(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0018452
(43) 공개일자 2014년02월13일

(51) 국제특허분류(Int. Cl.)
G06Q 50/10 (2012.01)
(21) 출원번호 10-2012-0078866
(22) 출원일자 2012년07월19일
심사청구일자 2012년07월19일

(71) 출원인
고려대학교 산학협력단
서울시 성북구 안암로 145 (안암동)
(72) 발명자
강재우
서울 송파구 잠실로 62, 332동 1003호 (잠실동, 트리지움)
최재훈
경기 용인시 수지구 성복2로 126, 310동 1902호 (성복동, 성동마을LG빌리지3차아파트)
(뒷면에 계속)
(74) 대리인
특허법인엠에이피에스

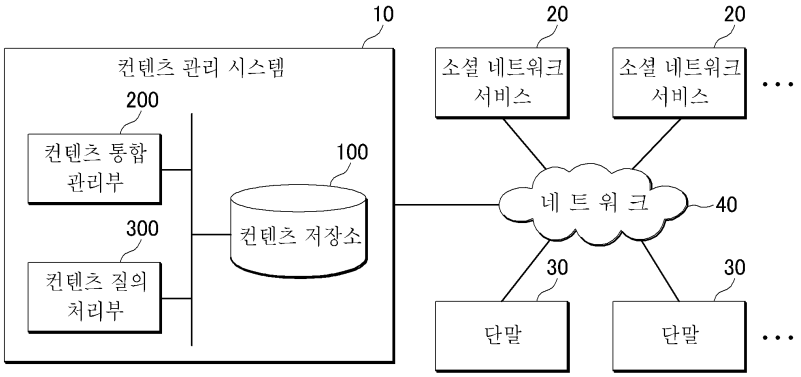
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 **컨텐츠 관리 시스템 및 방법**

(57) 요 약

본 발명은 네트워크에 분산되어 있는 사용자 생성 컨텐츠를 단일 사용자 인터페이스를 통해 통합 관리하는 컨텐츠 통합 관리부; 상기 사용자 생성 컨텐츠가 저장되는 컨텐츠 저장소; 상기 사용자 생성 컨텐츠에 대한 댓글에 기초하여, 사용자로부터의 질의에 대한 각 엔티티(entity)의 적합성을 심사하고, 상기 컨텐츠 저장소에서 상기 적합성 점수가 높은 엔티티 순으로 대응하는 상기 사용자 생성 컨텐츠를 반환하는 컨텐츠 질의 처리부;를 포함하되, 상기 컨텐츠 질의 처리부는 상기 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)를 구축하는 인덱스 구축부; 상기 인덱스가 저장되는 인덱스 저장부; 및 상기 인덱스를 기초로 상기 질의를 처리하는 질의 처리부;를 포함하며, 상기 엔티티는 상기 사용자 생성 컨텐츠 및 상기 사용자 생성 컨텐츠를 생성한 사용자를 포함하는 컨텐츠 관리 시스템을 제공한다.

대 표 도 - 도1



(72) 발명자

임상락

서울 송파구 송이로36길 2, 크로바 301호 (문정동)

최동희

서울 중구 동호로8길 28-5, 101호 (신당동, 약수원빌)

이 발명을 지원한 국가연구개발사업

과제고유번호 000476220111

부처명 중소기업청

연구사업명 산학연공동기술개발사업

연구과제명 모바일 사진 컨텐츠 분석을 통해 지능형 멀티미디어 SNS 기능을 극대화 하는 Insight Search Engine의 개발

기 여 율 1/1

주관기관 고려대학교 산학협력단

연구기간 2011.06.01 ~ 2012.05.31

특허청구의 범위

청구항 1

컨텐츠 관리 시스템에 있어서,

네트워크에 분산되어 있는 사용자 생성 컨텐츠를 단일 사용자 인터페이스를 통해 통합 관리하는 컨텐츠 통합 관리부;

상기 사용자 생성 컨텐츠가 저장되는 컨텐츠 저장소;

상기 사용자 생성 컨텐츠에 대한 댓글에 기초하여, 사용자로부터의 질의에 대한 각 엔티티(entity)의 적합성을 심사하고, 상기 컨텐츠 저장소에서 상기 적합성 점수가 높은 엔티티 순으로 대응하는 상기 사용자 생성 컨텐츠를 반환하는 컨텐츠 질의 처리부;를 포함하되,

상기 컨텐츠 질의 처리부는

상기 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)를 구축하는 인덱스 구축부;

상기 인덱스가 저장되는 인덱스 저장부; 및

상기 인덱스를 기초로 상기 질의를 처리하는 질의 처리부;를 포함하며,

상기 엔티티는 상기 사용자 생성 컨텐츠 및 상기 사용자 생성 컨텐츠를 생성한 사용자를 포함하는 컨텐츠 관리 시스템.

청구항 2

제 1 항에 있어서,

상기 사용자 생성 컨텐츠는 하나 이상의 소셜 네트워크 서비스 사이트에 등록된 사진 또는 동영상 컨텐츠를 포함하며,

상기 컨텐츠 통합 관리부는 상기 사용자 생성 컨텐츠에 대한 조회, 수정, 업로드, 및 다운로드 서비스를 제공하는 컨텐츠 관리 시스템.

청구항 3

제 1 항에 있어서,

상기 인덱스 구축부는

상기 댓글을 하나 이상의 세그먼트(segment)로 나누어 인덱싱(indexing)하되,

상기 세그먼트는 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장인 컨텐츠 관리 시스템.

청구항 4

제 1 항에 있어서,

상기 인덱스 구축부는

상기 인덱스를 역 인덱스(inverted index) 형태로 구축하는 컨텐츠 관리 시스템.

청구항 5

제 1 항에 있어서,

상기 인덱스 구축부는

상기 댓글에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱(parsing)을 수행하는 댓글 파싱부;및

상기 댓글을 상기 세그먼트로 분할하는 댓글 분할부;를 포함하는 컨텐츠 관리 시스템.

청구항 6

제 1 항에 있어서,

상기 질의 처리부는

상기 질의에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱을 수행하는 질의 파싱부;

상기 세그먼트에 대해 상기 질의에 대한 적합성 점수를 심사하는 세그먼트 심사부; 및

상기 세그먼트 심사부가 산출한 세그먼트 점수 및 상기 사용자 생성 컨텐츠에 대해 포스팅된 댓글의 세그먼트의 개수를 기초로, 상기 엔티티에 대해 상기 질의에 대한 적합성 점수를 심사하는 엔티티 심사부;를 포함하는 컨텐츠 관리 시스템.

청구항 7

제 6 항에 있어서,

상기 세그먼트 심사부는

상기 질의에 포함된 키워드를 포함하는 세그먼트 개수를 전체 세그먼트 개수와 비교하여 심사하는 세그먼트 개수 심사부;

댓글 작성자의 신뢰도를 심사하는 작성자 심사부;

상기 댓글의 상기 사용자 생성 컨텐츠와의 관련도를 심사하는 댓글 품질 심사부;

상기 세그먼트가 상기 엔티티에 대해 긍정적인 감상을 나타내는지 부정적인 감상을 나타내는지 심사하는 세그먼트 선호극성 심사부;

상기 댓글이 언제 작성되었는지를 심사하는 댓글 작성시기 심사부; 및

상기 댓글과 상기 질의에 포함된 키워드의 동의어 확장 집합을 심사하는 확장 집합 심사부;를 포함하는 컨텐츠 관리 시스템.

