



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년08월01일
(11) 등록번호 10-1170562
(24) 등록일자 2012년07월26일

(51) 국제특허분류(Int. Cl.)
G06Q 50/00 (2006.01)
(21) 출원번호 10-2005-0127222
(22) 출원일자 2005년12월21일
심사청구일자 2010년12월21일
(65) 공개번호 10-2006-0071361
(43) 공개일자 2006년06월26일
(30) 우선권주장
11/018,270 2004년12월21일 미국(US)
(56) 선행기술조사문헌
KR100452910 B1
KR1020040060457 A
WO0026795 A1
WO2004061698 A1

(73) 특허권자
알카텔-루센트 유에스에이 인코포레이티드
미국 뉴저지 07974 머레이 힐 마운틴 애비뉴
600-700
(72) 발명자
카이 이강
미국 일리노이주 60565 네이퍼빌 너트메그 서클
1919
큐팁 에스 쉬리어
미국 일리노이주 60194 호프만 이스테이즈 웨스턴
스트리트 725
샤마 알로크
미국 일리노이주 60532 라일 스코트 코트 2947
(74) 대리인
제일특허법인, 장성구

전체 청구항 수 : 총 10 항

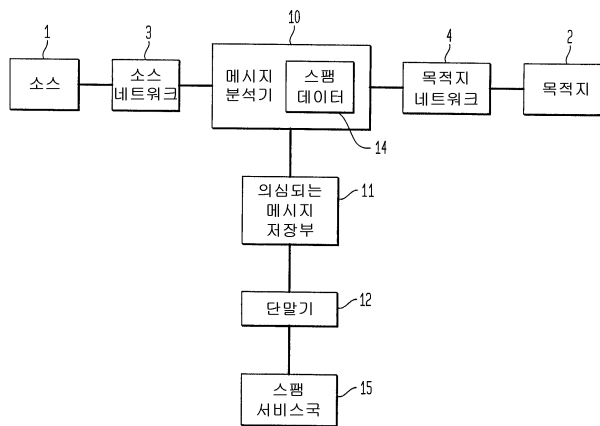
심사관 : 김석호

(54) 발명의 명칭 **스팸 메시지 검출 방법 및 스팸 메시지 검출 장치**

(57) 요약

본 발명은 원격 통신 네트워크에서 원하지 않는 (스팸) 메시지를 검출하는 방법에 관한 것이다. 의심되는 스팸 메시지의 내용을 분석하여, 메시지의 가중 특성들 및 특성들의 가중 합산이 임계값을 초과하는지를 판단한다. 만약 이러한 가중 합산이 임계값을 초과하면, 메시지는 스팸 메시지로서 취급되어 사람에게 의해 분석되며, 분석에 사용되는 가중 인수와 특성의 품질을 개선한다.

대표도 - 도1



특허청구의 범위

청구항 1

원격 통신 네트워크에서, 원하지 않는 (스팸) 메시지를 메시지 분석기에 의해 검출하는 방법으로서,
 잠재적 메시지의 각 특성에 대한 가중 인수(weighting factor), 지수(index) 및 임계값(threshold)을 저장하는 단계와,
 의심되는 스팸 메시지를 저장하는 단계와,
 상기 저장된 스팸 메시지의 특성을 도출하는 단계와,
 각 특성의 발생 횟수와, 해당 특성의 가중 인수 및 지수의 곱을 계산하는 단계와,
 상기 곱으로부터 배포된 스팸 프로파일 패턴을 형성하는 단계와,
 상기 배포된 스팸 프로파일 패턴이 메시지를 스팸 메시지로 분류하기 위한 기준을 충족하는지 여부를 판단하는 단계를 포함하는
 스팸 메시지 검출 방법.

청구항 2

제 1 항에 있어서,
 상기 배포된 스팸 프로파일 패턴이 상기 기준을 충족하는지 여부를 판단하는 동안, 만약 어떤 곱이라도 그 곱의 특성에 대한 임계값을 초과하면, 관련 메시지는 스팸 메시지로 선언되는
 스팸 메시지 검출 방법.

청구항 3

제 1 항에 있어서,
 상기 배포된 스팸 프로파일 패턴이 상기 기준을 충족하는지 여부를 판단하는 동안, 복수의 특성의 패턴에 대해서 각 패턴에 대한 임계값이 저장되고,
 만약 어떤 패턴이라도 그 패턴에 대한 임계값이 초과되면, 메시지는 스팸 메시지로 선언되는
 스팸 메시지 검출 방법.

청구항 4

제 1 항에 있어서,
 상기 배포된 스팸 프로파일 패턴이 상기 기준을 충족하는지 여부를 판단하는 동안, 만약 상기 메시지에 대한 모든 곱의 합이 사전결정된 임계값을 초과하면, 상기 메시지는 스팸 메시지로 취급되는
 스팸 메시지 검출 방법.

청구항 5

제 1 항에 있어서,
 특성의 상기 가중 인수 또는 임계값은 서비스국으로부터의 메시지에 응답하여 변경될 수 있는
 스팸 메시지 검출 방법.

