



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0059263
(43) 공개일자 2013년06월05일

- (51) 국제특허분류(Int. Cl.)
G06Q 50/30 (2012.01) G06F 17/27 (2006.01)
G06F 17/30 (2006.01)
- (21) 출원번호 10-2012-0103930
(22) 출원일자 2012년09월19일
심사청구일자 2012년09월19일
- (30) 우선권주장
1020110125386 2011년11월28일 대한민국(KR)
- (71) 출원인
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
- (72) 발명자
유성준
서울특별시 강동구 명일1동 삼익그린2차아파트
506동 1503호
강한훈
경기도 안산시 상록구 성포동 587번지 주공아파트
307동 101호
- (74) 대리인
특허법인엠에이피에스

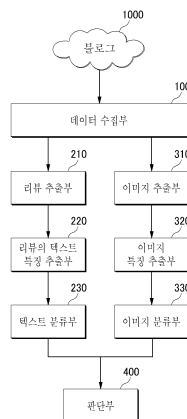
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 **상품 판매 업체 관련 글 판단 방법 및 그 장치**

(57) 요약

상품 판매 업체 관련 글 판단 장치가 개시되며, 상기 상품 판매 업체 관련 글 판단 장치는 질의어를 입력하여 검색된 상품 판매 업체 데이터베이스를 기초로 웹 상의 블로그에서 데이터를 수집하는 데이터 수집부; 상기 데이터 수집부에서 수집된 데이터에서 리뷰를 추출하는 리뷰 추출부; 상기 데이터 수집부에서 수집된 데이터에서 이미지를 추출하는 이미지 추출부; 및 상기 리뷰 추출부와 상기 이미지 추출부에서 추출된 정보를 기초로 상품 판매 업체와 관련된 정보 여부를 판단하는 판단부를 포함한다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	SS100006C0214831
부처명	서울시
연구사업명	서울전략산업 지원사업
연구과제명	시맨틱 멀티미디어 검색 엔진 개발
주관기관	건국대학교 산학협력단
연구기간	2010.12.01 ~ 2011.11.30

특허청구의 범위

청구항 1

소비자를 위한 상품 판매 업체 관련 글 판단 방법에 있어서,

- (a) 질의어를 입력하여 검색된 상품 판매 업체의 데이터베이스를 수집하는 단계;
- (b) 상기 수집된 상품 판매 업체의 데이터베이스를 기초로 웹 상의 블로그에서 데이터를 추출하는 단계;
- (c) 상기 추출된 데이터에서 리뷰를 분류하는 단계;
- (d) 상기 추출된 데이터에서 이미지를 분류하는 단계; 및
- (e) 상기 분류된 리뷰와 이미지를 바탕으로 상품 판매 업체 관련 데이터를 관련도에 따라 복수의 등급으로 분류하는 단계를 포함하는 상품 판매 업체 관련 글 판단 방법.

청구항 2

제1항에 있어서,

상기 (c) 단계는 상기 분류된 리뷰에 포함된 텍스트를 텍스트 전처리 단계, 텍스트 특징 추출 단계, 및 텍스트 분류 단계를 통해 상품 판매 업체와 관련된 리뷰 여부를 분류하는 상품 판매 업체 관련 글 판단 방법.

청구항 3

제2항에 있어서,

상기 텍스트 전처리 단계는 상기 추출된 리뷰의 요약으로부터 불필요한 HTML태그, 숫자, 기호 등을 제거하는 상품 판매 업체 관련 글 판단 방법.

청구항 4

제2항에 있어서,

상기 텍스트 특징 추출 단계는 상기 리뷰에서 특징적인 단어를 추출하는 단계로 단어의 품사에 따라 특징을 추출하는 상품 판매 업체 관련 글 판단 방법.

청구항 5

제2항에 있어서,

상기 텍스트 분류 단계는 상기 텍스트에서 추출된 특징값에 대한 상품 판매 업체와 관련된 리뷰 여부에 대한 확률을 구하여 분류 결과를 추출하는 상품 판매 업체 관련 글 판단 방법.