청구항 8

제 7 항에 있어서,

상기 작성자 심사부는 페이지랭크(PageRank) 기법을 사용하는 컨텐츠 관리 시스템.

청구항 9

제 7 항에 있어서,

상기 댓글 품질 심사부는 상기 댓글의 스팸(spam) 여부를 심사하는 컨텐츠 관리 시스템.

청구항 10

제 7 항에 있어서,

상기 세그먼트 선호극성 심사부는 상기 세그먼트가 나타내는 감상을 부호가 있는 점수로 나타내는 컨텐츠 관리 시스템.

청구항 11

컨텐츠 관리 시스템을 사용하는 컨텐츠 관리 방법에 있어서,

(a) 네트워크에 분산되어 있는 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)를 구축 및 갱신하는 단계;

(b) 상기 인덱스에 기초하여 사용자 질의에 적합한 사용자 생성 컨텐츠를 반환하는 단계;를 포함하되,

상기 (a) 단계는

상기 댓글에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱(parsing)을 수행하는 단계;및

상기 댓글을 하나 이상의 세그먼트(segment)로 분할하는 단계:

상기 세그먼트에 대한 역 인덱스(inverted index)를 구축하는 단계;를 포함하고,

상기 (b) 단계는

상기 질의에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱을 수행하는 단계;

상기 세그먼트에 대해 상기 질의에 대한 적합성 점수를 심사하는 단계; 및

상기 세그먼트 심사 단계가 산출한 세그먼트 점수 및 상기 사용자 생성 컨텐츠에 대해 포스팅된 댓글의 세그먼트의 개수를 기초로, 상기 사용자 생성 컨텐츠에 대해 상기 질의에 대한 적합성 점수를 심사하는 단계;를 포함하며,

상기 세그먼트는 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장인 컨텐츠 관리 방법.

청구항 12

제 11 항에 있어서,

상기 세그먼트 심사 단계는

상기 질의에 포함된 키워드를 포함하는 세그먼트 개수를 전체 세그먼트 개수와 비교하여 심사하는 단계;

댓글 작성자의 신뢰도를 심사하는 단계;

상기 댓글의 상기 사용자 생성 컨텐츠와의 관련도를 심사하는 단계;

상기 세그먼트가 상기 엔티티에 대해 긍정적인 감상을 나타내는지 부정적인 감상을 나타내는지 심사하는 단계;

상기 댓글이 언제 작성되었는지를 심사하는 단계; 및

상기 댓글과 상기 질의에 포함된 키워드의 동의어 확장 집합을 심사하는 단계;를 포함하는 컨텐츠 관리 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 컨텐츠 관리 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 최근 소셜 네트워크 서비스가 큰 인기를 끌고 있다. 특히 사진이나 동영상 등 사용자가 생성한 컨텐츠를 업로드하고 다른 사용자와 공유할 수 있는 서비스를 이용하는 사용자가 많다. 그러나 사용자 생성 컨텐츠가 여러 소셜 네트워크 서비스 사이트에 분산 등록되어 있어, 여러 다른 서비스 사이트에 게시한 컨텐츠의 관리가 힘들어졌다는 불편함이 있다.

[0003] 이에 따라 분산된 사용자 생성 컨텐츠를 한 곳에서 통합 관리할 수 있는 서비스가 필요하게 되었다. 더구나 최근 스마트폰이 널리 보급됨에 따라, 스마트폰을 통한 통합 관리 서비스를 제공할 필요가 생겨났다. 따라서 본 발명은 사용자 생성 컨텐츠를 통합 관리할 수 있는 환경을 제공하려 한다.

[0004] 한편, 페이스북®, 싸이월드® 등 소셜 네트워크 서비스의 산업적 성공에 힘입어, 사용자에게 편의를 제공하는 서비스들이 다양하게 시도되고 있다. 예를 들어, 사용자의 프로필을 분석하여 지인을 추천하는 서비스(같은 학교 출신, 같은 지역 출신, 비슷한 나이, 원하는 성별 등), 스마트폰 환경에서 위치 정보를 이용하여 비슷한 지역에 있는 사람을 추천해 주는 서비스 등이 있다.

[0005] 이에 따라 본 발명의 중요한 목표는 사용자 혹은 컨텐츠를 추천함에 있어서, 컨텐츠에 달린 댓글을 사용함으로써, 종래 기술보다 컨텐츠 검색 및 추천 성능을 한층 더 높이는 것이다. 단순히 문서에 포함되어 있는 특정 단어의 빈도에 기반하여 검색 또는 추천하는 것이 아니라, 댓글을 단 사용자의 해당 컨텐츠에 대한 감상(sentiment)까지도 의미 단위로 나누어 분석함으로써, 질의에 맞는 컨텐츠 및 사용자를 더욱 효과적으로 검색, 추천하려 한다.

[0006] 이와 관련하여 한국공개특허 제10-2011-0125905호("패턴분석을 이용한 댓글 관리 방법, 그 시스템, 및 웹 서버")은 스팸성 댓글 을 차단하는 구성을 개시하고 있다.

[0007] 또한, 한국공개특허 제10-2011-0051678("댓글 평가를 이용한 사용자 그룹 형성 방법과 시스템 및 그 시스템을 이용하기 위한 장치")는 댓글을 수집하고 평가하는 구성이 개시되어 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 문제를 해결하기 위한 것으로서, 그 목적은 컨텐츠에 달린 댓글에 기반하여 컨텐츠에 대한 효과적인 검색 및 추천 서비스를 제공하는 통합 컨텐츠 관리 시스템 및 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 상기와 같은 목적을 달성하기 위한 본 발명의 제 1 측면에 따른 컨텐츠 관리 시스템은, 네트워크에 분산되어 있는 사용자 생성 컨텐츠를 단일 사용자 인터페이스를 통해 통합 관리하는 컨텐츠 통합 관리부; 상기 사용자 생성 컨텐츠가 저장되는 컨텐츠 저장소; 상기 사용자 생성 컨텐츠에 대한 댓글에 기초하여, 사용자로부터의 질의에 대한 각 엔티티(entity)의 적합성을 심사하고, 상기 컨텐츠 저장소에서 상기 적합성 점수가 높은 엔티티 순으로 대응하는 상기 사용자 생성 컨텐츠를 반환하는 컨텐츠 질의 처리부;를 포함하되, 상기 컨텐츠 질의 처리부는 상기 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)를 구축하는 인덱스 구축부; 상기 인덱스가 저장되는 인덱스 저장부; 및 상기 인덱스를 기초로 상기 질의를 처리하는 질의 처리부;를 포함하며, 상기 엔티티는 상기 사용자 생성 컨텐츠 및 상기 사용자 생성 컨텐츠를 생성한 사용자를 포함하는 것을 특징으로 한다.

