 수신인(14)에게 전달한다. 메시지 수신 컴포넌트는 바람직하지 않은 메시지(예를 들어, 스팸)의 전달을 경감시키기 위해 다수의 메시지 수신 컴포넌트에서 관례적인 적어도 하나의 필터(16(예를 들어, 정크 필터)를 포함할 수 있다. 필터(16)와 관련된 메시지 수신 컴포넌트(12)는 메시지(IM)를 프로세스하여, 메시지의 필터링된 서브셋(IM')를 예정된 수신인(14)에게 제공한다.

[0035] 본 발명의 피드백 실시양상의 일부로서, 폴링 컴포넌트(18)는 모든 착신 메시지(IM)를 수신하고, 각각의 예정된 수신인(14)을 식별한다. 폴링 컴포넌트는 착신 메시지의 서브셋(IM"로 표시됨)을 예를 들어, 스팸 또는 비 스팸으로 분류하기 위해 예정된 수신인(14)(스팸 파이터(20)로 언급됨)의 서브셋을 선택한다. 분류 관련 정보(VOTING INFO로 표시됨)는 메시지 저장부/표결 저장부(22)에 제공되고, 이 저장부에는 각각의 IM"의 카피뿐만 아니라 표결 정보가 피드백 컴포넌트(24)에 의해 나중에 사용하기 위해 저장된다. 특히, 피드백 컴포넌트(24)는 기계 학습 기술(예를 들어, 신경망, SVM, Bayesian 네트워크, 또는 본 발명에 따라 사용하기 적절한 소정의 기계 학습 시스템)을 사용하고, 이러한 기계 학습 시스템은, 예를 들어 스팸 메일의 식별과 관련하여 필터(16)(및/또는 설정된 새로운 필터(들))를 트레이닝 및/또는 향상시키도록 표결 정보를 사용한다. 착신 메시지의 새로운 스트림이 새로 트레이닝된 필터(16)를 통해 프로세스되면, 보다 적은 스팸 및 더욱 합법적인 메시지(IM'로 표시됨)가 예정된 수신인(14)에게 전달된다. 그러므로, 시스템(10)은 스팸 파이터(20)에 의해 생성된 피드백을 사용함으로써 향상된 스팸 필터의 트레이닝 및 스팸의 식별을 용이하게 한다. 본 발명의 이러한 피드백 양상은 스팸 검출 시스템을 품위있게 하는 풍부하고 매우 동적인 스킴을 제공한다. 본 발명의 더욱 대략적인 양상에 관한 여러가지 상세는 후술된다.

[0036] 이제 도 1B를 참조하면, 본 발명에 따른 스팸 파이팅 및 스팸 방지와 관련된 피드백 루프 트레이닝 흐름도(10

0)가 도시되어 있다. 트레이닝 프로세스의 준비시에 및/또는 그 이전에, 사용자는 (예를 들어, 모든 이메일 사용자를 포함하는 마스터 세트로부터) 스팸 파이터로 선택되는데- 선택은 랜덤 샘플링, 또는 신뢰 레벨, 또는 본 발명에 따른 소정의 적절한 선택 스킴/기준에 기초하여 이루어질 수 있다. 예를 들어, 선택된 사용자들의 서브셋은 모든 사용자, 임의로 선택된 사용자들 세트, 스팸 파이터로서 선택된 사람들, 또는 탈퇴되지 않은 사람들, 및/또는 소정의 이들의 조합을 포함할 수 있고/있거나, 부분적으로 이들의 인구통계학 위치 및 관련된 정보에 기초할 수 있다.

[0037] 대안적으로, 선택된 이메일 사용자의 마스터 세트는 스패머(spammer)가 본 발명을 파괴시키기 위해 더욱 많은 돈을 들일 수 있는 유효 사용자에게 제한될 수 있다. 그러므로, 스팸 파이팅에 참여하도록 선택된 사용자들의 서브셋은 유효 사용자로만 구성될 수 있다. 선택된 사용자들(예를 들어, 스팸 파이터)의 이름 및 특성을 포함하는 리스트 또는 고객 목록은 이때 작성될 수 있다.

[0038] 착신 메시지(102)의 스트림이 수신되면, 104에서 모든 스팸 파이터의 리스트에 대해 각 메시지의 수신인이 체크된다. 수신인이 리스트 상에 있으면, 메시지는 폴링을 위해 고려된다. 그 다음에, 폴링용 메시지를 선택할 것인 지에 대한 판정이 행해진다. 종래의 스팸 필터와 달리, 본 발명은 적어도 모든 착신 메일이 폴링을 위해 고려될 때까지 소정의 메시지(예를 들어, 스팸)를 삭제하지 않는다. 즉, 메일은 소정의 라벨링(예를 들어, 스팸, 비스팸)이 되기 이전에 분류되고, 이것은 사용자 폴링에 이용가능한 메시지의 바이어스없는 샘플의 획득을 용이하게 한다.

[0039] 메시지 선택을 위한 컴포넌트(도시되지 않음)는 데이터의 바이어스를 경감시키기 위한 어떤 랜덤한 확률을 갖는 메시지를 선택하기 위해 사용될 수 있다. 다른 방법은 다른 사용자/수신인 속성 및 특성뿐만 아니라 인구통계 정보를 사용하는 것과 관련된다. 그러므로, 메시지는 적어도 부분적으로 사용자/수신인에 기초하여 선택될 수 있다. 메시지를 선택하기 위한 다른 대안적인 알고리즘이 존재한다. 그러나, 거기에는 사용자 당, 또는 기간 당 사용자 당 선택된 메시지의 수에 제한이 있을 수 있거나, 또는 소정의 주어진 사용자로부터의 메시지를 선택할 확률에 제한(즉, 폴링 제한(polling limitation))이 있을 수 있다. 이러한 제한이 없으면, 스패머는 어카운트(account)를 작성하여, 수백만의 스팸 메시지를 보내고, 이러한 모든 메시지를 정당한 메시지로 분류할 것이고: 이것은 스패머가 부정확하게 레이블된 메시지로 트레이닝 데이터베이스를 훼손시킬 수 있게 할 것이다.

[0040] 특히 블랙 홀 리스트로 언급된 소정의 스팸 필터링은 스킵할 수 없게 될 수 있다. 블랙 홀 리스트는 서버가 IP 어드레스의 리스트로부터 소정 메일을 수신하지 못하게 한다. 그러므로, 메시지의 선택은 블랙 홀 리스트로부터가 아닌 메일 세트로부터 선택될 수 있다.

[0041] 본 발명의 고유 양상은 현재 그 자리에서 필터에 의해 스팸으로서 표시되는 폴링을 위해 선택된 메시지가 삭제되지 않거나 또는 정크 메일 폴터로 이동되지 않는다는 것이다. 그 대신에, 이들은 다른 모든 메시지가 폴링을 고려하여 수신되는 통상의 인박스(inbox) 또는 메일박스 내에 배치된다. 그러나, 메시지의 2개의 카피가 있고, 메시지가 필터에 의해 스팸으로 간주되면, 하나의 카피는 스팸 폴더에게 전달되거나, 그렇지 않으면 설정된 파라미터(예를 들어, 삭제, 특별 표시, 또는 정크 폴더로 이동)에 따라 처리된다.

[0042] 하나의 메시지가 선택되면, 이것은 사용자에게 보내지고, 이것이 폴링 메시지라는 것을 나타내기 위해 어떤 특별한 방식으로 표시된다. 특히, 선택된 메시지는 메시지 변경 컴포넌트(106)에 의해 변경될 수 있다. 메시지 변경의 예는 분리된 폴터 내에 폴링 메시지를 위치시키는 것, '프롬(from)' 어드레스 또는 서브젝트 라인을 변경하는것, 및/또는 사용자에게로의 폴링 메시지와 같은 메시지를 식별할 수 있는 특별 아이콘 또는 특별 컬러를 사용하는 것을 포함하는데, 이것에 제한되는 것은 아니다. 선택된 메시지는 또한 캡슐화된 메시지를 표결 및/또는 분류하는 방법에 관해 사용자에게 명령어를 제공할 수 있는 다른 메시지 내에 캡슐화될 수 있다. 이들 명령어는 적어도 2개의 버튼 또는 링크: 예를 들어, 하나는 메시지를 스팸으로서 표결하기 위한 것 및 하나는 메시지를 비 스팸으로서 표결하기 위한 것을 포함할 수 있다.

[0043] 표결 버튼은 폴링 메시지의 한 카피를 사용자에게 보내기 전에 메시지의 내용을 변경함으로써 실현될 수 있다. 본 발명이 클라이언트 이메일 소프트웨어(메일 서버에 반대됨)에 관련하여 사용될 때, 사용자 인터페이스는 표결 버튼을 포함하도록 변경될 수 있다.

[0044] 게다가, 폴링 메시지는 거기에 부착된 선택된 메시지뿐만 아니라 명령어 및 표결 버튼을 포함할 수 있다. 폴링 메시지는 또한 서브젝트 라인, 프롬 어드레스, 송신 및/또는 수신된 날짜, 및 텍스트 또는 적어도 처음 몇줄의 텍스트와 같은 선택된 메시지의 요약을 포함할 수 있다. 다른 방법은 미리 부착된 표결 버튼 및 표결 명령어를 가진 메시지를 보내는 것과 관련된다. 실제로, 사용자가 폴링 메시지의 카피를 오픈 및/또는 다운로드하면, "

스팸" 및 "비 스팸"을 포함하고 이것에 제한되지 않는 버튼(또는 링크)은 사용자 인터페이스 상에 나타나거나, 폴링 메시지 내로 편입될 수 있다. 그러므로, 각각의 폴링 메시지가 한 세트의 명령어 및 적절한 표결 버튼을 포함하는 것이 가능하다. HTML 백그라운드 명령어(명령어 또는 버튼의 텍스트를 불명료하게 할 수 있음)를 되도록이면 제거하는 것을 포함하는 다른 변경이 필요하게 될 수 있다.

[0045] "요청된 상인 이메일"과 같은 다른 버튼은 또한 원하는 정보의 유형에 따라 제공될 수 있다. 메시지는 또한 장래의 폴링에서 탈퇴하는 버튼/링크를 포함할 수 있다. 명령어는 사용자가 선호하는 언어로 국한되고, 폴링 메시지 내에 내장될 수 있다.

[0046] 더우기, 폴링을 위해 선택된 메시지는 메시지 변경 컴포넌트(106)에 의해 또는 소정의 다른 적절한 바이러스 컴포넌트(도시되지 않음)에 의해 바이러스에 대해 스캔될 수 있다. 바이러스가 발견되면, 바이러스가 제거되거나, 메시지가 폐기될 수 있다. 바이러스 제거는 메시지가 선택되는 때 및 사용자가 메시지를 다운로드하기 바로 전을 포함하여, 시스템(100)의 소정의 시점에서 발생할 수 있다.

[0047] 메시지 변경 다음에, 메시지 전달 컴포넌트(108)는 폴링 메시지를 표결하기 위한 사용자에게 전달한다. 사용자 피드백(예를 들어, 폴링 메시지, 사용자의 표결, 및 이것과 연관된 소정의 사용자 특성)에는 고유 식별자(ID)(110)(예를 들어, 메타데이터)가 할당된다. ID(110) 및/또는 이것에 대응하는 정보는 사용자 분류/표결이 컴파일링되어 저장되는 메시지 저장부/표결 저장부(112)(예를 들어, 중앙 데이터베이스)에 제공된다.

[0048] 데이터베이스 레벨에서, 폴링에 이용가능한 선택된 메시지는 나중에 폴링 또는 사용하기 위해 유지될 수 있다. 또한, 데이터베이스는 특정 사용자가 지나치게 샘플링되지 않게 하고 데이터량이 사용자에 의해 지정된 한도 내에서 사용자로부터 수집할 수 있게 하기 위해 시간을 맞춰 분석을 자주 실행할 수 있다. 특히, 피드백 시스템(100)은 샘플링 및 데이터의 바이어스를 경감시키기 위해 샘플링 기간뿐만 아니라 사용자 메일의 퍼센트 한계를 모니터한다. 이것은 사용자들이 낮은 사용량과 높은 사용량을 포함한 모든 이용가능 사용자로부터 선택되는 경우에 특히 중요하다. 예를 들어, 낮은 사용량의 사용자는 전형적으로 높은 사용량의 사용자에 비해 상당히 낮은 분량의 메일을 수신하고 송신한다. 그러므로, 시스템(100)은 선택된 메시지가 T개의 메시지마다 그중의 대략 하나의 메시지가 되고 사용자에 의해 Z 시간마다 수신된 하나보다 많지 않은 메시지가 되도록 메시지 선택 프로세스를 모니터한다. 따라서, 시스템은 예를 들어, 샘플링되도록(예를 들어 폴링을 위해 고려되도록) 10개의 착신 메시지마다 그중의 하나를, 하지만 2시간마다 하나를 초과하지 않게 폴링할 수 있다. 빈도, 또는 퍼센트 한계는 높은 사용량의 사용자에 비해 낮은 사용량의 사용자에 대해 불균형한 양의 메시지를 샘플링하는 것을 경감시키고, 또한 사용자를 귀찮게 하는 오버레이(overlay)를 경감시킨다.

[0049] 빈번하게, 중앙 데이터베이스(112)는 폴링을 위해 시스템(100)에 의해 샘플되었지만 분류되지는 않은 메시지에 대해 스캔한다. 데이터베이스는 이들 메시지를 획득하고, 이들을 각 사용자의 인구통계학 특성에 관련하여 배치하며, 폴링 메시지를 작성하여 사용자(들)에게 메시지(들)을 표결하고 분류하라고 요청한다. 그러나, 스팸 필터는 모든 새로운 착신 분류의 수신 직후에 변경되거나 트레이닝될 수 없다. 오히려, 오프라인 트레이닝은 트레이너가 예정에 따라, 진행에 따라, 또는 날마다 데이터베이스(112) 내로 수신된 데이터를 계속 관찰할 수 있게 한다. 즉, 트레이너는 규정된 개시 시점에서 또는 과거에 설정된 양의 시간에서 개시하고, 필터를 트레이닝하도록 진행된 그 시점으로부터 모든 데이터를 관찰한다. 예를 들어, 규정된 기간은 자정에서 오전6시까지일 수 있다.

[0050] 새로운 스팸 필터는 기계 학습 기술(114)(예를 들어, 신경망, SVM)을 통해 데이터베이스(112) 내에 유지된 메시지 분류를 분석함으로써 진행에 따라 트레이닝될 수 있다. 기계 학습 기술은 정당한 메일과 스팸을 구별할 수 있도록 학습하기 위해 정당한 메일과 스팸의 두가지 예를 필요로 한다. 공지된 예의 스팸의 매칭에 기초한 대등한 기술은 이들이 우연히 정당한 메일을 포착하지 않는다는 것을 확신할 수 있도록, 정당한 메일의 예를 갖는다는 것으로부터 이익을 얻을 수 있다.