청구항 6

원격 통신 네트워크에서의 원하지 않는 (스팸) 메시지를 검출하는 장치로서,
 잠재적 메시지의 각 특성에 대한 가중 인수, 지수 및 임계값을 저장하는 수단과,
 의심되는 스팸 메시지를 저장하는 수단과,

상기 저장된 스팸 메시지의 특성을 도출하는 수단과,

각 특성의 발생 횟수와, 해당 특성의 가중 인수 및 지수의 곱을 계산하는 수단과,

상기 곱으로부터 배포된 스팸 프로파일 패턴을 형성하는 수단과,

상기 배포된 스팸 프로파일 패턴이 메시지를 스팸 메시지로서 분류하기 위한 기준을 충족하는지 여부를 판단하는 수단을 포함하는

스팸 메시지 검출 장치.

청구항 7

제 6 항에 있어서,

만약 어떤 곱이라도 그 곱의 특성에 대한 임계값을 초과하면, 관련 메시지를 스팸 메시지로서 취급하는 수단을 더 포함하는

스팸 메시지 검출 장치.

청구항 8

제 6 항에 있어서,

복수의 특성 패턴에 대해서 각 패턴에 대한 임계값을 저장하는 수단과,

만약 어떤 패턴이라도 그 패턴에 대한 임계값이 초과되면, 메시지를 스팸 메시지로서 취급하는 수단을 더 포함하는

스팸 메시지 검출 장치.

청구항 9

제 6 항에 있어서,

만약 상기 메시지에 대한 모든 곱의 합이 사전결정된 임계값을 초과하면, 상기 메시지를 스팸 메시지로서 취급하는 수단을 더 포함하는

스팸 메시지 검출 장치.

청구항 10

제 6 항에 있어서,

특성의 상기 가중 인수 또는 임계값을 서비스국으로부터의 메시지에 응답하여 변경하는 수단을 더 포함하는

스팸 메시지 검출 장치.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 메시지의 내용에 기초하여 스팸 메시지를 검출하는 방법에 관한 것이다.

인터넷의 출현으로, 발신자에게는 적은 비용 또는 무비용으로 다수의 수신자에게 메시지를 전송하는 것이 용이해졌다. 메시지는 SMS(short message service)의 짧은 메시지를 포함한다. 이들 메시지는 메시지를 삭제해야 하고 이 메시지가 중요한 것인지를 결정해야 하는 메시지 수신자에게는 귀찮은, 불필요하고 원치 않는 메시지를 포함한다. 또한, 이들 메시지는, 너무나 많은 스팸으로 인해 분노하는 고객과 관련된 고객 관련 문제를 나타낼 뿐 아니라, 일반적으로 거의 수익을 내지 못하거나 전혀 수익을 내지 못하는 이러한 메시지들이 네트워크 리소스를 사용하기 때문에, 이러한 메시지는 메시지를 전송하는데 사용되는 원격 통신 네트워크의 캐리어에게 방해

[0007]

[0008]

가 된다. 이러한 문제의 심각성에 대한 설명은 다음 2개의 통계에 의해 주어진다. 2003년 중국에서, 3조 개의 SMS 메시지가 중국의 원격 통신 네트워크를 통해 전송되었고, 이들 메시지 중에서 3/4의 메시지가 스팸 메시지로 추정된다. 두 번째 통계는 미국에서의 이메일 메시지 중 85-90%가 스팸으로 추정된다는 것이다.

- [0009] 전달되는 스팸 메시지의 개수를 줄이는 다양한 장치가 제안 및 구현되었다. 스팸 메시지를 전달하기 전에 메시지를 분석하는 다수의 장치가 제안되었다. 하나의 장치에 따르면, 발신자가 수신자에 의해 지정된 사전 선택된 그룹 중 하나가 아니면, 이 스팸 메시지는 차단된다. 또한 수신자가 N개 이상의 목적지로 향하는 메시지가 배송되지 않도록 표시할 수 있게 함으로써 스팸 메시지를 차단할 수 있다.
- [0010] 수신자는 자신의 전화 번호 또는 이메일 주소를 공표하는 것을 거부할 수 있다. 발신자가 수신자의 전화 번호 또는 이메일을 조사할 수 없게 하는 명백한 단점에 추가로, 이러한 장치는 비효율적일 수 있다. 목록에 없는 이메일 주소가 IP 네트워크로부터의 전문 해커에 의해 검출될 수 있으며, 예를 들어, 라우터에서 메시지의 헤더를 모니터링함으로써 검출될 수 있다. 목록에 없는 착신 번호는 단순히 발신자에게 국부호(office code)의 모두 10,000개의 전화 번호에 메시지를 전송할 수 있게 하며, 상술한 바와 같이, 이는 복수의 목적지로 메시지를 전송하기 위한 현재의 장치를 이용하면 매우 쉽다.
- [0011] 음란성 메시지 또는 원치 않는 광고를 수신자에게 전달하는 불쾌한 메시지는 포착하기 어려운 스팸 메시지이다. 발신자가 동일한 소스로부터 다수의 무해한 메시지를 전송할 수 있기 때문에, 종종 이러한 메시지는 메시지의 내용을 조사해야만 차단될 수 있다. 스팸 검출의 주요 문제점은 메시지의 내용에 근거하여 스팸을 검출하는 것이다.