청구항 6

제1항에 있어서,

상기 (d) 단계는 상기 분류된 이미지를 이미지 전처리 단계, 이미지 특징 추출 단계, 및 이미지 분류 단계를 통해 상품 판매 업체와 관련된 정보 여부를 분류하는 상품 판매 업체 관련 글 판단 방법.

청구항 7

제 6항에 있어서,

상기 이미지 전처리 단계는 이미지의 크기를 정규화하는 상품 판매 업체 관련 글 판단 방법.

청구항 8

제 6항에 있어서,

상기 특징 추출 단계는 엠펙7(MPEG7)을 사용하여 이미지로부터 특징값을 추출하는 상품 판매 업체 관련 글 판단 방법.

청구항 9

제 6항에 있어서,

상기 이미지 분류 단계는 SVM(support vector machine)분류 방법으로 상품 판매 업체와 관련된 정보 여부를 분류하는 상품 판매 업체 관련 글 판단 방법.

청구항 10

제 1항에 있어서,

상기 (e) 단계는 상기 (c) 단계 및 상기 (d) 단계 모두에서 상품 판매 업체 관련 데이터로 분류된 경우의 관련도를 '상', 상기 (c) 단계 및 상기 (d) 단계 중 하나가 상품 판매 업체 관련 데이터로 분류된 경우의 관련도를 '중', 및 상기 (c) 단계 및 상기 (d) 단계 모두 상품 판매 업체 관련 데이터가 아닌 경우의 관련도를 '하'로 분류하는 상품 판매 업체 관련 글 판단 방법.

청구항 11

상품 판매 업체 관련 글 판단 장치에 있어서,

질의어를 입력하여 검색된 상품 판매 업체 데이터베이스를 기초로 웹 상의 블로그에서 데이터를 수집하는 데이터 수집부;

상기 데이터 수집부에서 수집된 데이터에서 리뷰를 추출하는 리뷰 추출부;

상기 데이터 수집부에서 수집된 데이터에서 이미지를 추출하는 이미지 추출부; 및

상기 리뷰 추출부와 상기 이미지 추출부에서 추출된 정보를 기초로 상품 판매 업체와 관련된 정보 여부를 판단하는 판단부를 포함하는 상품 판매 업체 관련 글 판단 장치.

청구항 12

제 11항에 있어서,

상기 상품 판매 업체 관련 글 판단 장치는

상기 리뷰 추출부에서 추출된 리뷰의 텍스트 특징을 추출하는 리뷰의 텍스트 특징 추출부; 및

상기 리뷰의 텍스트 특징 추출부에서 추출된 특징을 기초로 상기 리뷰의 텍스트가 상품 판매 업체와 관련된 것 인지의 여부를 분류하는 텍스트 분류부를 더 포함하는 상품 판매 업체 관련 글 판단 장치.

청구항 13

제 11항에 있어서,

상기 상품 판매 업체 관련 글 판단 장치는

상기 이미지 추출부에서 추출된 이미지의 특징을 추출하는 이미지 특징 추출부; 및

상기 이미지 특징 추출부에서 추출된 이미지 특징을 기초로 상기 이미지를 상품 판매 업체와 관련된 것인지의 여부를 분류하는 이미지 분류부를 더 포함하는 상품 판매 업체 관련 글 판단 장치.

명세서

기술분야

[0001] 본 발명은 상품 판매 업체 관련 글 판단 방법 및 그 장치에 관한 것이다.

배경기술

[0002] 일반적으로 인터넷 사용자는 특정 정보를 찾기 위해 그와 관련된 의문점을 검색 엔진에 입력함으로써 원하는 정보를 찾을 수 있다. 그러나 질의어의 의미가 다양할수록 검색결과는 다양하게 나타나고, 사용자가 진정으로 원하는 문서를 찾기 위해 많은 시간을 필요로 할 수 있다.