[0051] 따라서, 타당한 불평 대신에, 포지티브 및 네가티브 스팸의 예를 갖는 것이 중요하다. 무료 메일링 리스트와 같은 많은 양의 스팸 및 정당한 메일을 발송하는 몇몇의 도메인들이 있다. 누군가 불평에만 기초해서 시스템을 만들었다면, 이들 도메인으로부터의 모든 메일이 필터링되어 상당한 잘못을 초래할 수 있다. 따라서, 도메인이 많은 양의 정당한 메일을 발송한다는 것을 아는 것은 중요하다. 또한, 사용자들은 그들이 무료 메일링 리스트상에 등록한 것을 잊어버리는 것과 같은 잘못을 종종 한다. 예를 들어, 뉴욕 타임즈와 같은 많은 합법적인 제공자는 합법적인 메일을 발송한다. 소수의 사용자들은 그들이 등록한 것을 잊어버리고 이들 메시지를 스팸으로 분류하는 것을 불평한다. 대부분의 사용자들이 이 메일이 합법적이라는 것을 실현하는 데이터가 없으면, 이 사이트로부터의 메일은 달리 차단될 수 있다.

[0052] 새로운 필터(116)는 분산 컴포넌트(118)에 의해 참여 인터넷 서비스 프로바이더(ISP)를 가로질러 이메일 또는 메시지 서버로, 개별 이메일 클라이언트로, 갱신 서버로, 및/또는 개별 회사의 중앙 데이터베이스로 진행에 따라 분산될 수 있다. 게다가, 피드백 시스템(100)은 폴링을 위해 고려되어 이용된 메시지의 샘플들이 시스템(100)에 의해 수신된 이메일의 실제 배포를 따를 수 있도록 진행에 따라 기능한다. 그 결과, 새로운 스팸 필터를 트레이닝하기 위해 사용된 트레이닝 데이터 세트는 적응 스패머와 관련하여 현재 유지된다. 새로운 필터가 만들어질 때, 폴링 데이터는 오래전에 이것이 얻어질 수 있었던 방법에 기초하여 폐기되거나 또는 다운 웨이팅될 수 있다(예를 들어, 무시될 수 있다).

[0053] 시스템(100)은 메일이 게이트웨이 서버, 이메일 서버, 및/또는 메시지 서버와 같은 서버에서 수신될 때 실현될 수 있다. 예를 들어, 메일이 이메일 서버에 들어오면, 서버는 예정된 수신인의 특성을 조사하여 수신인이 시스템(10)으로 선택되었는 지를 판정한다. 이들의 특성이 그 특성으로서 나타내면, 수신인의 메일은 잠재적으로 폴링에 이용가능하다. 클라이언트 전용 아키텍처가 또한 존재한다. 예를 들어, 클라이언트 메일 소프트웨어는 단일 사용자에 대해 폴링 판정을 행하여 중앙 데이터베이스로 이메일을 전달하거나, 또는 개인전용 필터의 성능을 향상시키기 위해 폴링 정보를 이용할 수 있다. 여기에 설명된 것 이외에, 이러한 시스템(100)의 다른 대안적인 아키텍처가 존재하고, 본 발명의 범위 내에 속하는 것으로 생각된다.

[0054] 이제 도 2를 참조하면, 본 발명의 한 실시양상에 따른 기본적인 피드백 루프 프로세스(200)의 흐름도가 도시되어 있다. 설명을 간단하게 하기 위해, 방법론은 일련의 동작으로서 도시되고 설명되었지만, 소정의 동작들이, 본 발명에 따라, 상이한 순서 및/또는 여기에 도시되고 설명된 것과는 다른 동작들과 동시에 발생할 수 있는 것과 같이, 본 발명은 동작의 순서에 의해 제한되지 않는다는 것을 이해할 수 있을 것이다. 예를 들어, 본 분야에 숙련된 기술자들은 방법론이 대안적으로 상태도에서와 같이 일련의 상관된 상태 또는 이벤트로서 나타내질 수 있다는 것을 이해할 수 있을 것이다. 게다가, 도시된 모든 동작들이 본 발명에 따른 방법을 실현하는데 요구되는 것은 아니다.

[0055] 프로세스(200)가 시작되어, 메일이 들어와서 서버와 같은 컴포넌트에 의해 수신된다(단계 202). 메일이 서버에 도착하면, 서버는 예정된 수신인의 특성을 식별하여, 예정된 수신인이 폴링을 위해 스팸 파이터로서 미리 선택되었는 지를 판정한다(단계 204). 그러므로, 프로세스(200)는 수신인이 피드백 시스템에 선택되었는 지를 나타낼 수 있는 사용자 특성 필드를 사용하거나, 또는 선택된 사용자들의 리스트를 참고한다. 사용자가 피드백 시스템 내의 참가자로 판정되고 폴링을 위해 선택되었으면(단계 206), 피드백 시스템은 폴링을 위해 어떤 메시지가 선택되는 지를 판정함으로써 동작을 행한다(단계208). 그렇지 않으면, 프로세스(200)는 착신 메시지의 적어도 하나의 예정된 수신인이 사용자(예를 들어, 스팸 파이터)인 것으로 판정될 때까지 단계 202로 돌아간다.

[0056] 실제로, 현재 사용된 필터(예를 들어, 개인전용 필터, 브라이트메일(Brightmail) 필터)에 의해 스팸으로 지정되거나 될 수 있는 메시지들을 포함하여, 모든 메시지가 폴링을 위해 고려된다. 그러므로, 메시지들이 폴링을 위해 고려되기 전에 삭제되거나, 폐기되거나, 정크 폴더로 보내지는 메시지는 없다.

[0057] 서버에 의해 수신된 각 메시지 또는 메일 아이템은 메일 트래잭션에 대응하는 한 세트의 특성을 갖는다. 서버는 이들 특성을 컴파일링하여, 이들를 폴링 메시지와 함께 중앙 데이터베이스로 보낸다. 특성 예는 수신인 리스트(예를 들어, "To:", "cc:", 및/또는 "bcc:" 필드에 리스트됨), 현재 사용된 필터의 판단(예를 들어, 필터가 메시지를 스팸으로 식별했는 지의 여부), 다른 선택 스팸 필터(예를 들어, 브라이트메일 필터), 및 사용자 정보(사용자명, 패스워드, 실명, 폴링된 메시지의 빈도, 사용 데이터, …)를 포함한다. 대응하는 사용자/수신인뿐만 아니라, 폴링 메시지 및/또는 그 내용에는 각각 고유 식별자가 할당된다. 식별자는 또한 데이터베이스에 보내진 다음에, 필요에 따라 갱신될 수 있다.

[0058] 단계 214에서, 폴링을 위해 선택된 메시지(들)(예를 들어, 오리지널 메시지$_{1-M}$(M은 1보다 크거나 같은 정수)는 메시지$_{1-M}$가 폴링 메시지$_{P1-PM}$이고, 다음에 폴링을 위해 사용자에게 전달된다는 것을 사용자에게 나타내기 위해 변경된다(단계 216). 예를 들어, 폴링 메시지는 어태치먼트로서 표결될 오리지널 메시지, 및 메시지를 표결하는 방법에 관한 한 세트의 명령어를 포함할 수 있다. 명령어 세트는 예를 들어, "정당한 메일" 버튼 및 "스팸" 버튼과 같은 적어도 2개의 버튼을 포함한다. 사용자가 메시지를 정당한 메일 또는 스팸으로서 분류하기 위해 버튼들 중의 하나를 클릭하면(단계 218), 사용자는 사용자가 제공하고 있는 분류를 위한 고유 식별자에 대응하는 URL(uniform resorce locator)로 향해진다. 이 정보는 게시되고, 그 오리지널 메시지$_{1-M}$를 위한 중앙 데이터베이스 내의 연관된 기록은 갱신된다.

[0059] 단계 216에서, 또는 프로세스(200) 중의 소정의 다른 적절한 시기에, 오리지널 메시지는 사용자에게 선택적으로

전달될 수 있다. 그러므로, 사용자는 메시지를 두번 수신한다-한번은 원래의 형태로, 또 한번은 변경된 폴링 형태로 수신한다.

[0060] 조금 나중에, 새로운 스팸 필터가 작성되어 적어도 부분적으로 사용자 피드백에 기초하여 트레이닝된다(단계 220). 새로운 스팸 필터가 일단 작성되어 트레이닝되면, 필터는 즉시 이메일 서버에서 사용될 수 있고/있거나, 클라이언트 서버, 클라이언트 이메일 소프트웨어 등으로 분산될 수 있다(단계 222). 새로운 또는 갱신된 스팸 필터의 트레이닝 및 분산은 진행해가는 활동이다. 그러므로, 프로세스는 착신 메시지의 새로운 스트림이 수신될 때 계속된다(단계 204). 새로운 필터가 만들어지면, 기존의 데이터는 폐기되거나, 또는 오래전에 이들이 얻어졌던 방법에 기초하여 가치가 떨어진다.

[0061] 피드백 시스템(100) 및 프로세스(200)는 참가하는 사용자의 피드백에 의존한다. 불행히도, 일부 사용자들은 신뢰받을 수 없거나, 또는 단순히 느려서 일관되고 정확한 분류를 제공하지 못한다. 중앙 데이터베이스(도 1A)는 사용자 분류의 이력을 유지한다. 그러므로, 피드백 시스템(100)은 폴링 메시지에 대한 사용자 응답의 수 또는 빈도뿐만 아니라, 부정의 수, 사용자가 자기의 마음을 바꾼 횟수, 공지된 정당한 메일 또는 공지된 스팸에 대한 사용자의 응답을 트랙할 수 있다.

[0062] 이들 수 중의 소정의 수가 규정된 임계치를 초과할 때, 또는 시스템의 모든 사용자에 대해, 피드백 시스템(100)은 특정 사용자 또는 사용자들의 신뢰성을 평가하기 위해 하나의 또는 몇가지 확인 기술을 호출할 수 있다. 한가지 방식은 본 발명의 다른 실시양상에 따른 도 3에 도시된 상호-확인 방법이다.

[0063] 상호 확인 기술은 단계 302에서 시작하여, 중앙 데이터베이스가 폴링 결과 및 각각의 사용자 정보와 같은 착신 데이터를 수신한다. 다음에, 단계 304에서, 적절한 수의 사용자를 테스트하기 위해 상호-확인이 요구되는 지를 판단한다. 요구되면, 새로운 스팸 필터는 착신 데이터의 일부분을 사용하여 트레이닝된다(단계 306). 즉, 테스트되고 있는 사용자로부터의 데이터는 트레이닝으로부터 제외된다. 예를 들어, 필터는 약 90%의 폴링된 사용자 데이터(90% 필터로 표시됨)로 트레이닝됨으로써, 테스트된 사용자에 의해 제공된 데이터에 대응하는 약 10%의 데이터(10% 테스트된 사용자로 표시됨)를 제외한다.

[0064] 단계 308에서, 90% 필터는 나머지 10% 테스트된 사용자 데이터에 대해 실행되어, 90% 사용자가 테스트된 사용자의 메시지를 표결할 것인 지를 판정한다. 90% 필터와 10% 테스트된 사용자 데이터 간의 불일치량이 규정된 임계치를 초과하면(단계 310), 사용자의 분류는 수동으로 조사될 수 있다(단계 312). 대안적으로 또는 또한, 테스트 메시지는 의심스럽거나 신뢰할 수 없는 사용자에게 보내질 수 있고/있거나, 이들 특정 사용자는 장래의 폴링에서 제외될 수 있고/있거나, 이들의 과거 데이터는 폐기된다. 그러나, 임계치가 초과되지 않으면, 프로세스는 단계 306으로 돌아간다. 실제로, 상호-확인 기술(300)은 표결/분류 데이터의 신뢰성을 판정하여 유지하기 위해 필요에 따라 상이한 사용자를 제외하고, 소정의 적절한 세트의 테스트 사용자로 이용될 수 있다.

[0065] 사용자 적합도 및 신뢰도를 평가하기 위한 두번째 방식은 주어진 기간에 모인 모든 데이터에 대해 필터를 트레이닝한 다음에, 필터를 사용하여 트레이닝 데이터에 대해 테스트하는 것을 포함한다. 이 기술은 테스트-온-트레이닝으로 공지되어 있다. 메시지가 트레이닝 내에 포함되었으면, 필터는 그 비율(rating)을 학습해야 하고, 예를 들어 학습된 필터는 사용자가 했던 것과 동일한 방식으로 메시지를 분류해야 한다. 그러나, 필터는 사용자가 비 스팸으로 레이블했을 때 이것을 스팸으로 레이블함으로써, 또는 그 반대로 함으로써 실수를 계속 할 수 있다. 필터가 그 트레이닝 데이터와 불일치하기 위해, 메시지는 다른 메시지와 강하게 불일치해야 한다. 그렇지 않으면, 트레이닝된 필터는 그것을 정확하게 분류하는 소정의 방식을 거의 정확하게 발견할 수 있었을 것이다. 그러므로, 메시지는 신뢰할 수 없는 레이블을 갖는 것으로 폐기될 수 있다. 이러한 기술 또는 상호 확인이 사용될수 있는데: 상호-확인은 분류시의 더 많은 실수를 보다 덜 확실하게 밝힐 수 있고; 이와 반대로 테스트-온-트레이닝은 더 적은 실수를 더욱 더 확실하게 찾는다.

[0066] 테스트-온-트레이닝 및 상호-확인 기술(300)은 개별 메시지에 적용될 수 있는데, 개별 사용자의 분류 또는 메시지의 비율은 (예를 들어, 대다수의 비율 다음에) 일반적인 일치에 의해 제외된다. 대안적으로, 두가지 기술은 잠재적으로 신뢰할 수 없는 사용자를 식별하는데 사용될 수 있다.

[0067] 상호 확인 및 테스트 온 트레이닝 기술 이외에 또는 그 대신에, 우리는 사용자 신뢰성을 검증하기 위해 "공지된-결과" 기술을 사용할 수 있다(단계 314에서 도 4로 진행). 도 3 및 도 4의 기술이 따로 설명되었지만, 두가지 방식은 동시에 이용될 수 있다는 것을 알 수 있을 것이다. 즉, 공지된 정당한 메시지 및 공지된 스팸 메시지로부터의 정보는 상호-확인 또는 테스트 온 트레이닝 결과와 결합되어 어떤 사용자를 폐기할 것인 지를 판정한다.