발명이 이루고자 하는 기술적 과제

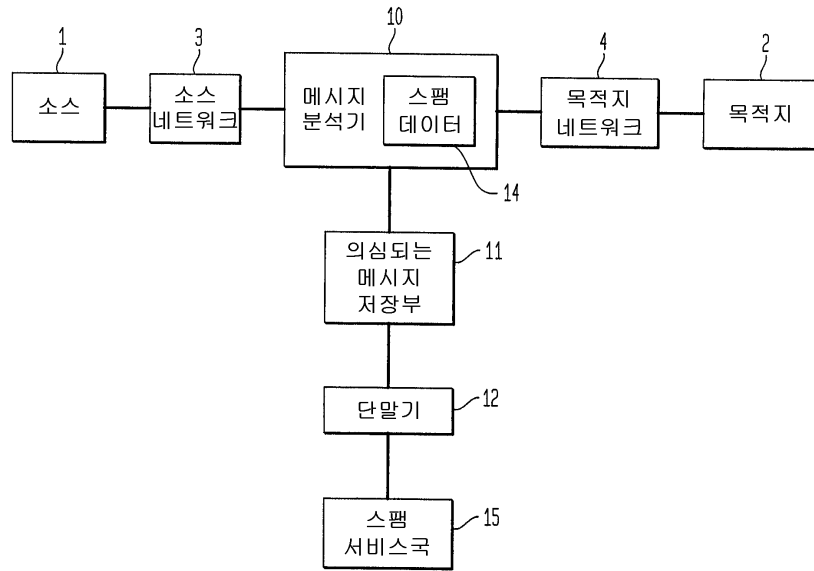
- [0012] 키워드와 같은 소정의 특성의 존재 및 이러한 특성의 빈도수에 대해 의심되는 메시지를 분석하는 본 발명에 따라서, 상술한 문제점이 해소되고 종래 기술에 대해 개선된다. 여기서, 각 특성에는 적절한 스팸 지수, 거의 고정적이고 사전 결정되며 사전제공되는 품질 및 트래픽 크기와 메시지/내용 타입에 따라서 동적으로 변화하는 가중 인자(weighting factor)가 주어진다. 메시지는 사용 빈도수가 임계값을 초과하는 특성과, 사전결정된 특성들의 조합의 사용이 임계값을 초과하는 사전결정된 특성들의 조합과, 조합된 사용이 임계값을 초과하는 모든 특성에 대해서 조사된다. 본 발명의 하나의 특징에 따르면, 각 특성의 가중 인수는 분석하는 사람에 의한 의심 메시지에 대한 조사 결과와 일치하도록 동적으로 조정된다. 바람직하게, 분석가를 통해서 검출 프로세스를 얻을 수 있다.

발명의 구성 및 작용

- [0013] 도 1은 본 발명의 동작을 예시한다. 소스(1)는 메시지를 목적지(2)에 전송하고자 한다. 메시지는 그 메시지가 스팸 메시지일 수 있음을 인식하는 네트워크(3)로 전송되지만, 이는 메시지를 내용을 분석하여 판단할 필요가 있다. 네트워크(3)는 메시지를 메시지 분석기(10)로 전달한다. 메시지 분석기에 의해 그 메시지가 스팸 메시지가 아니라고 결론나면, 그 메시지는 네트워크(4)를 통해 목적지(2)로 전송된다.
- [0014] 메시지 분석기(10)는 특성과, 각각의 특성에 대한 엄격 지수(severity index)와, 각각의 엄격 지수에 대한 가중 인수 및 그 특성에 대한 엄격 레벨 임계값의 데이터 표(14)를 포함한다.
- [0015] 스팸 특성은 스팸 메시지의 가능한 표시자인 단어, 구문, 문장, 이미지 또는 비디오 세그먼트이다. "마담"의 단어를 예로 든다. 그 메시지에서 발생하는 각각의 특성에 있어서, 그 특성의 발생 회수, 엄격 지수 및 가중 인수의 곱이 계산되어 엄격 레벨을 만든다. 엄격 레벨은 메시지가 스팸 메시지로서 취급되어야 하는지를 판단하는데 사용된다.
- [0016] 엄격 지수와 엄격 임계값은 상대적으로 일정하게 유지되지만, 가중 인수는, 스팸 활동성이 적은 특수 문제 지역(가중 인수를 증가시킴) 또는 지역(가중 인수를 감소시킴)에서의 검출에 응답하여, 스팸 서비스국(15)으로부터의 메시지에 응답하여 변경될 수 있다.
- [0017] 메시지 분석기는 메시지의 내용을 잡아서 예를 들어, "마담" 및 "에인" 등의 사전 저장된 특성을 조사한다. 각각의 사전 저장된 특성에 있어서, 엄격 레벨에 도달할 때 이러한 특성이 얼마나 과다하게 가중되는지를 표시하는 가중 인수가 있다. 엄격 레벨이 소정의 임계값을 초과하는 메시지는 차단되고 추후에 사람의 분석을 위해서

도면

도면1



도면2