[0003] 최근 들어, 인터넷 사용자들은 상품을 구매하기에 앞서 다른 사용자의 관련 리뷰도 검색 엔진을 통해 찾는다. 아울러, 레스토랑과 같이 비용을 지급하면서 방문해야 하는 장소에 대해서도 다른 사용자의 평가를 찾는다.

[0004] 이때, 일반적으로 레스토랑 명을 검색 엔진에 질의어로 입력해 주는데, 레스토랑 명이 다른 분야에서도 사용되는 단어, 즉 다양한 의미를 가질 경우 관련 없는 문서도 검색될 수 있다. 검색된 문서 중에는 텍스트 형태의 리뷰 내용뿐만 아니라 관련 이미지를 포함하는 관련성이 높은 문서가 있는가 하면 레스토랑 명과 전화번호 목록만 나타난 관련이 없는 문서도 포함되어 있다.

[0005] 이러한 관련성이 없는 문서가 포함되어 있는 이유는, 전통적인 정보검색 분야에서는 관련성 피드백을 통해서 문서를 찾아냈다. 전통적인 정보검색은 불완전한 초기 질의어를 통해 검색하고, 검색 결과 중 적합한 문서를 사용자가 직접 선택하였다. 이러한 과정이 수 차례 반복되고 질의어를 수정해야 관련성이 높은 문서를 찾아낼 수 있으며 짧은 질의어를 통해 적합 문서를 찾아내기도 어려웠다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 일 실시예는 진술한 바와 같은 문제점들을 해결하기 위해 창출된 것으로서, 본 발명이 해결하고자 하는 과제는 블로그 리뷰에서 텍스트와 이미지 분류를 통해 상품 판매 업체 관련 리뷰의 관련도를 분류하는 상품 판매 업체 관련 글 판단 방법을 제공하는 것이다.

과제의 해결 수단

[0007] 상기한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본원의 제 1측면에 따른 상품 판매 업체 관련 글 판단 방법은 (a) 질의어를 입력하여 검색된 상품 판매 업체의 데이터베이스를 수집하는 단계, (b) 상기 검색된 상품 판매 업체의 데이터베이스를 바탕으로 웹 상의 블로그에서 리뷰를 추출하는 단계, (c) 상기 추출된 리뷰의 텍스트를 분류하는 단계, (d) 상기 추출된 리뷰의 이미지를 분류하는 단계, 및 (e) 상기 분류된 텍스트와 이미지를 바탕으로 상품 판매 업체 관련 데이터를 관련도에 따라 복수의 등급으로 분류할 수 있다.

[0008] 본원의 제 2측면에 따른 상품 판매 업체 관련 글 판단 장치는 질의어를 입력하여 검색된 상품 판매 업체 데이터베이스를 기초로 웹 상의 블로그에서 데이터를 수집하는 데이터 수집부, 상기 데이터 수집부에서 수집된 데이터에서 리뷰를 추출하는 리뷰 추출부, 상기 데이터 수집부에서 수집된 데이터에서 이미지를 추출하는 이미지 추출부, 및 상기 리뷰 추출부와 상기 이미지 추출부에서 추출된 정보를 기초로 상품 판매 업체와 관련된 정보 여부를 판단하는 판단부를 포함할 수 있다.

발명의 효과

- [0009] 본 발명의 일 실시예에 의하면, 텍스트와 이미지를 분류함으로써, 관련 리뷰의 관련도를 분류하는데 보다 정확한 정보를 제공할 수 있다.
- [0010] 또한 관련 리뷰의 관련도를 복수의 등급으로 분류함으로써, 사용자가 질의어에 의해 검색된 리뷰를 관련도에 따라 선택적으로 확인할 수 있다.