[0068] 이제 도 4를 참조하면, 본 발명의 한 실시양상에 따라 사용자 표결의 적합도를 확인하기 위한 프로세스(400)의

흐름도가 도시되어 있다. 프로세스(400)는 도 3에 도시된 단계 314로부터 적용한다. 단계 402에서, 공지된 결과 테스트 메시지(들)은 의심스러운 사용자(들)(또는 모든 사용자들)에게 보내진다. 예를 들어, 테스트 메시지는 착신 메일에 삽입된 다음에, 데이터베이스가 "공지된" 결과를 수신하도록 수작업으로 분류될 수 있다. 그렇지 않으면, 프로세스(400)는 공지된 결과 메시지가 제3자에 의해 보내질 때까지 기다릴 수 있다. 사용자는 동일한 테스트 메시지를 표결하도록 허용될 수 있다. 표결 결과는 공지된 결과에 비교된다(단계 404). 사용자의 표결이 일치하지 않으면(단계 406), 이들의 현재 및/또는 미래 및/또는 과거 분류는 이들이 일관성 및 신뢰성을 설명할 때까지 적절한 기간동안 수작업으로 조사될 수 있다. 대안적으로, 이들의 현재 또는 미래 또는 과거 분류는 무시되거나 제거될 수 있다. 마지막으로, 사용자는 장래의 폴링에서 제거될 수 있다. 그러나, 이들의 표결 결과가 테스트 메시지 결과와 일치하면, 사용자는 신뢰성이 있는 것으로 간주될 수 있다(단계 410). 프로세스(400)는 단계 412에서 도 3으로 돌아가서, 다음 그룹의 의심스러운 사용자들에게 어떤 유형의 확인 기술이 바람직한 지를 판정한다.

[0069] 사용자 신뢰도를 평가하는 네번째 방식은 능동적인 학습이다. 능동적인 학습 기술에 따라, 메시지는 임의로 선택되지 않는다. 그 대신에, 피드백 시스템은 메시지가 시스템에 얼마나 유용해질 수 있는 지를 평가할 수 있다. 예를 들어, 필터가 스팸의 확률을 돌려보내면, 폴링을 위해 현재의 필터에 의해 가장 불확실하게 분류된 메시지, 즉 스팸의 확률이 50%에 가장 가까운 메시지를 우선적으로 선택할 수 있다. 메시지를 선택하는 다른 방식은 메시지가 얼마나 공통되는 지를 판정하는 것이다. 메시지가 더욱 공통되면, 폴링하는 데에 더욱 유용하다. 고유 메시지는 그들이 덜 공통되기 때문에 덜 유용하다. 능동적 학습은 현존 필터의 신용 레벨을 사용하고, 메시지의 특징이 얼마나 공통되는 지를 사용하며, 세팅 또는 컨텐트의 현존 필터의 신용 레벨(예를 들어, 메타컨피던스(emtaconfidence))을 사용함으로써 이용될 수 있다. 기계 학습 분야에 숙련된 기술자들에게 잘 알려진 쿼리-바이-커미티(query-by-committee)와 같은 다수의 다른 능동적 학습 기술이 있으며, 이들 기술 중의 소정의 기술이 사용될 수 있다.

[0070] 이제 도 5를 참조하면, 본 발명의 한 실시양상에 따라 스팸 필터 트레이닝 내로 사용자 피드백 이외에 허니팟 피드백을 편입시키기 위한 프로세스(500)의 흐름도가 도시되어 있다. 허니팟은 이메일을 보내야 하는 공지된 이메일 어드레스이다. 예를 들어, 새로 작성된 이메일 어드레스는 사적기밀로 유지될 수 있고, 선택된 개인들에게만 개시될 수 있다(단계 502). 이들은 또한 공적으로, 하지만 사람들이 볼 수 없는 제한된 방식으로(예를 들어, 메일 링크로서 하얀 활자체에 하얀 배경을 입혀서) 개시될 수 있다. 허니팟은 스패머에 의한 딕셔너리(dictionary) 공격시에 특히 유용하다. 딕셔너리 공격은 스패머가 상당히 많은 수의 어드레스, 즉 아마도 딕셔너리 내의 모든 어드레스, 또는 딕셔너리 내의 단어 쌍들로 이루어진 모든 어드레스를 이메일링하고자 하는 것, 또는 유효 어드레스를 찾기 위한 것과 유사한 기술이다. 허니팟으로 보내진 소정의 이메일(단계 504) 또는 몇몇의 선택된 개인들로부터가 아닌 소정의 이메일(단계 506)은 스팸으로 간주된다(단계 508). 이메일 어드레스는 또한 의심스러운 상인이 등록될 수 있다. 그러므로, 상인으로부터 수신된 소정의 이메일은 정당한 메일로 간주되지만(단계 510), 다른 모든 메일은 스팸으로 간주된다. 따라서 스팸 필터가 트레이닝될 수 있다(단계 512). 게다가, 의심스러운 상인은 제3자에게 사용자 정보(예를 들어, 적어도 이메일 어드레스)를 팔것인지 그렇지 않으면 개시할 것인지 판정된다. 이것은 다른 의심스러운 상인에 대해 반복될 수 있고, 리스트는 사용자에게 그들의 정보가 스패머에게 배포될 수 있다는 것을 경고하기 위해 생성될 수 있다. 이들은 안전하게 스팸으로 간주될 수 있는 허니팟에 보내진 이메일을 얻는 공정한 몇가지 기술이다. 실제로, 안전하게 스팸으로 간주될 수 있는 허니팟에 보내진 이메일을 얻는 다른 대안적인 방식이 있다.

[0071] 허니팟은 스팸의 양호한 소스이지만, 합법적인 메일의 무서운 소스이기 때문에, 허니팟으로부터의 데이터는 새로운 스팸 필터를 트레이닝하기 위해 피드백 루프 시스템(도 1)으로부터의 데이터와 결합될 수 있다. 상이한 소스 또는 상이한 분류로부터의 메일은 다르게 작용될 수 있다. 예를 들어, 10개의 허니팟, 및 메일의 10%에 관해 폴링하는 10명의 사용자가 있다고 하면, 스팸의 약 10배가 폴링으로부터 허니팟에서 예상될 수 있다. 그러므로, 폴링으로부터의 합법적인 메일은 이러한 차이를 벌충하기 위해 스팸의 10배 또는 11배로 웨이팅될 수 있다. 대안적으로, 허니팟은 선택적으로 다운 웨이팅될 수 있다. 예를 들어, 사용자 메일의 약 50%는 정당한 메일이고, 약 50%는 스팸이다. 동일한 분량의 스팸이 허니팟으로 진행한다. 그러므로, 허니팟이 스팸의 100%를 갖는 것처럼 보이고, 그 전부가 샘플링되지만, 단 10%는 그렇지 않다. 결합된 시스템 내에서 스팸과 정당한 메일의 정확한 비율로 트레이닝하기 위해, 허니팟 데이터는 95%만큼 다운 웨이팅되고, 사용자 스팸은 50%만큼 다운 웨이팅되어 1:1의 전체 비를 생기게 한다.

[0072] 스팸 리포트의 다른 소스는 피드백 루프 시스템 내의 참가자로서 포함되지 않은 사용자를 포함한다. 예를 들어, 거기에는 필터를 통해 만든 스팸을 리포트하기 위해 모든 메일에 대해 모든 사용자에게 이용가능한 "리포

트 스팸(Report Spam)" 버튼이 있을 수 있다. 이 데이터는 피드백 루프 시스템으로부터의 데이터와 결합될 수 있다. 또한, 이 스팸 소느는 바이어스되거나 여러가지 양상으로 신뢰할 수 없게 될 수 있기 때문에 다운 웨이팅되거나 다르게 웨이팅되어야 한다. 재웨이팅(re-weighting)은 필터링되지 않은 메일만이 "리포트-에즈-스팸(Report-as-spam)" 버튼에 의해 리포팅된다는 사실을 반영하도록 행해져야 한다.

[0073] 스팸 필터 이외에, 차단(quarantine) 필터가 작성되어 피드백 루프 시스템에 의해 사용될 수 있다. 차단 필터는 긍정적 및 부정적 메일 특징을 사용한다. 예를 들어, 인기있는 온라인 상인으로부터의 메일은 대체로 항상 정당한 메일이다. 스패머는 자기 스팸 내에서 정당한 상인 메일의 양상을 모방함으로써 시스템을 부당하게 이용한다. 다른 예는 스패머가 소량의 정당한 메일을 IP 어드레스를 통해 보냄으로써 의도적으로 피드백 시스템을 속이는 것이다. 피드백은 이 메일을 정당한 메일로 분류하는 학습을 하고, 그때, 스패머는 동일한 IP 어드레스로부터 스팸을 보내기 시작한다.

[0074] 그러므로, 차단 필터는 시스템이 이력 데이터에 기초하여 사용되는 것보다 훨씬 더 많은 양의 특정 긍정적 특징이 수신되고 있다는 것을 인지한다. 이것은 시스템이 메시지를 의심하게 하므로, 메일을 스팸으로서 전달하거나 표시하도록 선택하기 이전에 충분한 폴 결과가 얻어질 때까지 차단된다. 차단 필터는 또한 메일이 스팸인지 스팸이 아닌지 확실하지 않거나 알려져 있지 않고, 잠시동안 알 수 없는 새로운 IP 어드레스로부터 메일이 수신될 때 사용될 수 있다. 차단은 메일을 임시로 스팸으로서 표시하여 이것을 스팸 폴터로 이동하거나, 또는 이것을 사용자에게 전달하지 않거나, 또는 보이지 않을 어딘가에 저장하는 것을 포함하여, 여러가지 방식으로 실행될 수 있다. 차단은 스팸 필터 임계치에 가까운 메시지에 대해 행해질 수 있는데: 폴링으로부터의 추가 정보가 정확한 판정을 도울 수 있다는 것을 추정할 수 있다. 차단은 또한 다수의 유사한 메시지들이 수신될 때 행해질 수 있는데: 소수의 메시지들은 피드백 루프에 따라 폴링을 위해 보내질 수 있고, 재트레이닝된 필터는 메시지를 정확하게 분류하기 위해 사용될 수 있다.

[0075] 필터를 만드는 것 이외에, 여기에 설명된 피드백 루프 시스템은 그것들을 잘 평가하기 위해 사용될 수 있다. 즉, 스팸 필터의 파라미터는 필요에 따라 조정될 수 있다. 예를 들어, 필터는 지난 밤중에 자정까지 트레이닝된다. 자정 이후, 사용자의 분류에 비교하여 스팸 필터의 에러 비율을 판정하기 위해 데이터베이스에서 나오는 데이터를 획득한다. 또한, 피드백 루프는 스팸 필터의 잘못된 포지티브 및 캐치 비율을 판정하기 위해 사용될 수 있다. 예를 들어, 사용자 표결이 행해질 수 있고, 메일은 잘못된 포지티브 및 캐치 비율을 판정하기 위해 잠재적인 필터를 통해 전해질 수 있다. 이 정보는 이때 필터를 조정하여 최적화시키기 위해 사용될 수 있다. 상이한 파라미터 설정 또는 상이한 알고리즘은 몇개의 필터를 만드므로써 수동으로 또는 자동으로 시도될 수 있고, 각각의 필터는 가장 낮은 잘못된 포지티브 및 캐치 비율을 얻기 위해 상이한 설정 또는 알고리즘을 사용한다. 그러므로, 결과가 비교되어, 최상의 또는 최적의 필터 파라미터를 선택할 수 있다.

[0076] 피드백 루프는 항상 스팸으로서 표결되거나 항상 정당한 메일로 표결되거나 적어도 90% 정당한 메일로 표결되는 IP 어드레스 또는 도메인 또는 URL의 리스트를 만들어서 상주(populate)하게(또는, 채우게) 하는데 이용될 수 있다. 이들 리스트는 다른 방식의 스팸 필터링에 사용될 수 있다. 예를 들어, 적어도 90% 스팸으로 표결된 IP 어드레스의 리스트는 어떤 메일도 받아들이지 않는 어드레스의 블랙홀 리스트를 만드는데 사용될 수 있다. 피드백 루프는 또한 스패머의 어카운트를 종결시키는데 사용될 수 있다. 예를 들어, ISP의 특정 사용자가 스팸을 보내는 것으로 나타나면, ISP는 자동으로 통지받을 수 있다. 이와 유사하게, 특정 도메인이 대량의 스팸에 대한 책임이 있는 것으로 보이면, 도메인의 이메일 제공자는 자동으로 통지받을 수 있다.

[0077] 피드백 루프 시스템을 실현하기 위해 사용될 수 있는 다수의 아키텍처가 있다. 한가지 예시적인 아키텍처는 도 7에 설명되는 바와 같이, 메일이 이메일 서버에게 도달할 때 발생하는 선택 프로세스에 따른 서버 기반이다. 대안적인 아키텍처는 도 6에서 설명되는 바와 같은 클라이언트 기반이다. 클라이언트 기반의 피드백 루프에서, 폴링 정보는 개인전용 필터의 성능을 향상시키기 위해 사용될 수 있고, 또는 여기에 도시된 예시적인 실현에 있어서, 정보는 공유된 필터(예를 들어, 회사 전체, 또는 글로벌)를 위한 트레이닝 데이터로서 공유된 저장소에 보내질 수 있다. 후술되는 다음 아키텍처는 단지 예시적일 뿐이고, 여기에 설명되지 않은 추가 컴포넌트 및 특징을 포함할 수 있다는 것을 알 수 있을 것이다.

[0078] 이제 도 6을 참조하면, 클라이언트 기반의 아키텍처에서의 피드백 루프 기술의 예시적인 블록도가 도시되어 있다. 네트워크(600)는 하나 이상의 클라이언트(602, 604 및 606)($CLIENT_1$, $CLIENT_2$ ... $CLIENT_N$으로 표시되고, n은 1보다 크거나 동일한 정수)로/로부터 이메일의 통신을 용이하게 하기 위해 제공된다. 네트워크는 인터넷, 또는 WAN(wide area network), LAN(local area network), 또는 소정의 다른 네트워크 구성과 같은 글로벌 통신 네트워크(GCN)일 수 있다. 이러한 특정 실현에 있어서, SMTP(simple mail transfer protocol) 게이트웨이 서

버(608)는 네트워크(600)로 인터페이스하여 SMTP 서비스를 LAN(610)에 제공한다. LAN(610) 상에 동작적으로 배치된 이메일 서버(612)는 게이트웨이(608)로 인터페이스하여 클라이언트(602, 604 및 606)의 착신 및 발신 이메일을 제어하고 프로세스한다. 이러한 클라이언트(602, 604 및 606)는 또한 적어도 제공된 메일 서비스를 억세스하기 위해 LAN(610) 상에 배치된다.

[0079] $CLIENT_1$(601)은 클라이언트 프로세스를 제어하는 중앙 처리 장치(CPU)(614)를 포함한다. CPU(614)는 다수의 프로세서를 포함할 수 있다. CPU(614)는 상술된 소정의 하나 이상의 데이터 개더링/피드백 기능을 제공하는 것과 관련된 명령어를 실행한다. 명령어는 적어도 상술된 기본적인 피드백 루프 방법론을 실행하는 인코드된 명령어를 포함하고, 클라이언트 및 메시지 선택을 어드레싱하기 위해, 메시지 변경, 데이터 보유, 클라이언트 신뢰도 및 분류 확인을 폴링하기 위해, 피드백 루프 시스템을 포함하는 다수의 소스로부터의 데이터의 재웨이팅을 위해, 스팸 필터 최적화를 위해, 그리고 차단 필터, 스팸 리스트의 작성, 각 ISP 및 이메일 제공자로의 스패머의 자동 통지를 조정하기 위해 결합하여 사용될 수 있는 적어도 소정의 방법 또는 모든 방법들을 실행하는 인코드된 명령어를 포함하는데, 이것에 제한되는 것은 아니다. 사용자 인터페이스(616)는 $CLIENT_1$이 이메일을 억세스하여 폴링 메시지를 표결하기 위해 상호작용할 수 있도록 CPU(614) 및 클라이언트 운영 체계와의 통신을 용이하게 하기 위해 제공된다.