[0010] 상기와 같은 목적을 달성하기 위한 본 발명의 제 2 측면에 따른 컨텐츠 관리 시스템을 사용하는 컨텐츠 관리 방법은, (a) 네트워크에 분산되어 있는 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)를 구축 및 갱신하는 단계; (b) 상기 인덱스에 기초하여 사용자 질의에 적합한 사용자 생성 컨텐츠를 반환하는 단계;를 포함하되, 상기 (a) 단계는 상기 댓글에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱(parsing)을 수행하는 단계;및 상기 댓글을 하나 이상의 세그먼트(segment)로 분할하는 단계: 상기 세그먼트에 대한 역 인덱스(inverted index)를 구축하는 단계;를 포함하고, 상기 (b) 단계는 상기 질의에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱을 수행하는 단계; 상기 세그먼트에 대해 상기 질의에 대한 적합성 점수를 심사하는 단계; 및 상기 세그먼트 심사 단계가 산출한 세그먼트 점수 및 상기 사용자 생성 컨텐츠에 대해 포스팅된 댓글의 세그먼트의 개수를 기초로, 상기 사용자 생성 컨텐츠에 대해 상기 질의에 대한 적합성 점수를 심사하는 단계;를 포함하며, 상기 세그먼트는 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장인 것을 특징으로 한다.

발명의 효과

[0011] 본 발명은 컨텐츠 관리 시스템 및 방법에 있어, 네트워크에 분산되어 있는 사용자 생성 컨텐츠를 단일 사용자 인터페이스를 통해 통합 관리하고, 방대한 사용자 생성 컨텐츠를 효과적으로 검색, 추천하는 효과를 얻는다.

[0012] 구체적으로, 사용자 생성 컨텐츠에 포스팅된 댓글을 최소 의미 단위로 나누어 각각에 대해 사용자 질의와의 적합성을 심사함으로써 컨텐츠 검색 효과를 높일 수 있다.

[0013] 또한, 질의 처리시 사용자의 댓글에 포함되어 있는 컨텐츠에 대한 감상(sentiment)을 고려함으로써 사용자 질의에 좀더 적합한 컨텐츠를 추천할 수 있다.

[0014] 이에 따라, 방대한 사용자 생성 컨텐츠에 묻혀 현재는 거의 접근되지 않는 컨텐츠를 재활용하는 효과를 기대할 수 있다.

[0015] 또한, 여러 서비스에서의 인맥의 연동 등 사용자 편의 효과도 기대할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일실시예에 따른 컨텐츠 관리 시스템의 구조를 도시함.

도 2는 본 발명의 일실시예에 따른 컨텐츠 질의 처리부의 구조를 도시함.

도 3은 본 발명의 일실시예에 따른 세그먼트 심사부의 구조를 도시함.

도 4는 본 발명의 일실시예에 따른 인덱스의 구조를 도시함.

도 5는 본 발명의 일실시예에 따른 컨텐츠 관리 방법의 흐름을 도시함.

도 6은 본 발명의 일실시예에 따른 인덱스 구축 단계의 흐름을 도시함.

도 7은 본 발명의 일실시예에 따른 질의 처리 단계의 흐름을 도시함.

도 8은 본 발명의 일실시예에 따른 세그먼트 심사 단계의 흐름을 도시함.

발명을 실시하기 위한 구체적인 내용

[0017] 아래에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0018] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0019] 도 1은 본 발명의 일실시예에 따른 컨텐츠 관리 시스템(10)의 구조를 도시하고 있다.

[0020] 먼저 본 개시에서 사용할 용어를 정의하면, 사용자 생성 컨텐츠란 사용자가 만들어낸 창작물이나 허가받아 사용하는 타인의 자료를 지칭하며, 형태에 제한이 없다. 대표적인 사용자 생성 컨텐츠에는 사진, 글, 동영상이 있다. 엔티티(entity)란 사용자 생성 컨텐츠, 컨텐츠를 생성한 사용자, 사용자 그룹 등을 모두 포함하는 개념이다.

[0021] 컨텐츠 관리 시스템(10)은 사용자 생성 컨텐츠가 저장되는 컨텐츠 저장소(100), 사용자 생성 컨텐츠를 통합 관리하는 컨텐츠 통합 관리부(200), 및 질의를 처리하여 가장 적합한 사용자 생성 컨텐츠를 반환하는 컨텐츠 질의 처리부(300)를 포함하며, 네트워크(40)를 통해 하나 이상의 소셜 네트워크 서비스(20) 및 하나 이상의 단말(30)과 연결되어 있다.

[0022] 네트워크(40)는 모든 종류의 유, 무선 네트워크를 포함하며, 단말(30)은 네트워크(40)에 연결할 수 있는 모든 종류의 단말(30)을 포함한다. 예를 들어, 네트워크(40)는 유, 무선 인터넷일 수 있으며, 단말(30)은 개인용 컴퓨터, 스마트폰, 또는 PDA일 수 있다.

[0023] 컨텐츠 통합 관리부(200)는 네트워크(40)에 분산되어 있는 여러 소셜 네트워크 서비스(20)로부터 사용자 생성 컨텐츠를 수집하여 컨텐츠 저장소(100)에 저장한다. 각 사용자는 단말(30)을 통해, 컨텐츠 통합 관리부(200)로부터 단일 사용자 인터페이스를 제공받는다.

[0024] 이러한 통합 관리 환경은 사용자 생성 컨텐츠의 등록, 조회, 삭제, 수정 등의 기본적인 서비스 뿐 아니라, 분산되어 있는 사용자 생성 컨텐츠를 컨텐츠 저장소(100)에 통합 저장하거나 다른 소셜 네트워크 서비스(20)로 이동시키는 서비스도 제공할 수 있으며, 사용자 생성 컨텐츠를 여러 소셜 네트워크 서비스(20)에 일괄 등록하는 서비스도 제공할 수 있다. 또한 사용자는 컨텐츠 통합 관리부(200)가 제공하는 통합 관리 환경을 통해 각 소셜 네트워크 서비스(20)에 게시되어 있는 댓글이나 여타 메타데이터에 접근할 수도 있다.

[0025] 또한 계정을 관리하는 것이나 사용자 생성 컨텐츠에 대한 권한을 설정하는 것도 가능하다. 권한은 사용자의 동의에 의해 설정되며, 각 컨텐츠의 다운로드 허용 여부, 다른 사용자의 컨텐츠 조회 및 다운로드 허용 여부, 컨텐츠 검색 허용 여부 등을 포함한다. 또한 각 소셜 네트워크 서비스(20)에 등록되어 있는 컨텐츠와 자동 연동하여 컨텐츠 저장소(100)를 자동 업데이트하는 시점을 설정하게 할 수도 있다.

[0026] 이와 같이, 컨텐츠 통합 관리부(200)가 제공하는 사용자 편의 서비스의 범위에는 제한이 없다.

[0027] 예를 들어, 사용자는 개인용 컴퓨터에 설치된 어플리케이션, 웹 어플리케이션, 또는 스마트폰 앱을 통해 싸이월드®에 게시한 사진들을 페이스북®으로 일괄 이동할 수 있다. 또한 최근 여행에서 찍은 사진을 싸이월드®와 페이스북®에 동시에 업로드할 수도 있다. 또한 선택한 컨텐츠를 컨텐츠 관리 시스템(10)의 컨텐츠 저장소(10

0)로 저장하는 것도 가능하다.

[0028] 이러한 통합 관리 환경을 제공하기 위해, 컨텐츠 통합 관리부(200)는 각 소셜 네트워크 서비스(20)가 제공하는 API(application programming interface)를 사용할 수 있다.