도면의 간단한 설명

- [0011] 도 1은 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 장치에 대한 개념도이다.
- 도 2는 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법에 대한 순서도이다.
- 도 3은 본 발 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0012] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.
- [0013] 후술될 내용은 예시적으로 음식점 관련 글을 판단하는 것으로 설명되었으나, 이는 음식점 관련 글뿐만 아니라 모든 상점 및 사용자에게 소비성 물품을 제공하는 업체에 관한 내용으로 해석될 수 있다.
- [0014] 도 1은 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 장치에 대한 개념도이다.
- [0015] 도 1을 참조하면, 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 장치는 데이터 수집부(100), 리뷰 추출부(210), 리뷰의 텍스트 특징 추출부(220), 텍스트 분류부(230), 이미지 추출부(310), 이미지 특징 추출부(320), 이미지 분류부(330), 및 판단부(400)를 포함한다.
- [0016] 데이터 수집부(100)는 질의어를 입력하여 검색된 상품 관련 업체 데이터베이스를 기초로 웹 상의 블로그(1000)에서 데이터를 수집한다. 질의어는 예시적으로 웹 상의 포털 사이트나 검색 사이트를 통해 원하는 정보를 얻도록 입력하는 검색어일 수 있다.
- [0017] 리뷰 추출부(210)는 데이터 수집부(100)에서 수집된 데이터에서 리뷰를 추출한다. 웹 상의 블로그(1000)에서 데이터를 수집하기 때문에 사용자가 원하지 않는 정보도 많이 포함될 수 있다. 사용자가 원하는 정보를 찾기 위해서 블로그(1000)에서 수집된 데이터 중 블로그(1000) 사용자가 작성한 리뷰를 추출한다.
- [0018] 리뷰의 텍스트 특징 추출부(220)는 리뷰 추출부(210)에서 추출된 리뷰의 텍스트 특징을 추출한다. 특징을 추출하는 것은 리뷰에 포함된 텍스트에서 상품 판매 업체에 관련된 글에 포함될 수 있는 특징적인 단어를 추출하는 것이다. 예시적으로 한국어 형태소 분석기를 이용하여 특징적인 단어를 추출할 수 있다.
- [0019] 텍스트 분류부(230)는 리뷰의 텍스트 특징 추출부(220)에서 추출된 특징을 기초로 리뷰의 텍스트를 상품 판매 업체와 관련된 것인지의 여부를 분류한다.
- [0020] 이미지 추출부(310)는 데이터 수집부(100)에서 수집된 데이터에서 이미지를 추출한다. 데이터 수집부(100)를 통해 블로그(1000)에서 수집된 데이터 중 이미지가 포함된 데이터를 수집한다. 이미지를 추출하는 이유는 상품 판매 업체와 관련된 데이터에는 해당 상품 판매 업체에 관한 사진이나 해당 상품 판매 업체에서 판매하는 상품의 사진 등의 이미지가 포함되어 있을 수 있기 때문이다.
- [0021] 이미지 특징 추출부(320)는 이미지 추출부(310)에서 추출된 이미지의 특징을 추출한다. 이미지 특징 추출부(320)는 예시적으로 엠펙7(MPEG7)을 사용하여 이미지로부터 특징값을 추출할 수 있다. 이에 관한 설명은 후술될 부분에서 자세히 하도록 한다.
- [0022] 이미지 분류부(330)는 이미지 특징 추출부(320)에서 추출된 이미지 특징을 기초로 이미지를 상품 판매 업체와 관련된 것인지의 여부를 분류한다. 이미지 분류부(330)는 예시적으로 SVM(support vector machine)분류 방법을 사용할 수 있다.
- [0023] 판단부(400)는 리뷰 추출부(210), 리뷰의 텍스트 특징 추출부(220), 및 텍스트 분류부(230)를 통해 분류된 리뷰

와 이미지 추출부(310), 이미지 특징 추출부(320), 및 이미지 분류부(330)를 통해 분류된 이미지를 바탕으로 상품 판매 업체와 관련된 정보 여부를 판단한다.

[0024] 예시적으로 리뷰 및 이미지가 모두 상품 판매 업체와 관련된 정보일 경우의 관련도를 ‘상’, 리뷰 및 이미지 중 하나가 상품 판매 업체와 관련된 정보일 경우의 관련도를 ‘중’, 리뷰 및 이미지 모두 상품 판매 업체와 관련된 정보가 아닐 경우의 관련도를 ‘하’ 로 분류한다.