[0080] 서버(612)로부터 검색된 클라이언트 메시지의 샘플링은 메시지 셀렉터(620)에 의해 폴링을 위해 선택될 수 있다. 메시지는 예정된 수신인(클라이언트)이 미리 참가하기로 동의했으면, 폴링을 위해 선택되어 변경된다. 메시지 변경자(622)는 폴링 메시지가 되도록 메시지를 변경한다. 예를 들어, 메시지(들)은 위에서 제공된 메시지 변경 설명에 따라 표결 명령어 및 표결 버튼 또는 링크를 포함하도록 변경될 수 있다. 표결 버튼 및/또는 링크는 클라이언트 이메일 소프트웨어의 사용자 인터페이스(616)를 변경함으로써 실현된다. 또한, 메시지 변경자(622)는 메시지들이 클라이언트(602)에 의해 보기 위해 오픈되거나 다운로드되기 전에 메시지(폴링 및 비폴링 메시지) 내의 소정의 바이러스를 제거할 수 있다.

[0081] 한 실현에 있어서, 스팸 파이팅 클라이언트(602)의 사용자는 일부 메시지들이 특히 폴링 메시지로서 표시되고 표결 버튼 등을 포함하는 각각의 메시지를 한번만 본다. 이러한 실현에 있어서, 스팸 파이팅 클라이언트(602)의 사용자는 일부 메시지들을 2번 보는데, 하나는 통상의 메시지이고, 다른 하나는 폴링 메시지이다. 이것은 몇가지 방식으로 실현될 수 있다. 예를 들어, 폴링 메시지는 서버(612)에게 복귀되어 폴링된 메시지 저장부 내에 저장될 수 있다. 대안적으로, 클라이언트(602)는 이메일 서버(612) 내에 추가 메시지를 저장할 수 있다. 대안적으로, 클라이언트(602)는 사용자에게 각 메시지를 두번, 한번은 통상 메시지로, 또 한번은 변경된 형태로 보여줄 수 있다.

[0082] 폴링 결과(626)는 CPU(614)로, 그 다음에 데이터베이스(630)로 보내질 수 있고, 데이터베이스(630)는 클라이언트 피드백 아키텍처의 특정 배열에 따라 하나의 클라이언트로부터 또는 2개 이상의 클라이언트로부터 데이터를 저장하도록 구성될 수 있다. 중앙 데이터베이스(630)는 각 클라이언트-사용자 정보뿐만 아니라, 폴링 메시지, 폴링 결과를 저장한다. 관련된 컴포넌트는 폴링 빈도, 클라이언트-사용자 신뢰성(예를 들어, 사용자 확인(632)), 및 다른 클라이언트 통계를 판정하는 것과 같은 정보를 분석하기 위해 이용될 수 있다. 확인 기술은 특히 클라이언트 표결의 신뢰도가 문제될 때 사용될 수 있다. 부정의 수, 변경된 마음의 수, 및 특정 사용자 또는 사용자들을 위해 폴링된 메시지의 수를 분석함으로써 의심이 생길 수 있으며: 대안적으로, 확인 기술은 모든 사용자에 대해 사용될 수 있다. 중앙 데이터베이스 내에 저장된 소정의 적절한 양의 데이터는 새로운 및/또는 향상된 스팸 파이터의 트레이닝을 용이하게 하기 위해 기계 학습 기술(634)에 사용될 수 있다.

[0083] 클라이언트(604 및 606)는 특정 클라이언트(들)에게 개인화되는 필터를 얻어서 트레이닝하기 위해 상술된 것과 유사한 컴포넌트를 포함한다. 폐기된 것 이외에, 폴링된 메시지 스크러버(scrubber)(628)는 폴링된 메시지의 양상이 데이터 집합, 데이터 압축 등과 같은 여러가지 이유로 제거될 수 있도록 CPU(614)와 중앙 데이터베이스(630) 사이를 인터페이스할 수 있다. 폴링된 메시지 스크러버(628)는 메시지의 관련없는 부분뿐만 아니라 그것과 연관된 소정의 바람직하지 않은 사용자 정보를 버릴 수 있다.

[0084] 이제 도 7을 참조하면, 본 발명의 피드백 루프 기술에 따라 다수의 사용자 로그인을 용이하게 하고 폴링 데이터를 얻는 예시적인 서버 기반의 피드백 루프 시스템(700)이 도시되어 있다. 네트워크(702)는 하나 이상의 사용자(또한, $USER_1$(704_1), $USER_2$(704_2) ... 및 $USER_N$(704_N)으로 표시되고, N은 1보다 크거나 동일한 정수)로/로부터 이메일의 통신을 용이하게 하기 위해 제공된다. 네트워크(702)는 인터넷, 또는 WAN, LAN, 또는 소정의 다른 네트워크 구성과 같은 GCN(글로벌 통신 네트워크)일 수 있다. 이러한 특정 실현에 있어서, SMTP 게이트웨이 서버

(710)는 네트워크(702)에 인터페이스하여 SMTP 서비스를 LAN(712)에 제공한다. LAN(712) 상에 동작적으로 배치된 이메일 서버(714)는 게이트웨이(710)에 인터페이스하여 사용자(704)의 착신 및 발신 이메일을 제어 및 프로세스한다.

[0085] 시스템(700)은 사용자 및 메시지 선택(716), 메시지 변경(718), 및 메시지 폴링(720, 722, 724)이 시스템(700)으로 로그인하는 각각의 서로다른 사용자마다 발생하도록 다수의 로그인 능력을 제공한다. 그러므로, 컴퓨터 운영 체계의 부트-업 프로세스의 일부로서 로그인 스크린을 제공하거나, 또는 요구에 따라, 사용자(704)가 자기의 착신 메시지를 억세스하기 전에 연관된 사용자 프로필을 채우기 위한 사용자 인터페이스(726)가 제공된다. 그러므로, 제1 사용자(704_1)($USER_1$)가 메시지를 억세스하기 위해 선택할 때, 제1 사용자(704_1)는 전형적으로 사용자명과 패스워드의 형태로 억세스 정보를 입력함으로써 로그인 스크린(728)을 통해 시스템으로 로그인한다. CPU(730)는 억세스 정보를 프로세스하여, 사용자가 메시지 통신 어플리케이션(예를 들어, 메일 클라이언트)을 통해 제1 사용자 인박스 위치(732)로만 억세스할 수 있게 한다.

[0086] 착신 메일이 메시지 서버(714) 상에 수신되면, 그들은 폴링을 위해 임의로 선택되는데, 이것은 적어도 하나의 메시지가 폴링을 위해 태그된다는 것을 의미한다. 태그된 메시지의 예정된 수신인(들)은 소정의 한 수신인이 또한 지정된 스팸 파이팅 사용자 인지를 판정하기 위해 관찰된다. 그러한 정보를 나타내는 수신인 특성은 메시지 서버(714) 상에 또는 적절한 시스템(700)의 소정의 다른 컴포넌트 상에 유지될 수 있다. 일단 예정된 수신인들 중 어느 수신인이 또한 스팸 파이터인 지가 결정되면, 메일 트랜잭션에 관한 소정의 다른 정보뿐만 아니라 이들의 각 메일의 한 카피는 저장을 위해 중앙 데이터베이스(734)에 보내질 수 있다. 폴링을 위해 태그된 메시지는 상술된 소정의 방식으로 메시지 변경자(718)에 의해 변경된다. 폴링을 위해 선택된 메시지는 또한 사용자(704)에게 지정될 수 있다. 예를 들어, 사용자(704)는 어떤 유형의 메시지만이 폴링에 이용가능하다는 것을 나타낸다. 이것이 데이터의 바이어스된 샘플링을 초래할 수 있기 때문에, 이러한 데이터는 불균형의 트레이닝 데이터 세트의 설정을 경감시키기 위해 다른 클라이언트 데이터에 대해 재웨이팅될 수 있다.

[0087] 폴링 메시지의 바이러스 스캐닝은 또한 이때 실행되거나, 또는 폴링 메시지가 사용자(704)에 의해 오픈되거나 다운로드되기 이전의 소정의 다른 시기에 실행될 수 있다. 일단 메시지가 적절한 방식으로 변경되었으면, 이들은 $INBOX_1$(732), $INBOX_2$(736) 및 $INBOX_N$(738)으로 표시되고 폴링을 위해 오픈될 수 있는 각 사용자의 인박스에 전달된다. 폴링 프로세스를 용이하게 하기 위해, 각각의 폴링 메시지는 사용자에 의해 선택될 때, 폴링 메시지 및 폴링 결과에 관련된 정보를 생성하는 2개 이상의 표결 버튼 또는 링크를 포함한다. 각 폴링 메시지의 텍스트는 표결 버튼 또는 링크를 포함시키도록 변경될 수 있다.

[0088] 분류에 기인하는 소정의 정보(예를 들어, 폴링 메시지 또는 이것과 연관된 ID, 사용자 특성)를 포함하는 메시지 폴 결과(MESSAGE $POLL_1$(720), MESSAGE $POLL_2$(722) 및 MESSAGE $POLL_n$(724)로 표시됨)는 LAN(712) 상의 네트워크 인터페이스(720)를 통해 중앙 데이터베이스(734)로 보내진다. 중앙 데이터베이스(734)는 새로운 및/또는 향상된 스팸 필터(742)를 만들거나 최적화시키도록 기계 학습 기술에 적용하기 위해 각 사용자로부터 폴링 및 사용자 정보(720, 722, 724)를 저장할 수 있다. 그러나, 사적기밀의 이유 및/또는 안전의 이유로, 비밀 정보는 중앙 데이버베이스(714)로 보내지기 전에 정보에서 제거되거나 해제될 수 있다. 폴링을 통해 사용자(들)(704)에 의해 생성된 정보는 통계적 데이터 내로 집합될 수 있다. 그러므로, 정보를 전송하는데 보다 작은 대역폭이 사용된다.

[0089] 그 다음, 새로 트레이닝된 스팸 필터(742)는 새로운 필터가 특정 요청에 의해 또는 자동으로 이용가능한 경우와 같은 진행에 따라 LAN(712)와 인터페이스하는 중앙 이메일 소프트웨어(도시되지 않음)뿐만 아니라 다른 서버(도시되지 않음)에 분산될 수 있다. 예를 들어, 가장 새로운 스팸 필터는 자동으로 배포될 수 있고/있거나, 웹사이트를 통한 다운로딩을 위해 이용가능하게 될 수 있다. 새로운 트레이닝 데이터 세트가 새로운 스팸 필터를 만들도록 생성됨에 따라, 기존의 데이터 세트(예를 들어, 이전에 얻어지고/지거나 필터를 트레이닝하는데 사용된 정보)는 데이터의 수명에 따라 폐기되거나 무시될 수 있다.

[0090] 이제, 스팸 파이팅에 전념한 조직이 다수의 상이한 필터-사용 조직에 의해 공유된 필터를 이용할 수 있게 하는 대안적인 시나리오를 고려해보자. 본 발명의 한 실시양상에서, 필터 제공자는 또한 매우 광범위한 이메일 서비스(예를 들어, 유료 및/또는 무료 이메일 어카운트)의 제공자이다. 오로지 자기 조직으로부터의 이메일에만 의존하기 보다는 오히려, 필터 제공자는 정당한 메일 및 스팸의 범위를 더욱 잘 포착하기 위해 일부의 필터링-사용 조직으로부터의 어떤 데이터를 또한 사용하도록 선택한다. 상술된 피드백 루프 시스템은 서버 또는 클라이언트 기반의 아키텍처에서 이러한 상호-조직 시나리오에서 사용될 수 있다. 우리는 자신의 사용자들로부터 그

리고 "내부" 조직인 서로다른 필터-사용 조직들로부터 데이터를 모으는 필터 제공자를 호출할 수 있고, "외부" 조직인 참여 필터 사용 조직들 중의 하나에 상주하는 컴포넌트를 호출할 수 있다. 일반적으로, 상호-조직 시스템은 핫메일(Hotmail)과 같은(이것에 제한되지 않음) 필터 제공자측의 메일 데이터베이스 서버(내부), 및 하나 이상의 개별 회사 내에 상주할 수 있는 것과 같은 하나 이상의 메시지 서버(외부)를 포함한다. 이 경우에, 내부 메일 데이터베이스 서버는 또한 자신 고객으로부터 실질적인 이메일 피드백을 저장한다. 본 발명의 이러한 양상에 따라, 트레이닝 데이터 세트는 각각의 외부 서버와 연관된 하나 이상의 외부 데이터베이스 상에 저장된 정보뿐만 아니라 내부 데이터베이스 상에 저장된 정보(예를 들어, 핫메일 또는 MSN 서버 상의 무료 이메일/메시징)에 기초하여 생성될 수 있다. 외부 데이터베이스 상에 유지된 정보는 예를 들어 기계 학습 기술에서의 사용을 위해 인터넷과 같은 네트워크를 통해 내부 서버에 통신될 수 있다. 궁극적으로, 외부 데이터베이스로부터의 데이터는 새로운 스팸 필터를 트레이닝하고/하거나, 외부에 위치되거나(예를 들어, 각 회사 내에) 내부 메일 서버와 연관된 현존하는 스팸 필터를 향상시키기 위해 이용될 수 있다.

[0091] 하나 이상의 외부 데이터베이스로부터의 데이터는 폴링 메시지, 폴링 결과(분류), 사용자 정보/특성, 및 사용자당, 사용자의 그룹 당 또는 각 회사마다 평균한 표결 통계 데이터 중에서 적어도 하나를 포함해야 한다. 표결 통계 데이터는 외부 데이터의 바이어스를 경감시키는 것뿐만 아니라 각 회사에 의해 생성된 정보의 신뢰도를 판정하는 것을 용이하게 한다. 그러므로, 하나 이상의 외부 데이터베이스(회사)로부터의 데이터는 재웨이팅되거나 하나 이상의 다른 외부 데이터베이스로부터 다르게 웨이팅될 수 있다. 게다가, 외부 엔티티는 상술된 것과 유사한 확인 기술을 사용하여 신뢰도 및 신뢰성이 테스트될 수 있다.

[0092] 회사 보안, 사적기밀 및 비밀성을 위해, 각 회사에서 이메일 서버로 인터넷을 통해 통신되는 정보 또는 데이터는 예를 들어 원래의 형태에서 제거, 단축, 및/또는 압축될 수 있다. 원래의 형태는 각각의 외부 데이터베이스 상에 유지될 수 있고/있거나, 그렇지 않으면 각 회사의 선호에 따라 처리된다. 그러므로, 이메일 서버 또는 소정의 다른 내부 메일 서버는 스팸 분류, 송신자 도메인, 송신자 이름, 스팸으로 분류된 메시지의 내용 등과 같은 트레이닝 데이터를 생성하는데 필요한 적절한 정보만을 수신한다.