[0029] 컨텐츠 질의 처리부(300)는 사용자 생성 컨텐츠에 대한 댓글에 기초하여, 사용자로부터의 질의에 대한 각 엔티티(entity)의 적합성을 심사하고, 상기 컨텐츠 저장소(100)에서 상기 적합성 점수가 높은 엔티티 순으로 대응하는 상기 사용자 생성 컨텐츠를 반환한다.

[0030] 도 2는 본 발명의 일실시예에 따른 컨텐츠 질의 처리부(300)의 구조를 도시하고 있다.

[0031] 컨텐츠 질의 처리부(300)는 사용자 생성 컨텐츠에 포스팅된 댓글에 대한 인덱스(index)가 저장되는 인덱스 저장부(310), 인덱스를 구축, 갱신하는 인덱스 구축부(320), 및 인덱스를 기초로 질의를 처리하는 질의 처리부(330)를 포함한다.

[0032] 인덱스 구축부(320)는 댓글을 하나 이상의 세그먼트(segment)로 나누어 인덱싱(indexing)한다. 세그먼트는 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장이다. 이렇게 전체 댓글이 아닌 최소 의미 단위(minimal semantic unit)로 나누어 각각을 인덱싱함으로써, 본 발명의 일실시예에 따른 컨텐츠 질의 처리부(300)는 사용자의 질의에 더욱 적합한 사용자 생성 컨텐츠를 포함하는 엔티티를 찾아낼 수 있다.

[0033] 예를 들어, 사용자 생성 컨텐츠가 강아지와 고양이를 찍은 사진이고, 거기에 달린 댓글이 "강아지는 예쁜데, 고양이는 못생겼다"라고 하자. 이 댓글은 강아지와 고양이라는 서로 다른 두가지 특징에 대해 상반된 감상(sentiment)를 표현하고 있다.

[0034] 종래의 검색 엔진은 이를 의미적 단위로 나누어 인덱싱하지 않고 전체 문서로 인덱싱하기 때문에, "예쁜 고양이가 나온 사진"이라는 질의에 대해 이 사진이 반환될 수 있었다. 질의에 포함되어 있는 "예쁘다"와 "고양이"라는 키워드를 해당 댓글이 모두 포함하고 있기 때문이다.

[0035] 반면, 본 발명의 일실시예에 따른 컨텐츠 질의 처리부(300)는 "강아지는 예쁜데"와 "고양이는 못생겼다"를 각각의 세그먼트로 분할하여 인덱싱하기 때문에 이 사진이 해당 질의에 적합하지 않은 엔티티임을 판단해낼 수 있다. 따라서 좀더 적합한 다른 사용자 생성 컨텐츠를 반환할 수 있게 된다.

[0036] 인덱스 구축부(320)는 댓글 파싱부(322)와 댓글 분할부(324)를 포함한다.

[0037] 댓글 파싱부(322)는 댓글에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱(parsing)을 수행한다. 예를 들어, "강아지는 예쁜데"는 "강아지는 예쁘다" 또는 "강아지가 예쁘다"와 의미상 차이가 없으므로, 동일하게 처리될 수 있도록 전처리된다. 파싱은 어구를 분석하는 것을 의미하며, 해당 댓글에 어떤 키워드(keyword) 또는 용어(term)가 들어있는지를 해부하는 것이다.

[0038] 댓글 분할부(324)는 댓글을 전술한 바와 같이 최소 의미 단위의 세그먼트로 분할하는 일을 담당한다. 예를 들어, 댓글 파싱부(322)가 전처리 및 파싱한 댓글이 "강아지, 예쁘다, 고양이, 못생기다"라면, 댓글 분할부(324)가 이를 최소 의미 단위로 분할한 결과는 "강아지, 예쁘다"와 "고양이, 못생기다"일 수 있다.

[0039] 이때 여기에 기술한 것은 일실시예일 뿐이며, 전처리 방법이나 파싱 방법, 댓글 분할 방법에는 제한이 없으며 구현자의 선택에 달려 있다. 예를 들어, 댓글 파싱부(322)와 댓글 분할부(324)를 거친 결과가 "강아지 예쁘"와 "고양이 못생기"일 수도 있는 것이다.

[0040] 각 세그먼트는 역 인덱스(inverted index)로 인덱싱된다. 자세한 내용은 도 4를 통해 후술한다.

[0041] 질의 처리부(330)는 질의 파싱부(332), 세그먼트 심사부(334), 및 엔티티 심사부(336)를 포함한다.

[0042] 질의 파싱부(332)는 질의에 대해 불용어 및 어미 제거를 포함하는 전처리 및 파싱을 수행한다. 즉, 전술한 댓글 파싱부(322)와 유사한 일을 담당하는데, 그 대상이 질의이다. 예를 들어, "예쁜 고양이가 나온 사진"이라는 질의에 대해 질의 대상이 사진으로 한정되는 경우 댓글 파싱부(322)는 "고양이, 예쁘다"를 산출할 수 있다. 마찬가지로 전처리 방법이나 파싱 방법에는 제한이 없으며 구현자의 선택에 달려 있다.

[0043] 파싱된 질의에 적합한 엔티티를 찾기 위해 질의 처리부(330)는 우선 역 인덱싱된 각 세그먼트에 대해 질의와의 적합성 여부를 심사한 후 그 결과를 바탕으로 각 엔티티에 대해 적합성 여부를 심사한다. 각 세그먼트에 대한 심사는 세그먼트 심사부(334)가, 엔티티에 대한 심사는 엔티티 심사부(336)가 수행한다.

[0044] 도 3은 본 발명의 일실시예에 따른 세그먼트 심사부(334)의 구조를 도시하고 있다.

[0045] 세그먼트 심사부(334)는 세그먼트 개수 심사부(334a), 작성자 심사부(334b), 댓글 품질 심사부(334c), 세그먼트 선호극성 심사부(334d), 댓글 작성시기 심사부(334e), 및 확장 집합 심사부(334f)를 포함한다.

[0046] 세그먼트 개수 심사부(334a)는 전체 댓글 문서 집합에서 전체 세그먼트의 개수와 질의 키워드가 포함된 세그먼트의 개수를 비교하여 그 비율을 구한다. 즉, (전체 세그먼트 개수+1)/(질의 키워드가 포함된 세그먼트 개수)를 산출한다. 이때 이에 대한 로그값을 취할 수 있다.

[0047] 작성자 심사부(334b)는 댓글 작성자의 신뢰도를 심사한다. 댓글을 단 사용자가 믿을 만한 사용자라면 더 높은 점수를 주는 것이다. 예를 들어, 페이지랭크(PageRank) 기법을 사용하여 댓글 작성자의 신뢰도를 심사할 수 있다. 이는 월드 와이드 웹과 같은 하이퍼링크 구조를 가지는 문서에 상대적 중요도에 따라 가중치를 부여하는 방법이다. 즉, 어떤 문서를 참조한 다른 문서가 많을 수록 해당 문서의 가중치 점수가 올라간다. 본 발명의 일실시예는 이러한 링크 분석(link-analysis) 기반 방식을 사용자의 신뢰도를 심사하는 데 사용한다. 그러나 댓글 작성자의 신뢰도를 심사하는 방법이 여기에 제한되는 것은 아니며, 다른 방법을 사용할 수도 있다.