[0025] 이하에서는 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법에 대해 살핀다.

[0026] 참고로, 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법은 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 장치로 상품 판매 업체 관련 글을 판단하는 방법에 관한 것으로, 앞서 살핀 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 장치에서 설명한 구성과 유사한 구성에 대해서는 동일한 도면 부호를 적용하고 이에 대한 설명은 간략히 하거나 생략하기로 한다.

[0027] 도 2는 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법에 대한 순서도이다.

[0028] 이러한 본 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법은 먼저, 질의어를 입력하여 검색된 상품 판매 업체의 데이터베이스를 수집(S10)한다.

[0029] 질의어는 사용자가 원하는 정보를 검색하기 위해 입력하는 단어로, 예시적으로 상품 판매 업체의 이름이나 상품의 이름 또는 그와 관련된 단어를 포털 사이트 검색창에 입력한다. 이를 통해 상품 판매 업체의 데이터베이스를 수집(S10)할 수 있다.

[0030] 다음으로, 검색된 상품 판매 업체의 데이터베이스를 바탕으로 웹 상의 블로그(1000)에서 리뷰를 추출(S20)한다. 예시적으로, 블로그(1000)의 게시물 중에서 리뷰에 관한 요약물 추출하여 기계 학습 알고리즘의 한 종류인 문서 분류 방법을 통해 리뷰에 상품 판매 업체와 관련된 단어의 존재 유무를 판단하여 블로그에서 리뷰를 추출한다.

[0031] 다음으로, 추출된 리뷰의 텍스트를 분류(S30)한다. 텍스트 분류(S30)는 텍스트 전처리 단계, 특징 추출 단계, 및 확률 기반 분류 단계를 포함한다.

[0032] 텍스트 전처리 단계는 추출된 리뷰의 요약물로부터 HTML 태그, 숫자, 기호 등을 제거한다.

[0033] HTML(hypertext markup language)는 웹 문서를 만들기 위하여 사용하는 기본적인 프로그래밍 언어의 한 종류이다. 문서의 글자크기, 글자색, 글자모양, 그래픽, 문서이동 등을 정의하는 명령어로 홈페이지를 작성하는데 쓰인다. 이 때 HTML의 명령어로 태그(tag)라고 하는 꺾쇠괄호를 사용하는데 이를 텍스트 전처리 단계를 통해 제거한다. 또한 리뷰의 요약물로부터 숫자, 기호 등 상품 판매 업체 관련 리뷰를 판단하기 위해 필요한 정보가 아닌 것을 제거한다.

[0034] 텍스트 특징 추출 단계는 리뷰에서 특징적인 단어를 추출하는 단계로 한국어 형태소 분석을 통해 단어의 품사에 따라 특징을 추출한다. 형태소란 의미의 기능을 부여하는 언어의 형태론적 수준에서의 최소단위를 말한다.

[0035] 예시적으로, 상품 판매 업체 관련 리뷰에는 특정 단어가 많이 포함되어 있을 수 있다. 이러한 특정 단어를 분석하고 정리하기 위해 단어의 최소 단위인 형태소를 분석하고 정리하여 비교 대상으로 사용할 수 있는 훈련 모델을 만들 수 있다. 이러한 형태소 분석을 통해 상품 판매 업체 관련 리뷰의 특징을 추출하여 후술될 텍스트 분류 단계를 통해 상품 판매 업체 관련 텍스트를 분류하는 자료로 사용할 수 있다.