[0093] 이제 도 8을 참조하면, 예시적인 상호-조직 피드백 시스템(800)이 도시되어 있는데, 이 시스템에서 내부 데이터베이스 서버 및 외부 메일 서버는 향상된 스팸 필터를 설정하기 위해 기계 학습 기술에 사용된 트레이닝 데이터 세트의 생성을 용이하게 하기 위해 네트워크를 통해 데이터베이스 정보를 통신하고 교환할 수 있다. 시스템(800)은 적어도 하나의 외부 메시지 서버(802)(예를 들어, 적어도 하나의 호사와 연관됨) 및 내부 데이터베이스 서버(804)를 포함한다. 상호-조직 시스템의 특성으로 인해, 외부 서버(802) 및 내부 이메일 서버(804)는 각각 자신의 데이터베이스를 유지한다. 즉, 이메일 서버(804)는 또한 새로운 스팸 필터(808)를 트레이닝하는데 사용될 수 있는 내부 데이터베이스(806)와 연관된다. 이와 마찬가지로, 내부 서버(802)는 이메일 서버(804)와 관련하여 내부에 위치한 스팸 필터(808)뿐만 아니라 적어도 하나의 새로운 스팸 필터(812)를 트레이닝하기 위해 사용될 수 있다. 그러므로, 외부 데이터베이스(810) 상에 저장된 정보는 이메일 서버 상에 위치한 스팸 필터(808)를 트레이닝하기 위해 사용될 수 있다.

[0094] GCN(814)은 내부 이메일 서버(804) 및 하나 이상의 외부 데이터베이스 서버(802)로/로부터 정보의 통신을 용이하게 하기 위해 제공된다. 상호-조직 시스템의 외부 서버(들) 컴포넌트는 서버 기반의 피드백 루프 시스템(예를 들어, 상기 도 7 참조)을 동작시키는 것과 유사한 방식으로 동작한다. 예를 들어, 메시지 서버(802), 외부 데이터베이스(810) 및 필터(812)는 LAN(815) 상에 위치될 수 있다. 또한, 컴퓨터 운영 체계의 부트-업 프로세스의 일부로서 로그인 스크린(818)을 제공하거나, 또는 요구에 따라, 사용자(들)이 자기의 착신 메시지를 억세스하기 전에 연관된 사용자 프로필을 채우기 위한 사용자 인터페이스(816)가 제공된다.

[0095] 이러한 서버 기반의 시스템에서, 하나 이상의 사용자($USER_1$(820), $USER_2$(822), $USER_N$(824))는 이용가능한 메일 서비스를 사용하기 위해 동시에 시스템으로 로그인할 수 있다. 실제로, 제1 사용자(820)($USER_1$)가 메시지를 억세스하기 위해 선택할 때, 제1 사용자(820)는 전형적으로 사용자명 및 패스워드의 형태로 억세스 정보를 입력함으로써 로그인 스크린(818)을 통해 시스템으로 로그인한다. CPU(826)는 억세스 정보를 프로세스하여, 사용자가 메시지 통신 어플리케이션(예를 들어, 메일 클라이언트)을 통해 제1 사용자 인박스 위치(828)만으로 억세스할 수 있게 한다.

[0096] 착신 메일이 메시지 서버(802) 상에 수신되면, 메시지는 폴링을 위채 임의로 또는 특별하게 타겟된다. 메시지가 폴링을 위해 선택되기 전에, 이러한 타겟된 메시지의 예정된 수신인은 스팸-파이터 사용자 리스트에 비교되어, 소정의 한 수신인이 또한 지정된 스팸 파이팅 사용자 인지를 판정한다. 이러한 정보를 나타내는 수신인 특성은 메시지 서버(802), 데이터베이스(810), 또는 적절한 시스템(800)의 소정의 다른 컴포넌트 상에 유지될 수

있다. 일단 예정된 수신인 중의 어떤 수신인이 또한 스팸 파이터인지 결정되면, 메시지(들)은 폴링을 위해 선택되고, 메일 트랜잭션에 관련된 소정의 다른 정보뿐만 아니라 폴링 메시지(들)의 한 카피는 데이터베이스(810)로 보내질 수 있다.

[0097] 폴링을 위해 선택된 메시지는 상술된 소정 수의 방식으로 메시지 변경자(830)에 의해 변경된다. 실제로, 고유 식별번호(ID)는 각각의 폴링 메시지에, 각각의 스팸 파이터에, 및/또는 각각의 폴링 결과에 할당되고, 데이터베이스(810) 내에 저장될 수 있다. 상술된 바와 같이, 폴링을 위해 선택된 메시지는 임의로 선택되거나, 또는 각 사용자(들)(820, 822 및 824)에게 지정될 수 있다. 예를 들어, $USER_1$(820)은 어떤 형태의 메시지만이 폴링에 이용가능하다는 것을 나타낸다(예를 들어, 회사의 외부에서 보내진 메시지). 이러한 특정 메시지로부터 생성된 데이터는 재웨이팅되고/되거나, 바이어스된 데이터의 샘플링의 획득을 경감시키기 위해 무시된다.

[0098] 폴링 메시지의 바이러스 스캐닝은 또한 이때 실행되거나, 폴링 메시지가 사용자에 의해 다운로드되고/되거나 오픈되기 전의 소정의 다른 시기에 실행될 수 있다. 일단 메시지가 적절한 방식으로 변경되었으면, 이들은 $INBOX_1$(828), $INBOX_2$(832) 및 $INBOX_N$(834)으로 표시되고 폴링을 위해 오픈될 수 있는 각 사용자(들)의 인박스로 전달된다. 폴링 프로세스를 용이하게 하기 위해, 각 폴링 메시지는 사용자에 의해 선택될 때, 폴링 메시지 및 폴링 결과에 관련된 정보를 생성하는 2개 이상의 표결 버튼 또는 링크를 포함한다. 각 폴링 메시지의 텍스트는 그 안에 표결 버튼 또는 링크를 포함하도록 변경될 수 있다.

[0099] 분류에 기인하는 소정의 정보(예를 들어, 폴링 메시지 또는 그것과 연관된 ID, 사용자 특성)를 포함하는 메시지 폴 결과(MESSAGE $POLL_1$(836), MESSAGE $POLL_2$(838) 및 MESSAGE $POLL_N$(840)로 표시됨)는 LAN(815) 상에 위치한 네트워크 인터페이스(842)를 통해 데이터베이스(810)에 보내진다. 데이터베이스(810)는 새로운 및/또는 향상된 스팸 필터(들)(812, 808)을 작성 및/또는 최적화하는데 사용되는 기계 학습 기술에서 각 사용자로부터의 폴링 및 사용자 정보를 나중에 사용하기 위해 저장한다.

[0100] 사적기밀의 이유로, 각각의 회사는, 예를 들어 폴링된 메시지 및/또는 사용자 정보를 자신의 데이터베이스(810)로 및/또는 GCN(814)을 통한 이메일 데이터베이스(806)로 보내기 전에 핵심적인 정보를 해체하고자 할 수 있다. 한가지 방법은 스팸 메시지에 관해 데이터베이스(806 및/또는 810)로의 피드백만을 제공함으로써, 합법적인 메일에 관한 피드백을 배제하는 것이다. 다른 방법은 송신자 및 송신자의 IP 어드레스와 같은 합법적인 메일에 관한 정보의 부분적인 서브셋만을 제공하는 것이다. 다른 방법은, 사용자에 의해서는 정당한 것으로 표시되지만 필터에 의해서는 부당한 것으로 표시되거나, 또는 그 반대로 되는 선택된 메시지에 대해, 필터에 보내기 전에 명백하게 사용자 허가를 받는 것이다. 소정의 이러한 방법 또는 그 조합은 스팸 필터(들)(808 및/또는 812)을 트레이닝하기 위해 데이터를 연속으로 제공하는 동안에 참여 클라이언트에 대한 비밀 정보의 사적기밀을 유지하는 것을 용이하게 한다.

[0101] 상술된 것과 같은 사용자 확인 스킴은 또한 회사 내의 각 사용자뿐만 아니라 각 회사에 적용될 수 있다. 예를 들어, 사용자는 의심스러운 사용자(들)의 분류가 필터 트레이닝으로부터 배제되는 상호-확인 기술이 개별적으로 필요하게 될 수 있다. 필터는 나머지 사용자(들)로부터의 데이터를 사용하여 트레이닝된다. 그 다음, 트레이닝된 필터는 배제된 사용자(들)로부터의 메시지를 통과하게 하여, 메시지를 어떻게 분류하였는 지를 판정한다. 불일치의 수가 임계 레벨을 초과하면, 의심스러운 사용자(들)은 신뢰할 수 없는 것으로 간주된다. 신뢰할 수 없는 사용자(들)로부터의 장래의 메시지 분류는 데이터베이스 및/또는 필터에 의해 받아들여지기 전에 수동으로 조사될 수 있다. 그렇지 않으면, 사용자(들)은 장래의 폴링에서 제거될 수 있다.

[0102] 이제 도 9를 참조하면, 본 발명의 여러가지 실시양상을 실현하는 예시적인 환경(910)은 컴퓨터(912)를 포함한다. 컴퓨터(912)는 프로세싱 유닛(914), 시스템 메모리(916) 및 시스템 버스(918)를 포함한다. 시스템 버스(918)는 시스템 메모리(916)를 포함하는(이것에 제한되지 않음) 시스템 구성요소를 프로세싱 유닛(914)에 연결한다. 프로세싱 유닛(914)은 다양한 이용가능 프로세서들 중의 소정의 프로세서일 수 있다. 이중(dual) 마이크로프로세서 및 다른 마이크로프로세서 아키텍처는 또한 프로세싱 유닛(914)으로서 사용될 수 있다.

[0103] 시스템 버스(918)는 메모리 버스 또는 메모리 제어기, 주변 버스 또는 외부 버스를 포함하고, 및/또는 11-비트 버스, ISA(Industrial Standard Architecture), MSA(Micro-Channel Architecture), EISA(Extended ISA), IDE(Intelligent Drive Electronics), VLB(VESA Local Bus), PCI(Peripheral Component Interconnect), USB(Universal Serial Bus), AGP(Advanced Graphics Port), PCMCIA(Personal Computer Memory Card International Association bus) 및 SCSI(Small Computer Systems Interface)를 포함하지만 이것에 제한되지 않는 여러가지 이용가능한 버스 아키텍처들 중의 소정의 것을 사용하는 로컬 버스를 포함하는 몇가지 종류의 버

스 구조(들) 중의 소정의 것일 수 있다.

[0104] 시스템 메모리(916)는 휘발성 메모리(920) 및 불휘발성 메모리(922)를 포함한다. 개시 중과 같이, 컴퓨터(912) 내의 소자들 간의 정보를 전달하기 위한 기본적인 루틴을 포함하는 기본적인 입/출력 시스템(BIOS)은 불휘발성 메모리(922) 내에 저장된다. 실례로서(제한적인 것은 아님), 불휘발성 메모리(922)는 ROM(read only memory), PROM(programmable ROM), EPROM(electrically programmable ROM), EEPROM(electrically erasable programmable ROM) 또는 플래시 메모리를 포함할 수 있다. 휘발성 메모리(920)는 외부 캐시 메모리로서 동작하는 RAM(random access memory)을 포함한다. 실례로서(제한적인 것은 아님), RAM은 SRAM(synchronous RAM), DRAM(dynamic RAM), SDRAM(synchronous DRAM), DDR SDRAM(double data rate SDRAM), ESDRAM(enhanced SDRAM), SLDRAM(synchlink DRAM) 및 DRRAM(direct rambus RAM)와 같은 다수의 형태로 이용가능하다.

[0105] 컴퓨터(912)는 또한 착탈가능/착탈불가능, 휘발성/불휘발성 컴퓨터 저장 매체를 포함한다. 도 9는 예를 들어 디스크 저장장치(924)를 도시하고 있다. 디스크 저장 장치(924)는 자기 디스크 드라이브, 플로피 디스크 드라이브, 테이프 드라이브, 재즈(Jaz) 드라이브, 지프(Zip) 드라이브, LS-100 드라이브, 플래시 메모리 카드 또는 메모리 스틱과 같은 장치를 포함하는데, 이것에 제한되지는 않는다. 또한, 디스크 저장장치(924)는 저장 매체를 따로, 또는 컴팩트 디스크 ROM 드라이브(CD-ROM), CD 기록가능 드라이브(CD-R Drive), CD 재기입가능 드라이브(CD-RW Drive) 또는 DVD-ROM(digital versatile disk ROM) 드라이브와 같은 광 디스크 드라이브를 포함하지만 이것에 제한되지 않는 다른 저장 장치와 조합하여 포함할 수 있다. 시스템 버스(918)로의 디스크 저장 장치(924)의 접속을 용이하게 하기 위해, 착탈가능 또는 착탈불가능 인터페이스는 전형적으로 인터페이스(926)와 같이 사용된다.

[0106] 도 9는 적절한 동작 환경(910)에서 설명된 기본적인 컴퓨터 자원과 사용자 사이의 중개자로서 작용하는 소프트웨어를 설명한다는 것을 알 수 있을 것이다. 이러한 소프트웨어는 운영 체계(928)를 포함한다. 디스크 저장 장치(924) 상에 저장될 수 있는 운영 체계(928)는 컴퓨터 시스템(912)의 자원을 제어하여 할당하도록 동작한다. 시스템 어플리케이션(930)은 시스템 메모리(916) 내에 또는 디스크 저장 장치(924) 상에 저장된 프로그램 모듈(932) 및 프로그램 데이터(934)를 통해 운영 체계에 의한 자원 관리의 이점을 얻는다. 본 발명은 다양한 운영 체계 또는 운영 체계의 조합으로 실현될 수 있다는 것을 알 수 있을 것이다.

[0107] 사용자는 입력 장치(들)(936)을 통해 컴퓨터(912) 내로 커맨드 또는 정보를 입력시킨다. 입력 장치(936)는 마우스와 같은 포인팅 장치, 트랙볼, 스타일러스, 터치 패드, 키보드, 마이크로폰, 조이스틱, 게임 패드, 위성 접시, 스캐너, TV 튜너 카드, 디지털 카메라, 디지털 비디오 카메라, 웹 카메라 등을 포함하지만, 이것에 제한되지는 않는다. 이들 및 다른 입력 장치는 인터페이스 포트(들)(938)을 경유하여 시스템 버스(918)를 통해 프로세싱 유닛(914)에 접속한다. 인터페이스 포트(들)(938)은 예를 들어, 직렬 포트, 병렬 포트, 게임 포트 및 USB(universal serial bus)를 포함한다. 출력 장치(들)(940)은 입력 장치(들)(936)과 동일한 형태의 포트들의 몇몇을 사용한다. 그러므로, 예를 들어, USB 포트는 입력을 컴퓨터(912)에 제공하고, 컴퓨터로부터 출력 장치(940)로 정보를 출력시키기 위해 사용될 수 있다. 출력 어댑터(942)는 특정 어댑터를 요구하는 다른 출력 장치들(940) 중에서 모니터, 스피커 및 프린터와 같은 몇몇의 출력 장치(940)가 있다는 것을 도시하기 위해 제공된다. 출력 어댑터(942)는 실례로서(제한은 아님), 출력 장치(940)와 시스템 버스(918) 사이에 접속 수단을 제공하는 비디오 및 사운드 카드를 포함한다. 다른 장치 및/또는 장치의 시스템은 원격 컴퓨터(들)(944)과 같이 입력 및 출력 능력을 제공한다는 것을 알기 바란다.