[0048] 댓글 품질 심사부(334c)는 댓글의 사용자 생성 컨텐츠와의 관련도를 심사한다. 예를 들어, 스팸(spam)성 댓글에는 점수를 낮게 책정한다. 이때 스팸 여부를 심사하는 방법에는 제한이 없다.

[0049] 세그먼트 선호극성 심사부(334d)는 세그먼트가 엔티티에 대해 긍정적인 감상을 나타내는지 부정적인 감상을 나타내는지를 심사한다. 예를 들어, 전술한 예에서 "강아지는 예쁜데, 고양이는 못생겼다"라는 댓글에서 "강아지는 예쁘다"는 것은 긍정적인 감상이고, "고양이는 못생겼다"라는 것은 부정적인 감상이 될 수 있다. 이러한 감상에는 얼마나 선호하느냐에 따라 점수를 책정할 수도 있고, 긍정적이냐 부정적이냐에 따라 부호를 줄 수도 있다. 이는 댓글을 단 사용자들이 해당 엔티티(예: 사용자 생성 컨텐츠)에 대해 어떤 생각을 가지고 있는지를 나타내는 척도로 쓰이게 된다.

[0050] 댓글 작성시기 심사부(334e)는 댓글이 언제 작성되었는지를 심사한다. 예를 들어, 댓글 작성시기 심사부(334e)는 최근에 달린 댓글에 예전에 달린 댓글보다 더 높은 점수를 줄 수 있고, 그 반대일 수도 있다.

[0051] 확장 집합 심사부(334f)는 댓글과 질의에 포함된 키워드의 동의어 확장 집합을 심사한다. 이때의 댓글 및 질의는 전처리 및 파싱을 거친 것이다. 예를 들어, 파싱된 질의에서 추출한 키워드가 "강아지"인 경우, 이에 대한 동의어에는 "개", "멍멍이" 등이 있을 것이다. 확장 집합 심사부(334f)는 "강아지"를 포함하는 세그먼트 뿐만이 아니라, 상기한 동의어들을 포함하는 세그먼트들도 심사한다. 따라서 강아지를 찍은 사진인데 댓글에는 "멍멍이 예쁘다"라는 식의 강아지가 아닌 다른 어휘를 사용한 감상이 적혀 있을 경우 본 발명에 따른 질의 처리부(330)는 해당 사진도 검색해낼 수 있다.

[0052] 세그먼트 심사부(334)는 각 구성 요소가 산출한 심사 점수를 통합한다. 일실시예에서 세그먼트 심사부(334)는 다음과 같은 수학식을 통합 심사에 사용할 수 있다.

[0053] <수학식1>

$$S_{SEG}(s,Q) = \sum_{t \in Q \cap s} \log\frac{n+1}{n_t} \times auth(w) \times (1+q(r))^{k_1}$$

$$\times sign(p(s)) \times (1+|p(s)|)^{k_2} \times f(r)$$

$$\times \begin{cases} k_3 & if\ t = a(s)\ and\ m(s) \cap x(m(Q)) \neq \emptyset \\ 1 & otherwise \end{cases}$$

[0054]

[0055] 또한 일실시예에서 엔티티 심사부(336)는 다음과 같은 수학식을 심사에 사용할 수 있다.

[0056] <수학식2>

$$S_{ENT}(e,Q) = \sum_{S_i \in S_{Q \cap e}} \frac{S_{SEG}(s_i,Q)^K}{|S_e|}$$

[0057]

[0058] 상기한 두 수학식에서 각 심벌은 다음과 같은 의미를 가지고 있다.

[0059] e, ENT: 엔티티(entity, 사용자 생성 컨텐츠(예: 사진), 사용자, 또는 그룹)

[0060] s, SEG: 세그먼트(segment, 각 댓글의 최소 의미 단위)

[0061] Se: 특정 엔티티에 대한 댓글의 세그먼트 집합

[0062] Si: Se 중 i번째 세그먼트

[0063] S: 점수(score)

[0064] r: 댓글(reply)

[0065] Q: 사용자 질의(query)

[0066] t: 쿼리 키워드(term)

[0067] n: 전체 세그먼트 수

[0068] nt: t를 포함하는 세그먼트수

[0069] w: 댓글을 단 사용자(writer)

[0070] p: 선호극성(polarity, 긍정적인가, 부정적인가에 대한 점수)

[0071] sign: 선호극성의 부호(예: 극성이 긍정적이면1,부정적이면 -1)

[0072] f: 신선도(freshness, 댓글 작성 시기)

[0073] auth: 신뢰도(authority)

[0074] m: 수정(modified, 전처리 및 파싱을 거침)

[0075] x: 확장(extented, 동의어확장)

[0076] a: 애스팩트(aspect, 세그먼트의 애스팩트)

[0077] k1, k2, k3, K: 상수.

[0078] 따라서, 상기한 첫번째 수학식을 통해 세그먼트 심사부(334)는 질의 및 세그먼트에 공통적으로 포함되어 있는 키워드에 대해(시그마 기호 아래 참조), 각 구성 요소가 산출한 점수를 곱한 후 이를 모두 합하여 세그먼트의 질의와의 적합성을 심사한다는 것을 알 수 있다.

[0079] 곱셈의 첫번째 항목은 세그먼트 개수 심사부(334a)가 산출한, 전체 세그먼트 수와 질의 키워드가 포함된 세그먼트 수의 비를 나타낸다.

[0080] 두번째 항목은 작성자 심사부(334b)가 산출한, 댓글을 단 사용자의 신뢰도 점수를 나타낸다.

[0081] 세번째 항목은 댓글 품질 심사부(334c)가 산출한, 댓글의 품질 점수를 나타낸다. 전술한 바와 같이, 이는 댓글의 사용자 생성 컨텐츠와의 연관성이 높을수록, 스팸성 댓글이 아닐수록 높게 책정된다. 세그먼트 개수 심사부(334a)가 로그를 취했듯, 이 값도 일실시예에서 수학식에 기술된 바와 같은 부가적인 처리를 할 수 있으며, 다른 항목들도 마찬가지이다.

[0082] 네번째 항목 및 다섯번째 항목은 세그먼트 선호극성 심사부(334d)가 산출한, 선호극성 부호 및 점수이다.

[0083] 여섯번째 항목은 댓글 작성시기 심사부(334e)가 산출한, 댓글 작성 시기에 따른 점수이다.

[0084] 일곱번째 항목은 확장 집합 심사부(334f)가 산출한, 질의에 포함된 키워드의 동의어 확장 집합을 심사한 점수이

다. 수학식을 통해, 전처리 및 파싱된 질의의 동의어 확장 집합과 전처리 및 파싱된 세그먼트 사이에 공통적인 키워드가 존재하지 않는지 여부 등을 심사하여 전술한 점수들에 곱할 상수를 결정함을 알 수 있다.

[0085] 또한, 상기한 두번째 수학식을 통해 엔티티 심사부(336)는 세그먼트 심사부(334)가 산출한 점수를 엔티티를 심사하는 데 사용함을 알 수 있다.

[0086] 시그마 기호 아래의 의미는 해당 엔티티에 대해 포스팅된 댓글의 세그먼트들 중 질의에 맞는 것들만 대상으로 삼는다는 뜻이다. 즉, Si는 세그먼트 중 특정 엔티티에 관련된 세그먼트이면서 질의에 있는 키워드를 포함하는 세그먼트를 의미한다.