[0036] 텍스트 분류 단계는 구축된 확률 모델을 바탕으로 새롭게 추출된 특징값에 대한 확률을 구하여 분류 결과를 추출한다. 구축된 확률 모델은 전술한 형태소 분석을 통해 상품 판매 업체 관련 리뷰에 포함될 수 있는 특정 형태소, 특정 품사, 및 특정 단어를 정리하여 비교 대상으로 사용한다. 이러한 확률 모델을 바탕으로 새로운 리뷰가 검색되면 그 안에서 형태소 분류를 통해 특정 형태소, 특정 품사, 및 특정 단어를 분류하게 되고 이를 확률 모델과 비교하여 결과를 추출한다. 결과는 상품 판매 업체와 관련이 있는 리뷰 또는 상품 판매 업체와 관련이 없는 리뷰로 분류할 수 있다.

[0037] 다음으로, 추출된 리뷰의 이미지를 분류(S40)한다.

[0038] 리뷰의 이미지 분류(S40)는 이미지 전처리 단계, 특징 추출 단계, 및 분류 단계를 포함한다.

[0039] 이미지 전처리 단계는 후술될 이미지의 특징 추출을 위해 이미지의 크기를 정규화한다.

[0040] 이미지 특징 추출 단계는 엠펙7(MPEG7)을 사용하여 이미지로부터 정량화된 특징값을 추출한다. 엠펙(MPEG)은 Moving Picture Experts Group의 약자로 동영상 압축하고 코드로 표현하는 방법의 표준을 만드는 전문가 그룹

을 뜻한다. 엠팩7은 이 전문가 그룹에서 개발한 것으로 동영상 데이터 검색과 전자상거래 등에 적합하도록 개발된 차세대 동영상 압축 재생기술이다. 원하는 그림이나 영화의 한 장면 또는 특정 음악의 일부 등을 검색할 용도로 시작되었다. 색상이나 물체의 모양에 관한 정보를 입력하는 것만으로 웹에서 필요로 하는 멀티미디어 자료를 찾을 수 있는 기술이다.

이러한 애플7은 세분화된 디스크립터(descriptor)를 포함하고 있다. 예시적으로, 컬러 스트럭처 디스크립터(color structure descriptor), 컬러 레이아웃 디스크립터(color layout descriptor), 스케러블 컬러 디스크립터(scalable color descriptor), 에지 히스토그램 디스크립터(edge histogram descriptor), 호모지니어스 텍스처 디스크립터(homogeneous texture descriptor), 리전 베이스 셰이프 디스크립터(region-based shape descriptor) 등을 포함한다.

여러 종류의 디스크립터는 색, 에지, 및 형태 중 하나로 기반을 설정하여 특징값을 추출한다.

특히 방향성 예지 4종류와 비방향성 예지 하나로 이미지가 갖는 예지의 공간적 분포를 표현하는 예지 히스토그램 디스크립터와 색상의 공간적 분포를 나타내는 컬러 레이아웃 디스크립터를 이용하여 이미지로부터 정량화된 특징값을 추출할 수 있다.

이미지 분류 단계는 SVM분류 방법으로 기존에 구축된 훈련 모델을 적용하여 분류한다. 기존에 구축된 훈련 모델이란 전술한 эм펙7을 통해 음식 이미지로 분류할 수 있는 대상을 훈련 모델로 설정해 둔 것이다.

또한 SVM은 support vector machine의 약자로, 기계학습알고리즘의 한 종류이다. 이미지 분류에 사용할 수 있는 기계학습알고리즘은 SVM외에도 Naive Bayes, RBFnetwork, C4.5, kNN 등이 있다. 하지만 최대 마진을 고려하여 초평면을 찾아 긍정과 부정을 분류할 수 있는 모든 정보를 갖고 있는 벡터를 찾고 나머지 벡터는 폐기하는 SVM이 이미지 분류 단계에 사용하기에 가장 적합한 기계학습알고리즘이다.

상품 판매 업체 관련 글 판단 방법은 마지막으로, 분류된 텍스트와 이미지를 바탕으로 상품 판매 업체 관련 리뷰의 관련도를 복수의 등급으로 분류(S50)한다.