[0108] 컴퓨터(912)는 원격 컴퓨터(들)(944)과 같은 하나 이상의 원격 컴퓨터에 논리적 접속을 사용하는 네트워크된 환경에서 동작할 수 있다. 원격 컴퓨터(들)(944)은 퍼스널 컴퓨터, 서버, 라우터, 네트워크 PC, 워크스테이션, 마이크로프로세서 기반의 전기제품, 피어 장치 또는 다른 공통 네트워크 노드 등일 수 있고, 전형적으로 컴퓨터(912)와 관련하여 설명된 다수의 또는 모든 소자들을 포함한다. 간결하게 하기 위해, 메모리 저장 장치(946)만이 원격 컴퓨터(들)(944)와 도시된다. 원격 컴퓨터(들)(944)은 네트워크 인터페이스(948)를 통해 컴퓨터(912)에 논리적으로 접속된 다음에, 통신 접속(950)을 통해 물리적으로 접속된다. 네트워크 인터페이스(948)는 LAN 및 WAN과 같은 통신 네트워크를 포함한다. LAN 기술은 FDDI(Fiber Distributed Data Interface), CDDI(Copper Distributed Data Interface), Ethernet/IEEE 1102.3, Token Ring/IEEE 1102.5 등을 포함한다. WAN 기술은 포인트-투-포인트 링크, ISDN(Integrated Services Digital Networks) 및 그 변종과 같은 회로 스위칭 네트워크, 패킷 스위칭 네트워크 및 DSL(Digital Subscriber Lines)을 포함하는데, 이것에 제한되지는 않는다.

[0109] 통신 접속(들)(950)은 네트워크 인터페이스(948)를 버스(918)에 접속시키는데 사용된 하드웨어/소프트웨어를 칭한다. 통신 접속(950)이 컴퓨터(912) 내부에 있는 것으로 도시되었지만, 컴퓨터(912) 외부에도 있을 수 있다.

네트워크 인터페이스(948)에 접속하는데 필요한 하드웨어/소프트웨어는 예시적으로, 정규 텔레폰 그레이드 모뎀, 케이블 모뎀 및 DSL 모뎀을 포함하는 모뎀, ISDN 어댑터, 및 이더넷 카드와 같은 내부 및 외부 기술을 포함한다.

[0110] 도 10은 본 발명이 상호작용할 수 있는 샘플 컴퓨팅 환경(1000)의 개략적인 블록도이다. 시스템(1000)은 하나 이상의 클라이언트(들)(1010)를 포함한다. 클라이언트(들)(1010)은 하드웨어 및/또는 소프트웨어(예를 들어, 스레드, 프로세스, 컴퓨팅 장치)일 수 있다. 시스템(1000)은 또한 하나 이상의 서버(들)(1030)을 포함한다. 서버(들)(1030)은 또한 하드웨어 및/또는 소프트웨어(예를 들어, 스레드, 프로세스, 컴퓨팅 장치)일 수 있다. 서버(1030)는 예를 들어 본 발명을 사용함으로써 변형을 실행하기 위해 스레드를 수용할 수 있다. 클라이언트(1010)와 서버(1030) 간의 한가지 가능한 통신은 2개 이상의 컴퓨터 프로세스 사이에서 전송되도록 되어 있는 데이터 패킷의 형태로 될 수 있다. 시스템(1000)은 클라이언트(들)(1010)과 서버(들)(1030) 사이의 통신을 용이하게 하기 위해 사용될 수 있는 통신 프레임워크(1050)를 포함한다. 클라이언트(들)(1010)은 클라이언트(들)(1010)에 국한된 정보를 저장하기 위해 사용될 수 있는 하나 이상의 클라이언트 데이터 저장부(들)(1060)에 동작가능하게 접속된다. 이와 유사하게, 서버(들)(1030)은 서버(1030)에 국한된 정보를 저장하기 위해 사용될 수 있는 하나 이상의 서버 데이터 저장부(들)(1040)에 동작가능하게 접속된다.

[0111] 본 발명은 종래의 스팸 필터와 달리, 정당한 메일과 스팸 간의 구별을 학습하도록 본 발명의 피드백 기술에 따라 스팸 필터를 트레이닝함으로써 더욱 정확한 스팸 필터가 작성될 수 있고, 이에 의해 바이어스되고 부정확한 필터링이 경감된다.

[0112] 상술된 것은 본 발명의 예들을 포함한다. 물론, 본 발명을 설명하기 위해 컴포넌트 또는 방법론의 가능한 모든 조합을 설명할 수는 없지만, 본 분야에 숙련된 기술자들은 본 발명의 더 많은 조합과 변경이 가능하다는 것을 인식할 수 있을 것이다. 따라서, 본 발명은 첨부된 청구범위의 정신 및 범위 내에 속하는 모든 변경, 변형 및 변화를 포함하고자 한다.

도면의 간단한 설명

[0018] 도 1A는 본 발명의 실시양상에 따른 피드백 루프 트레이닝 시스템의 블록도이다.

[0019] 도 1B는 본 발명의 실시양상에 따른 예시적인 피드백 루프 트레이닝 프로세스의 흐름도이다.

[0020] 도 2는 본 발명의 실시양상에 따라 스팸 필터를 작성하기 위해 사용자에 의한 메일 분류를 용이하게 하는 예시적인 방법의 흐름도이다.

[0021] 도 3은 본 발명의 실시양상에 따라 도 2의 방법에 관련되는 사용자의 상호-확인을 용이하게 하는 예시적인 방법의 흐름도이다.

[0022] 도 4는 본 발명의 실시양상에 따라 사용자가 신뢰할수 없는 지의 판정을 용이하게 하는 예시적인 방법의 흐름도이다.

[0023] 도 5는 본 발명의 실시양상에 따라 스팸을 포착하여 스팸 발신자를 판정하는 것을 용이하게 하는 예시적인 방법의 흐름도이다.

[0024] 도 6은 본 발명의 실시양상에 따른 클라이언트 기반의 피드백 루프의 블록도이다.

[0025] 도 7은 본 발명의 실시양상에 따라 트레이닝 데이터를 생성하는 하나 이상의 사용자를 갖는 서버 기반의 피드백 루프 시스템의 블록도이다.

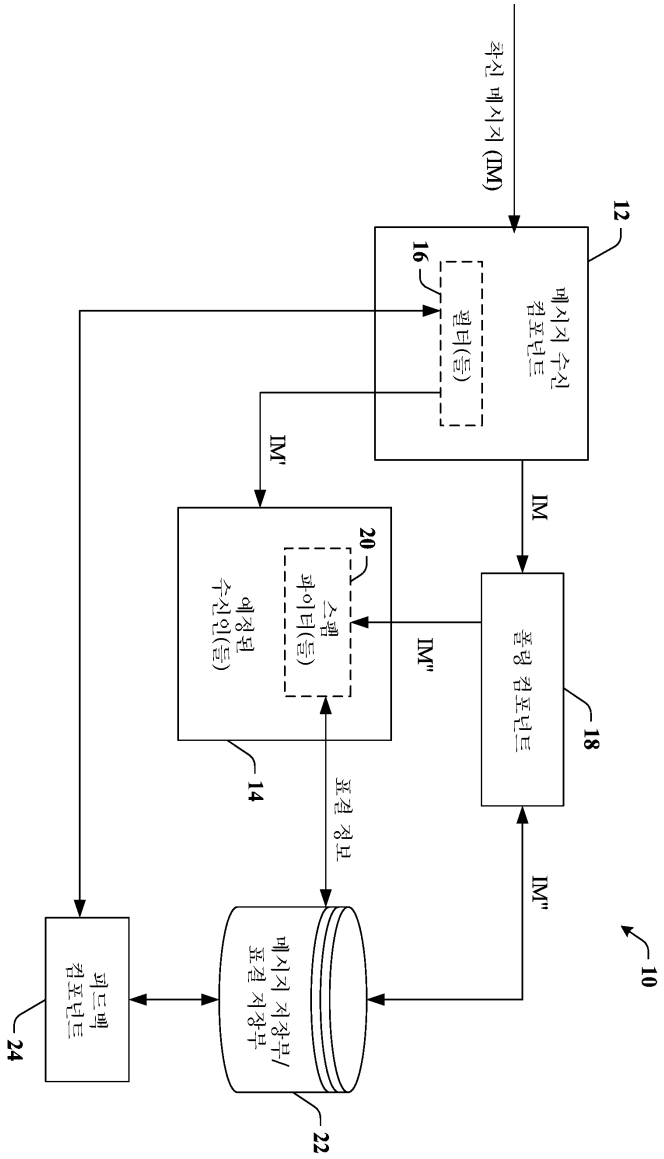
[0026] 도 8은 본 발명의 실시양상에 따라 상호-조직 서버 기반의 피드백 루프 시스템을 도시한 블록도로, 이 시스템은 본 발명의 실시양상에 따라 외부 사용자 데이터베이스 상에 저장된 트레이닝 데이터를 획득하기 위해 자체 데이터베이스를 가진 내부 서버를 포함한다.

[0027] 도 9는 본 발명의 여러가지 실시양상을 실현하기 위한 예시적인 환경을 도시한 도면이다.

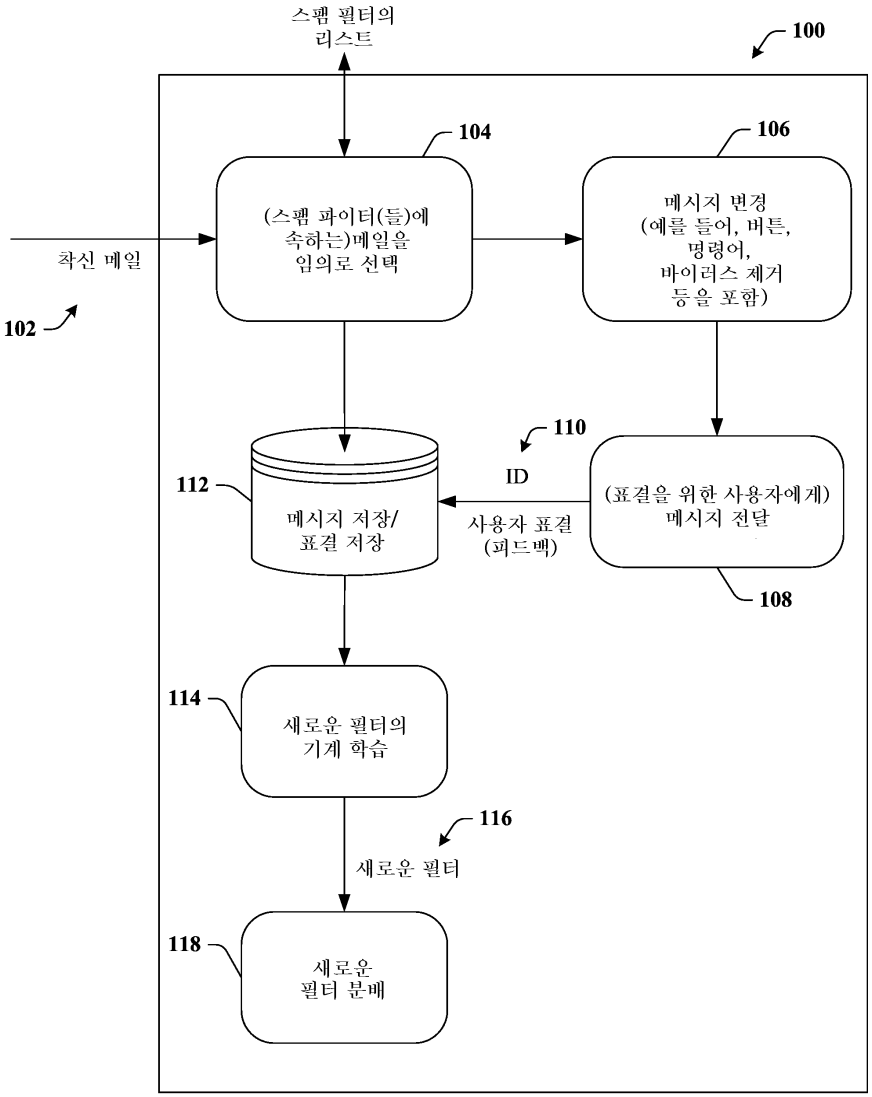
[0028] 도 10은 본 발명에 따른 예시적인 통신 환경의 개략적인 블록도이다.