[0087] 상기한 바와 같이, Se는 특정 사용자 생성 컨텐츠에 대해 포스팅된 댓글의 세그먼트이다. Se의 절대값은 그 개수이며, K는 0에서 1 사이의 상수이다. 수학식에 기술되어 있는 바와 같이, K를 사용하여 부가적인 처리를 한 Se를 세그먼트 심사 점수와 함께 사용하여 엔티티의 질의에 대한 적합성 점수를 산출한다.

[0088] 도 4는 본 발명의 일실시예에 따른 인덱스의 구조를 도시하고 있다.

[0089] 전술한 바와 같이, 본 발명의 일실시예에 따른 컨텐츠 관리 시스템(10)은 사용자 생성 컨텐츠에 달린 댓글을 최소 의미 단위로 분할하고, 각 분할된 세그먼트들에 대해 역 인덱스를 구축한다. 역 인덱스를 구축한다는 것은 문서에 해당 키워드가 어디 있는지로 인덱싱하는 것이 아니라, 해당 키워드가 어느 문서에 포함되어 있는지, 더 나아가 그 문서의 어디에 위치하는지로 인덱싱한다는 뜻이다.

[0090] 역 인덱스의 개념을 쉽게 설명하기 위해 일상 생활에서의 예를 들자면, 일반적으로 전문 도서의 맨 뒤쪽에는 독자가 중요한 키워드가 설명되어 있는 페이지를 쉽게 찾을 수 있도록 어떤 키워드가 어느 페이지에 있는지를 나타내는 인덱스 페이지가 있는데, 이 인덱스 페이지를 개념적으로 역 인덱스라고 할 수 있다.

[0091] 도면을 보면, 각 키워드ID(t)에 대해 댓글(r)이 나열되어 있고, 각 댓글ID(r)에 대해 세그먼트ID(s)가 나열되어 있다. 즉, 키워드1(t)은 댓글1(r), 댓글2(r),... 등을 포함하며, 댓글1(r)은 자신의 세그먼트1(s), 세그먼트2(s),... 등을, 댓글2(r)는 자신의 세그먼트1(s), 세그먼트2(s),... 등을 포함한다.

[0092] 이때의 각 댓글(r)이나 세그먼트(s)는 해당 댓글이나 세그먼트를 나타내는 ID이다. 전술한 전문 도서의 인덱스 페이지가 해당 키워드를 포함하는 페이지의 내용 자체가 아닌 해당 페이지 번호만을 기재하는 것과 같은 원리이다. 내용을 모두 기재하기 위해서는 필요한 공간이 너무 큰 것이다.

[0093] 예를 들어, 전체 댓글 문서에서 추출한 키워드가 "강아지"와 "고양이"일 때, "강아지"는 키워드1, "고양이"는 키워드2가 된다. "강아지"라는 키워드가 추출된 첫번째 댓글이 "강아지는 예쁜데 고양이는 못생겼다"이고, 이 댓글은 예컨대 ID가 335인 사진의 4번째 댓글이어서 댓글ID가 335_4라고 하자(이는 설명을 위한 예일 뿐이며, ID(아이디)를 지정하는 방식에는 제한이 없다). 그렇다면 키워드1(t)이 포함하는 댓글1(r)은 "강아지는 예쁜데 고양이는 못생겼다"라는 댓글 문장 자체가 아닌 해당 댓글의 ID인 335_4이다.

[0094] 즉, 본 발명의 일실시예에 따른 컨텐츠 관리 시스템(10)이 사용하는 역 인덱스는 [키워드, [< [댓글 ID1 , [세그먼트 ID1, 세그먼트 ID2, …]] > , < [댓글 ID2, [세그먼트 ID1, 세그먼트 ID2, …]] > …]]와 같은 형식으로 저장된다.

[0095] 도 5는 본 발명의 일실시예에 따른 컨텐츠 관리 방법의 흐름을 도시하고 있다.

[0096] 네트워크에 분산되어 있는 사용자 생성 컨텐츠를 수집하고 통합 관리한다(S510). 통합 관리에는 전술한 바와 같이, 사용자 생성 컨텐츠의 조회, 다운로드, 삭제 및 여러 소셜 네트워크 서비스 사이트로의 일괄 등록 등이 포함된다.

[0097] 상기 단계(S510)에서 통합 관리하는 사용자 생성 컨텐츠에 대한 인덱스를 구축하고 업데이트한다(S520). 전술한 바와 같이, 인덱스는 사용자 생성 컨텐츠 등 엔티티에 대한 댓글을 최소 의미 단위인 세그먼트로 분할하고, 각 세그먼트에 대해 구성한 역 인덱스(inverted index)를 의미한다.

[0098] 사용자 질의를 수신하면(S530), 역 인덱스에 기초하여 질의를 처리하여 질의에 대한 적합성 점수가 높은 순으로 사용자 생성 컨텐츠를 반환한다(S540). 예를 들어, 질의가 "예쁜 고양이가 나온 사진"일 때, 역 인덱스에 저장되어 있는 "고양이"를 포함하는 댓글 세그먼트가 질의 처리에 사용된다.

[0099] 이때, 전술한 바와 같이, 각 세그먼트별로 댓글을 단 사용자의 감상(sentiment)을 고려하여 질의를 처리하므로, "강아지는 예쁜데 고양이는 못생겼다"와 같은 댓글은 질의와의 적합성 점수가 낮게 책정된다. 따라서 질의에 맞

는 사용자 생성 컨텐츠를 좀더 정확하게 검색 및 추천할 수 있다.

[0100] 또한 해당 댓글이 달린 사용자 생성 컨텐츠에 대한 다른 댓글에서는 "고양이가 귀엽다", "고양이가 사랑스럽다" 등 고양이에 대한 긍정적인 감상이 포함되어 있다면, 해당 사용자 생성 컨텐츠는 질의에 적합한 컨텐츠로 판단되어 반환된다. 즉, 본 발명의 일실시예에 따른 컨텐츠 관리 방법은 각 엔티티에 대해 달린 댓글의 세그먼트를 심사하고 이를 사용하여 엔티티에 대해 질의와의 적합성을 심사하므로, 사용자 생성 컨텐츠를 좀더 정확하게 검색 및 추천할 수 있다.

[0101] 도 6은 본 발명의 일실시예에 따른 인덱스 구축 단계(S520)의 흐름을 도시하고 있다.

[0102] 댓글을 전처리 및 파싱한다(S610). 이 과정에서 댓글에 포함되어 있는 불용어 및 어미가 제거된다.

[0103] 다음, 댓글을 세그먼트로 분할한다(S620). 세그먼트는 동일한 의미적 간계를 갖는 최소의 구, 절, 문장이다.

[0104] 다음, 각 분할된 세그먼트를 인덱싱한다(S630). 이에 따라 도 4와 같은 구조를 갖는 역 인덱스가 구축된다.

[0105] 도 7은 본 발명의 일실시예에 따른 질의 처리 단계(S540)의 흐름을 도시하고 있다.

[0106] 질의를 전처리 및 파싱한다(S710). 이 과정에서 질의에 포함되어 있는 불용어 및 어미가 제거된다.