예시적으로, 리뷰의 관련도 분류(S50)는 리뷰의 텍스트 분류(S30) 및 리뷰의 이미지 분류(S40) 모두 상품 판매 업체 관련 리뷰로 분류된 경우의 관련도를 ‘상’으로 분류한다. 또한 리뷰의 텍스트 분류(S30) 및 리뷰의 이미지 분류(S40) 중 하나가 상품 판매 업체 관련 리뷰로 분류된 경우의 관련도를 ‘중’으로 분류한다. 그리고 리뷰의 텍스트 분류(S30) 및 리뷰의 이미지 분류(S40) 모두 상품 판매 업체 관련 리뷰가 아닌 경우의 관련도를 ‘하’로 분류한다.

도 3은 본 발 발명의 일 실시예에 따른 상품 판매 업체 관련 글 판단 방법을 설명하기 위한 도면이다.

도 3을 참조하여 전술한 내용을 요약하여 설명하면, 예시적으로 질의어를 입력하여 크로울러를 통해 상품 판매 업체 정보 사이트로부터 상품 판매 업체 이름을 수집한다. 예시적으로 수집한 상품 판매 업체 이름을 각각의 질의어로 하여 bing.com에서 검색한다. Bing.com 외에 검색엔진, 포털 사이트나 인터넷 상의 검색 사이트 어느 곳에서 검색 작업을 수행하여도 무관하다.

Bing.com에서 검색된 검색 결과 중 블로그 요약의 텍스트와 이미지를 추출하게 되고, 추출된 텍스트와 이미지를 분류하여 상품 판매 업체 관련 리뷰의 관련도를 분류하게 된다.

전술한 본원의 설명은 예시를 위한 것이며, 본원이 속하는 기술분야의 통상의 지식을 가진 자는 본원의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

본원의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본원의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

100: 데이터 수집부 210: 리뷰 추출부

220: 리뷰의 텍스트 특징 추출부 230: 텍스트 분류부

310: 이미지 추출부

330: 이미지 분류부

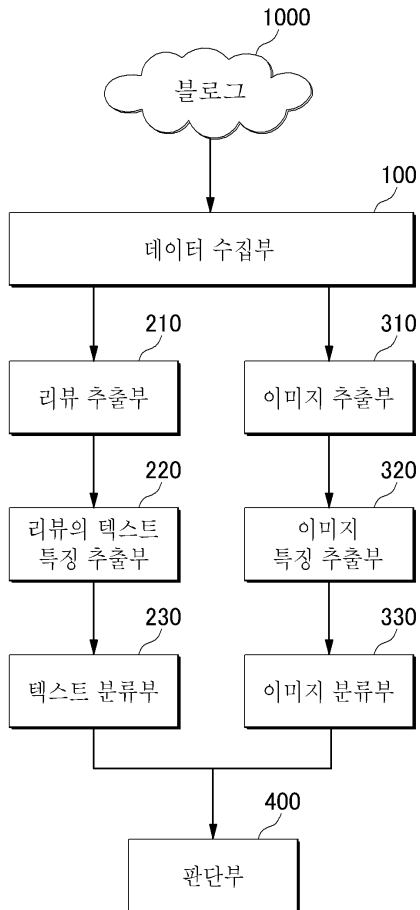
1000: 블로그

320: 이미지 특징 추출부

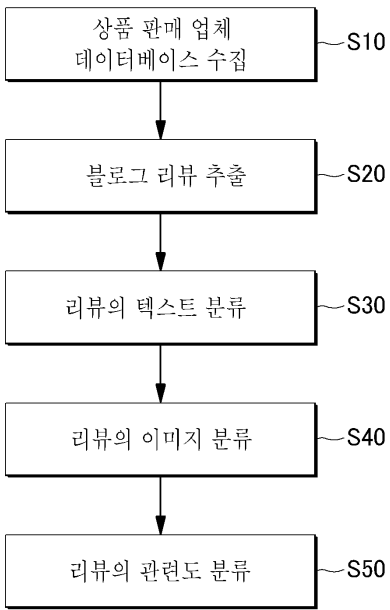
400: 판단부

도면

도면1



도면2



도면3