도면

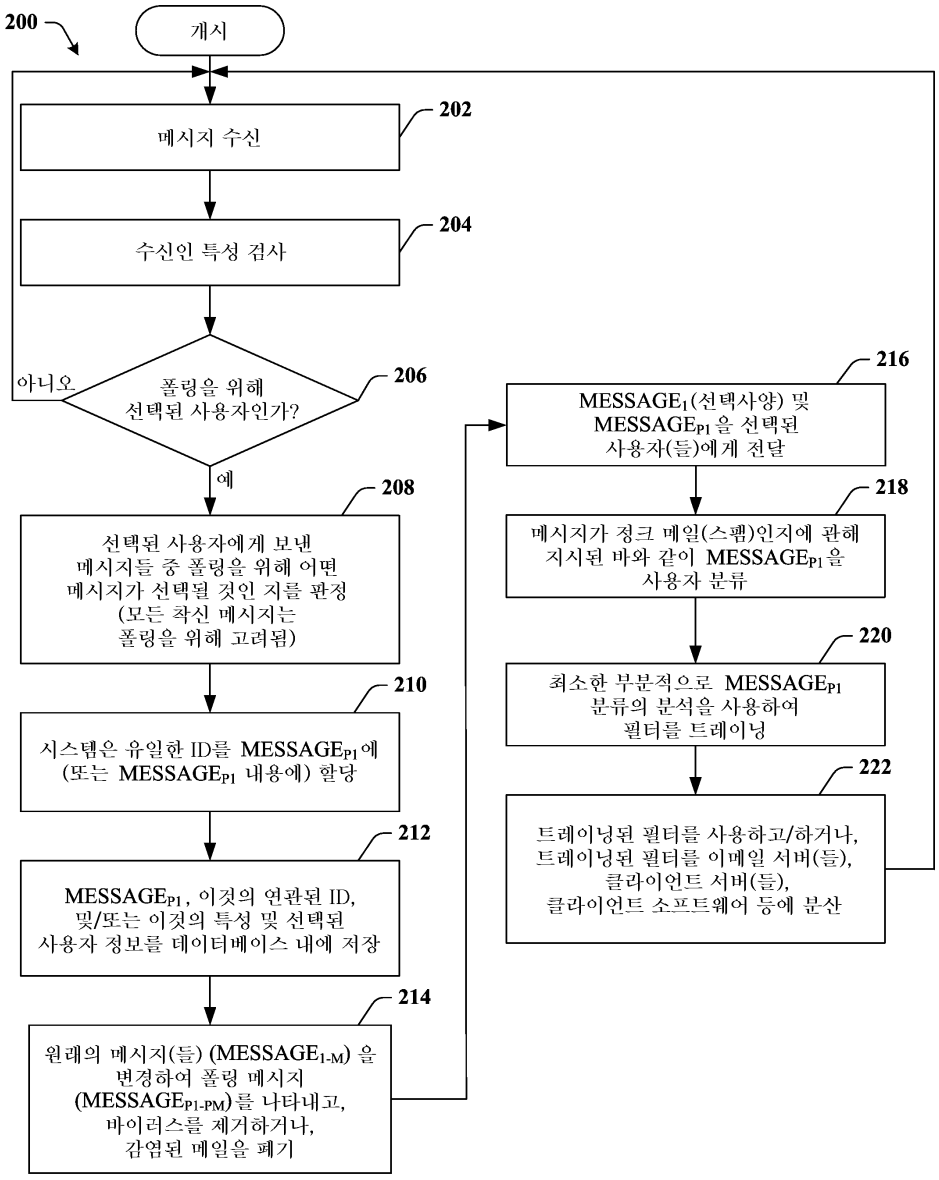
도면1A


착신 메시지 (IM)
12
메시지 수신 컴포넌트
16
필터(들)
IM'
IM
20
스팸 필터(들)
예정된 수신인(들)
14
IM"
폴링 컴포넌트
18
IM"
10
표결 정보
메시지 저장부/ 표결 저장부
22
피드백 컴포넌트
24

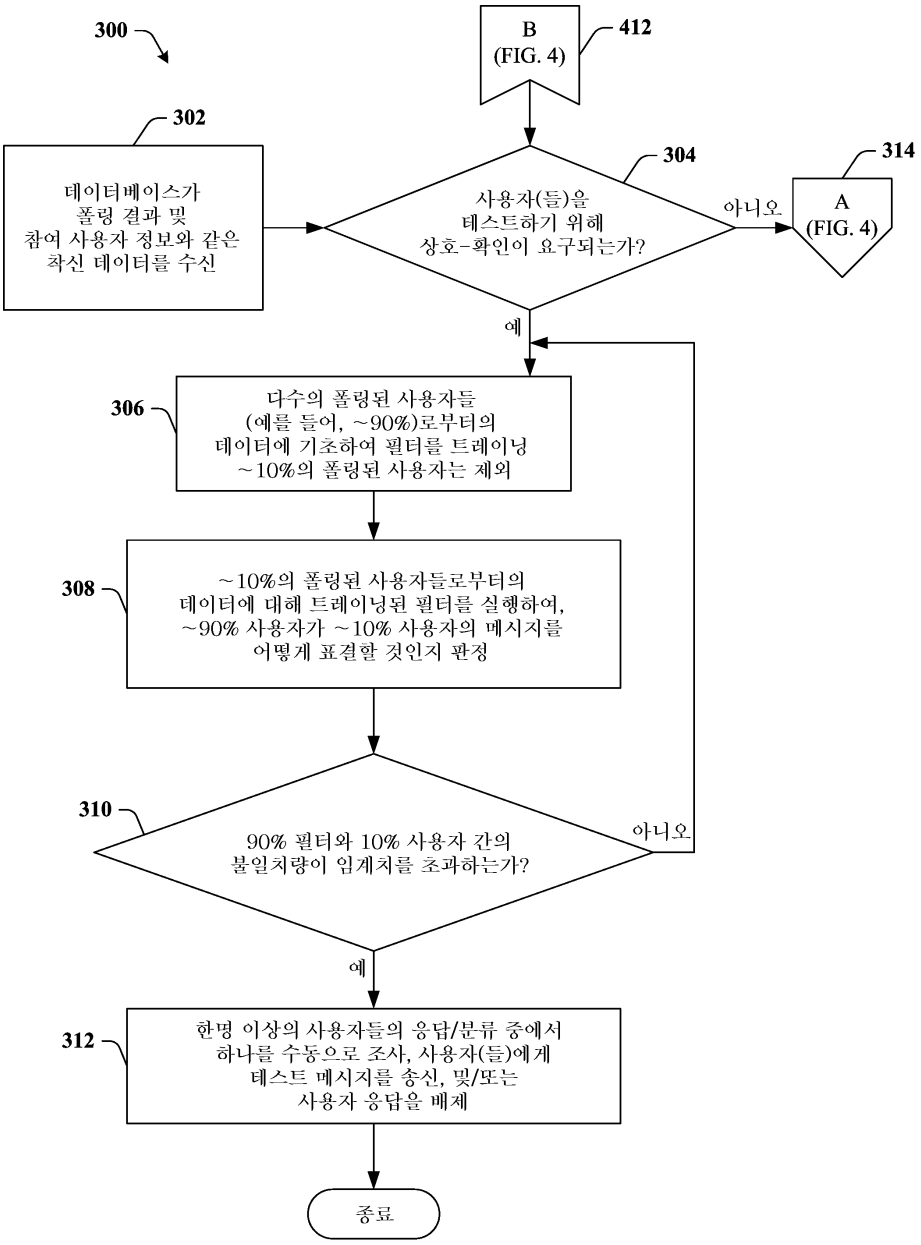
도면1B


스팸 필터의
리스트
100
104
106
(스팸 파이터(들)에
속하는)메일을
임의로 선택
메시지 변경
(예를 들어, 버튼,
명령어,
바이러스 제거
등을 포함)
착신 메일
102
110
ID
112
메시지 저장/
표결 저장
사용자 표결
(피드백)
(표결을 위한 사용자에게)
메시지 전달
108
114
새로운 필터의
기계 학습
116
새로운 필터
118
새로운
필터 분배

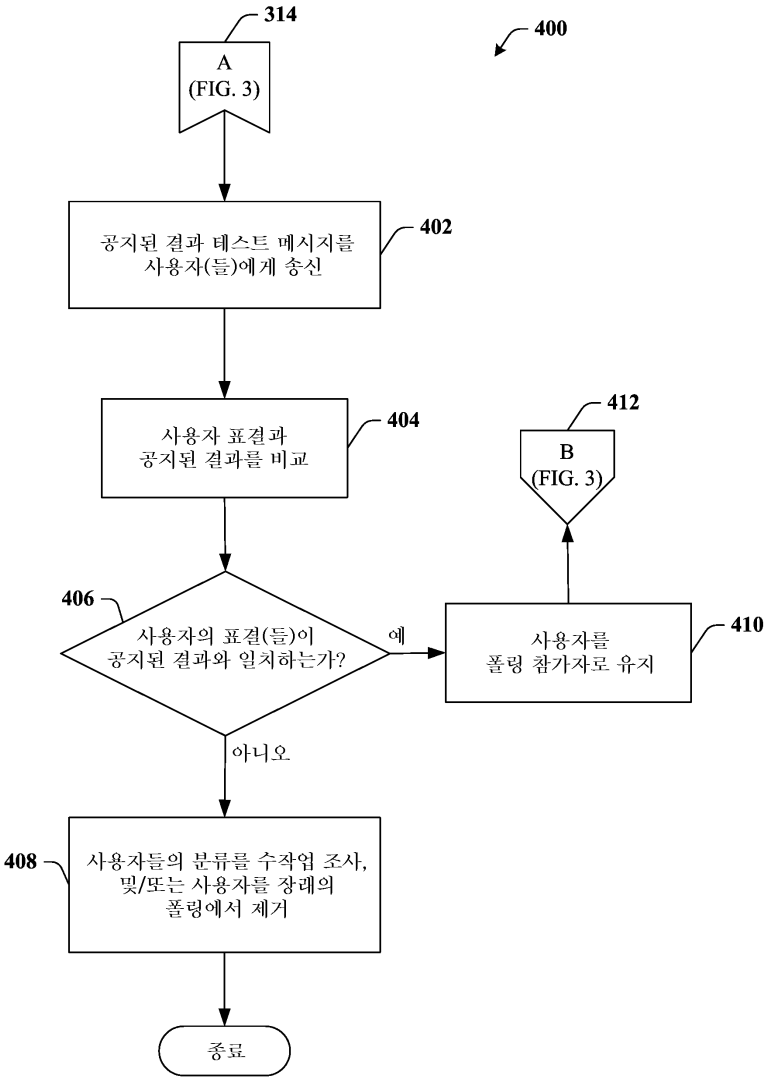
도면2


200
개시
202
메시지 수신
204
수신인 특성 검사
206
폴링을 위해 선택된 사용자인가?
아니오
예
208
선택된 사용자에게 보낸 메시지들 중 폴링을 위해 어떤 메시지가 선택될 것인 지를 판정 (모든 착신 메시지는 폴링을 위해 고려됨)
210
시스템은 유일한 ID를 MESSAGE$_{P1}$에 (또는 MESSAGE$_{P1}$ 내용에) 할당
212
MESSAGE$_{P1}$, 이것의 연관된 ID, 및/또는 이것의 특성 및 선택된 사용자 정보를 데이터베이스 내에 저장
214
원래의 메시지(들) (MESSAGE$_{1-M}$) 을 변경하여 폴링 메시지 (MESSAGE$_{P1-PM}$)를 나타내고, 바이러스를 제거하거나, 감염된 메일을 폐기
216
MESSAGE$_1$(선택사양) 및 MESSAGE$_{P1}$ 을 선택된 사용자(들)에게 전달
218
메시지가 정크 메일(스팸)인지에 관해 지시된 바와 같이 MESSAGE$_{P1}$ 을 사용자 분류
220
최소한 부분적으로 MESSAGE$_{P1}$ 분류의 분석을 사용하여 필터를 트레이닝
222
트레이닝된 필터를 사용하고/하거나, 트레이닝된 필터를 이메일 서버(들), 클라이언트 서버(들), 클라이언트 소프트웨어 등에 분산

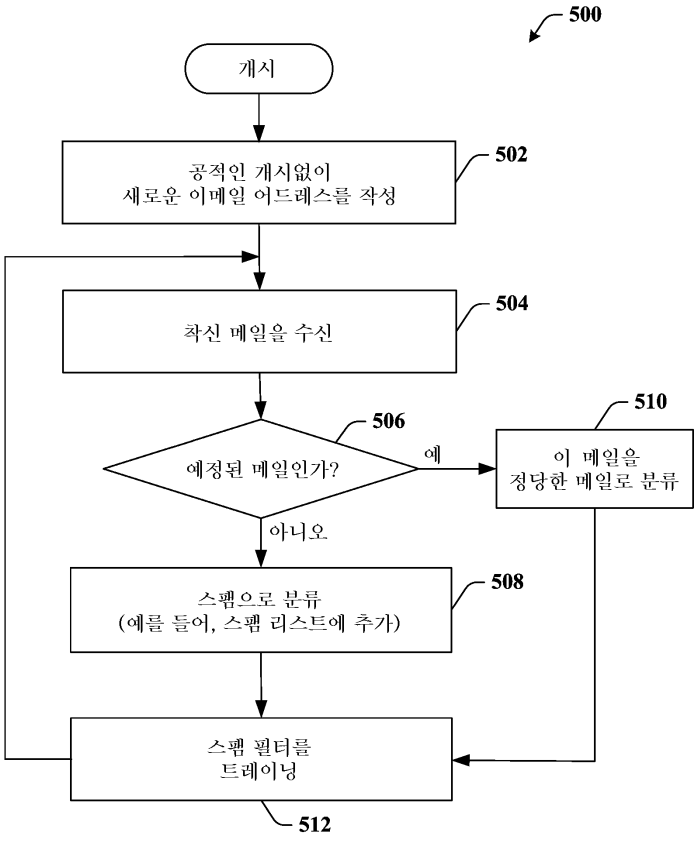
도면3


300
B
(FIG. 4)
412
302
데이터베이스가
폴링 결과 및
참여 사용자 정보와 같은
착신 데이터를 수신
304
사용자(들)을
테스트하기 위해
상호-확인이 요구되는가?
아니오
314
A
(FIG. 4)
예
306
다수의 폴링된 사용자들
(예를 들어, ~90%)로부터의
데이터에 기초하여 필터를 트레이닝
~10%의 폴링된 사용자는 제외
308
~10%의 폴링된 사용자들로부터의
데이터에 대해 트레이닝된 필터를 실행하여,
~90% 사용자가 ~10% 사용자의 메시지를
어떻게 표결할 것인지 판정
310
90% 필터와 10% 사용자 간의
불일치량이 임계치를 초과하는가?
아니오
예
312
한명 이상의 사용자들의 응답/분류 중에서
하나를 수동으로 조사, 사용자(들)에게
테스트 메시지를 송신, 및/또는
사용자 응답을 배제
종료

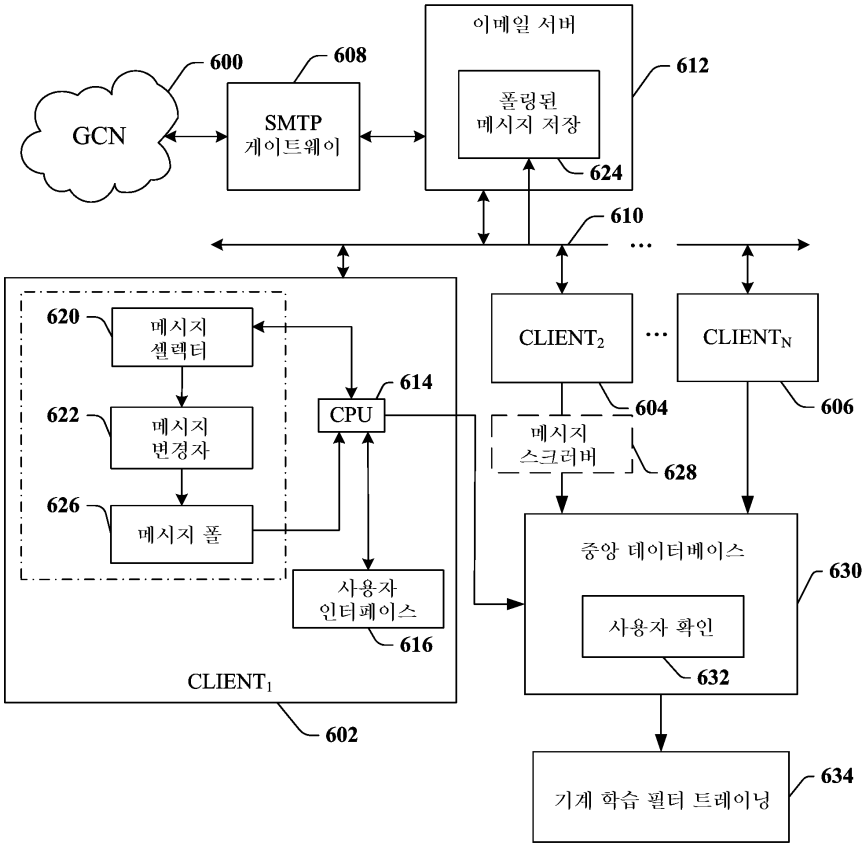
도면4


314
A
(FIG. 3)
400
공지된 결과 테스트 메시지를
사용자(들)에게 송신
402
사용자 표결과
공지된 결과를 비교
404
412
B
(FIG. 3)
406
사용자의 표결(들)이
공지된 결과와 일치하는가?
예
사용자를
폴링 참가자로 유지
410
아니오
408
사용자들의 분류를 수작업 조사,
및/또는 사용자를 장래의
폴링에서 제거
종료