[0107] 세그먼트에 대해 질의와의 적합성을 심사한다(S720).

[0108] 상기 단계(S720)의 결과를 기초로, 엔티티에 대해 질의와의 적합성을 심사한다(S730). 실제로는 엔티티 심사 단계(S730)에서 각 엔티티에 대해 포스팅된 댓글의 세그먼트에 대해 상기 세그먼트 심사 단계(S720)를 호출하는 식으로 구현될 수 있다. 즉, 각 엔티티에 대해 포스팅된 댓글의 세그먼트에 대한 적합성 점수를 기초로 엔티티의 적합성 순위를 산출하고, 이 순위가 높은 순으로 해당 사용자 생성 컨텐츠를 반환할 수 있다.

[0109] 도 8은 본 발명의 일실시예에 따른 세그먼트 심사 단계(S720)의 흐름을 도시하고 있다.

[0110] 세그먼트 개수를 심사한다(S810). 즉, 질의에 포함된 키워드를 포함하는 세그먼트 개수를 전체 세그먼트 개수와 비교한다.

[0111] 댓글 작성자의 신뢰도를 심사한다(S820). 즉, 댓글 작성자의 신뢰도를 심사하여 믿을 만한 사용자의 댓글이면 점수를 높게 책정한다.

[0112] 댓글 품질을 심사한다(S830). 즉, 댓글의 사용자 생성 컨텐츠와의 관련도를 심사하여, 스팸성 댓글이면 점수를 낮게 책정한다.

[0113] 세그먼트 선호극성을 심사한다(S840). 즉, 세그먼트가 엔티티에 대해 긍정적인 감상을 나타내는지 부정적인 감상을 나타내는지 심사한다.

[0114] 댓글 작성시기를 심사한다(S850). 즉, 댓글이 언제 작성되었는지를 심사한다.

[0115] 확장 집합을 심사한다(S860). 즉, 댓글과 질의에 포함된 키워드의 동의어 확장 집합을 이용하여 세그먼트를 심사한다.

[0116] 상기 단계들(S810에서 S860)에서 산출한 심사 점수들을 통합하여 세그먼트의 적합성 순위를 계산한다(S870). 이때 상기 단계들은 순서대로 진행되는 것으로 도시되었지만, 실제로는 임의의 순서로 진행될 수 있다.

[0117] 전술한 본 발명의 설명은 예시를 위한 것이며, 본 발명이 속하는 기술분야의 통상의 지식을 가진 자는 본 발명의 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.

[0118] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

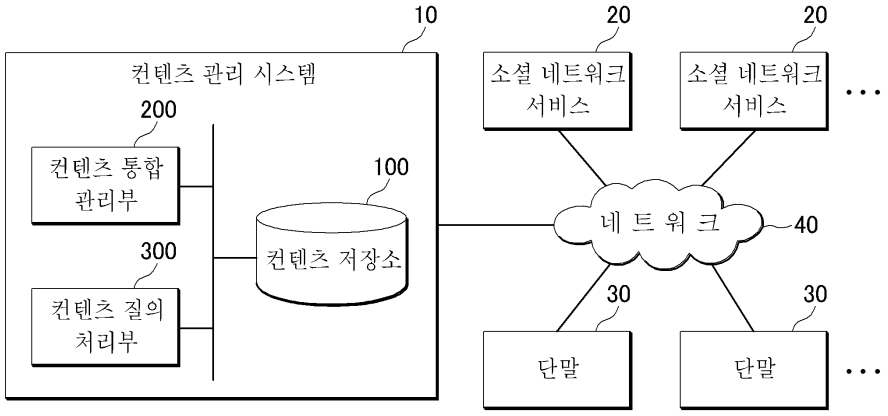
[0119] 10: 컨텐츠 관리 시스템

200: 컨텐츠 통합 관리부

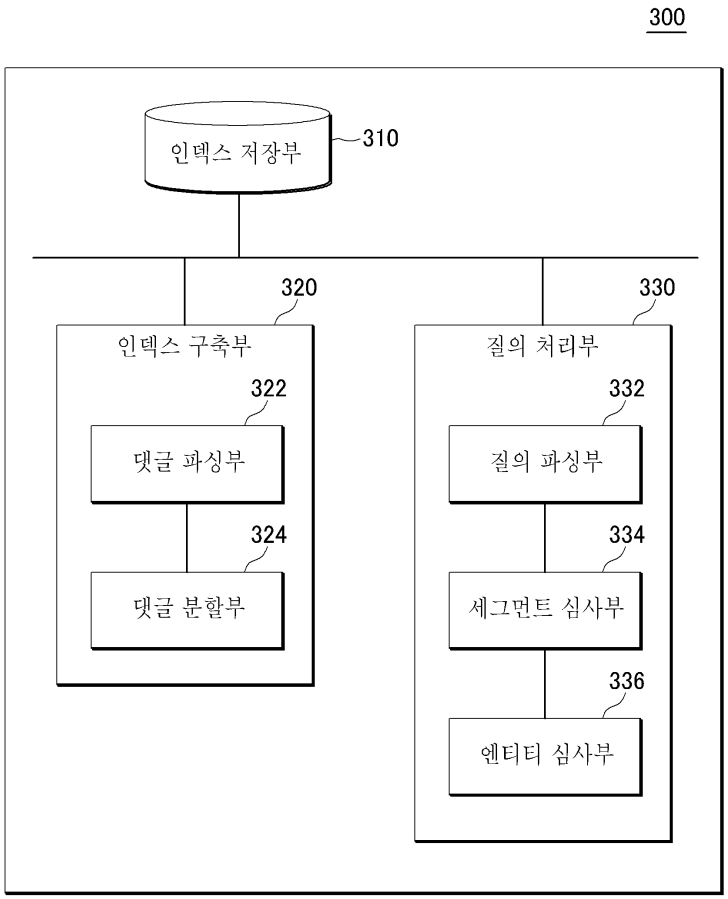
300: 컨텐츠 질의 처리부

도면

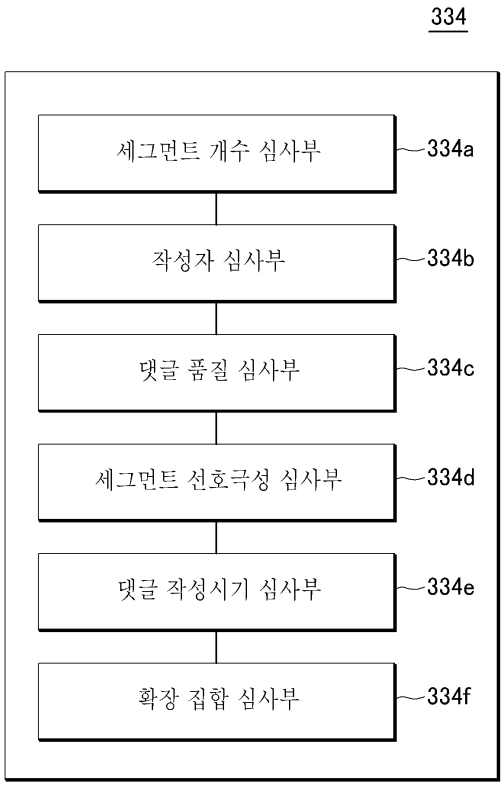
도면1



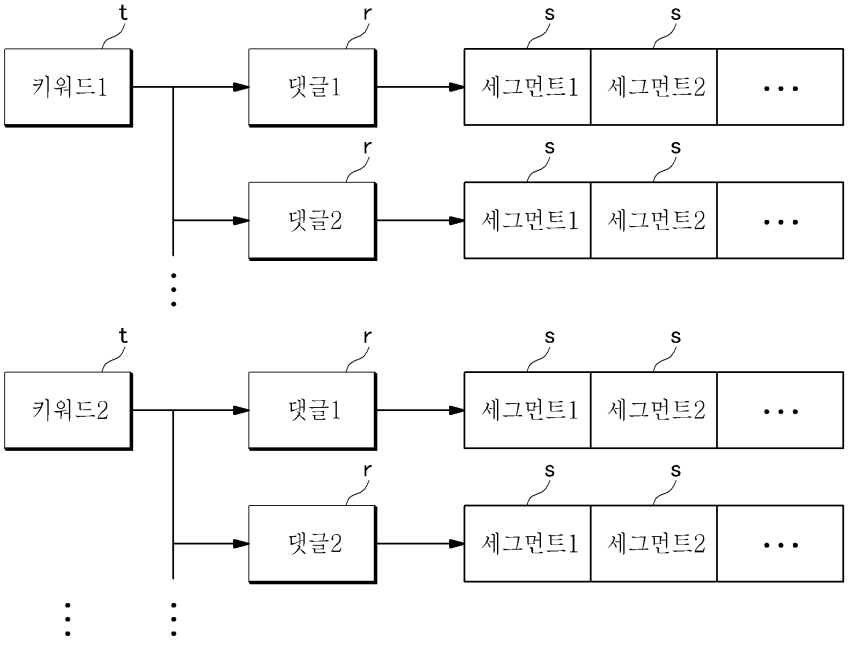
도면2



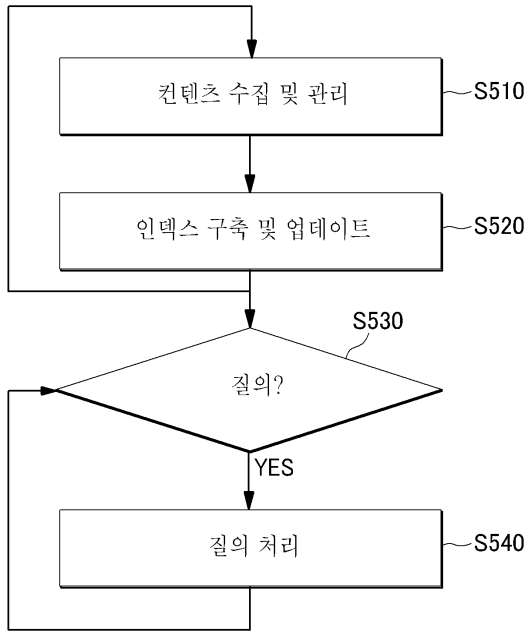
도면3


334
세그먼트 개수 심사부
334a
작성자 심사부
334b
댓글 품질 심사부
334c
세그먼트 선호극성 심사부
334d
댓글 작성시기 심사부
334e
확장 집합 심사부
334f

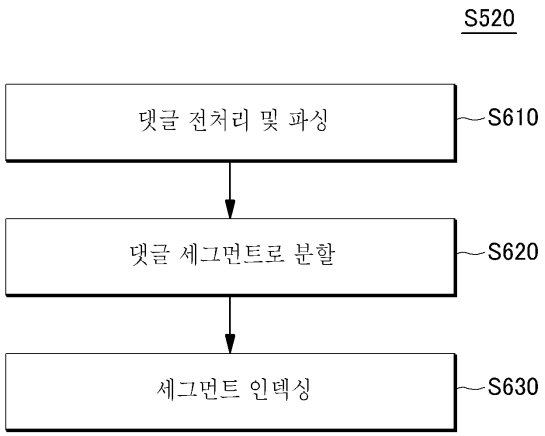
도면4


t
키워드1
r
댓글1
s
세그먼트1
세그먼트2
r
댓글2
키워드2

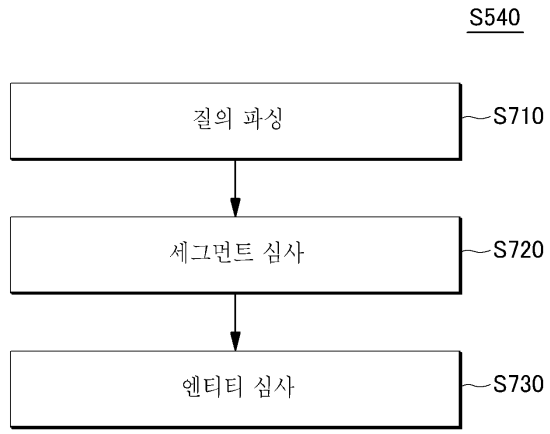
도면5


컨텐츠 수집 및 관리
S510
인덱스 구축 및 업데이트
S520
S530
질의?
YES
질의 처리
S540

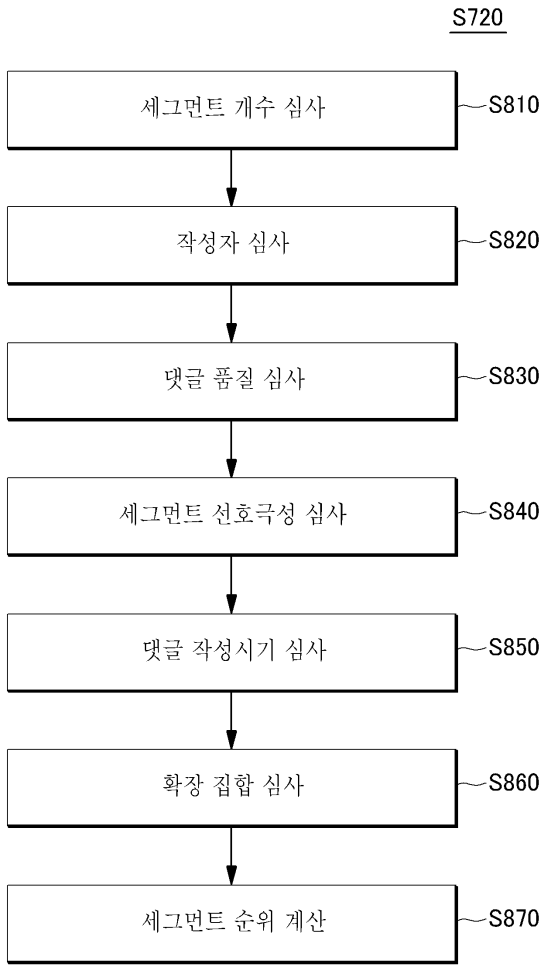
도면6


S520
댓글 전처리 및 파싱
S610
댓글 세그먼트로 분할
S620
세그먼트 인덱싱
S630

도면7


S540
질의 파싱
S710
세그먼트 심사
S720
엔티티 심사
S730

도면8


S720
세그먼트 개수 심사
S810
작성자 심사
S820
댓글 품질 심사
S830
세그먼트 선호극성 심사
S840
댓글 작성시기 심사
S850
확장 집합 심사
S860
세그먼트 순위 계산
S870