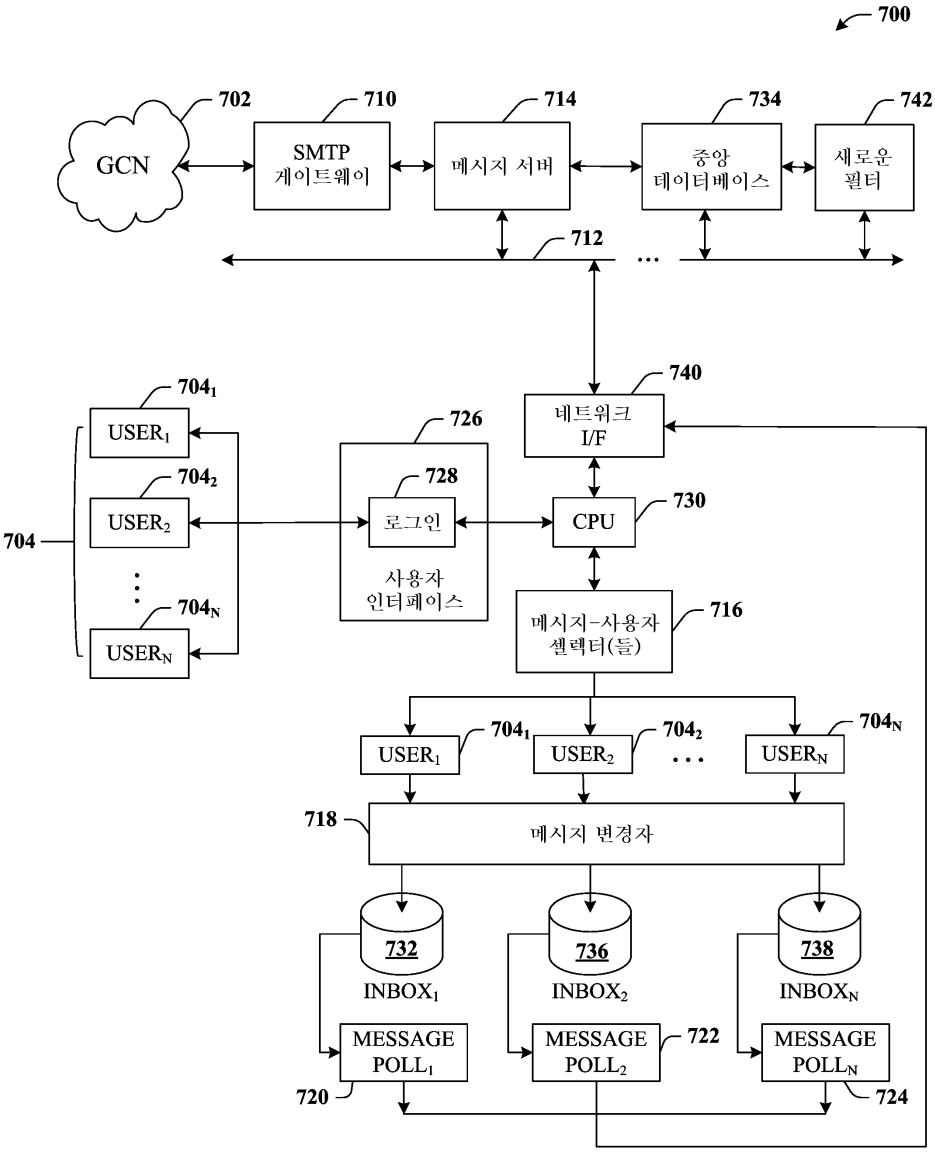
도면5


500
개시
공적인 개시없이
새로운 이메일 어드레스를 작성
502
착신 메일을 수신
504
506
예정된 메일인가?
예
510
이 메일을
정당한 메일로 분류
아니오
스팸으로 분류
(예를 들어, 스팸 리스트에 추가)
508
스팸 필터를
트레이닝
512

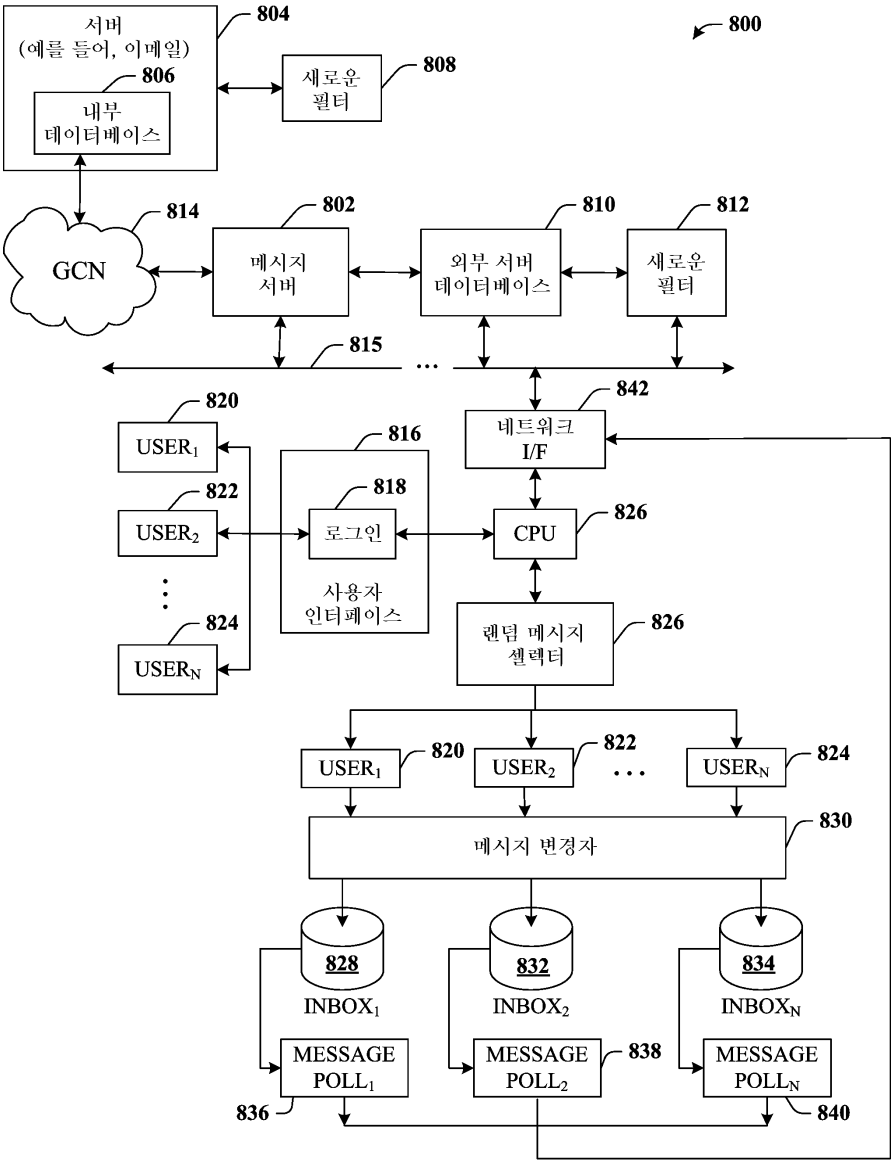
도면6


이메일 서버
612
폴링된
메시지 저장
624
600
608
GCN
SMTP
게이트웨이
610
620
메시지
셀렉터
622
메시지
변경자
626
메시지 폴
614
CPU
사용자
인터페이스
616
CLIENT1
602
CLIENT2
CLIENTN
604
606
메시지
스크리버
628
중앙 데이터베이스
630
사용자 확인
632
기계 학습 필터 트레이닝
634

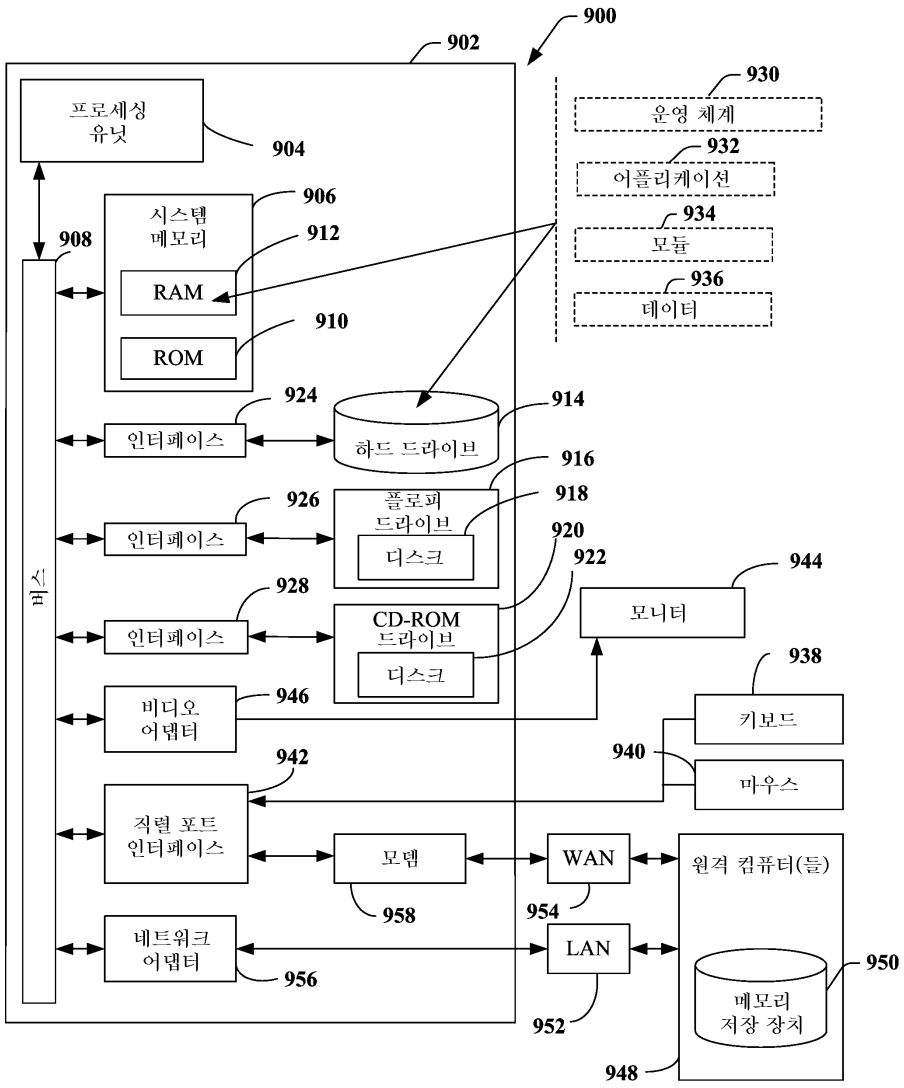
도면7


700
GCN
702
SMTP 게이트웨이
710
메시지 서버
714
중앙 데이터베이스
734
새로운 필터
742
712
740
네트워크 I/F
726
728
로그인
사용자 인터페이스
CPU
730
704
USER1
USER2
USERN
메시지-사용자 셀렉터(들)
716
718
메시지 변경자
732
736
738
INBOX1
INBOX2
INBOXN
MESSAGE POLL1
MESSAGE POLL2
MESSAGE POLLN
720
722
724

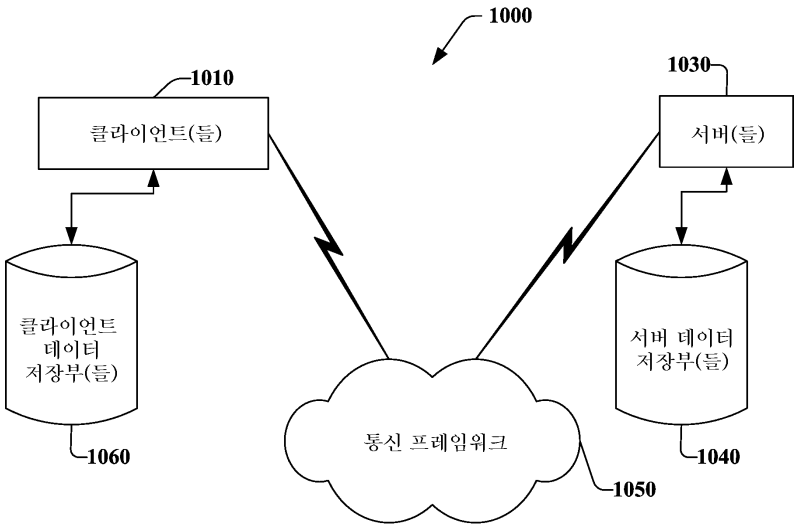
도면8


서버
(예를 들어, 이메일)
804
806
내부
데이터베이스
새로운
필터
808
800
814
GCN
802
메시지
서버
810
외부 서버
데이터베이스
812
새로운
필터
815
820
USER1
822
USER2
824
USERN
816
818
로그인
사용자
인터페이스
842
네트워크
I/F
826
CPU
랜덤 메시지
셀렉터
826
USER1
820
USER2
822
USERN
824
830
메시지 변경자
828
INBOX1
832
INBOX2
834
INBOXN
MESSAGE
POLL1
MESSAGE
POLL2
838
MESSAGE
POLLN
836
840

도면9


900
902
프로세싱 유닛
904
930
운영 체계
932
어플리케이션
934
모듈
936
데이터
906
시스템 메모리
912
908
RAM
910
ROM
924
인터페이스
914
하드 드라이브
916
926
플로피 드라이브
918
인터페이스
디스크
920
922
944
928
CD-ROM 드라이브
모니터
인터페이스
디스크
938
946
비디오 어댑터
키보드
940
942
마우스
직렬 포트 인터페이스
모뎀
WAN
원격 컴퓨터(들)
958
954
네트워크 어댑터
LAN
950
956
메모리 저장 장치
952
948
버스

도면10


1000
1010
클라이언트(들)
1030
서버(들)
클라이언트
데이터
저장부(들)
서버 데이터
저장부(들)
통신 프레임워크
1060
1040
1050