(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2018-0024837
(43) 공개일자 2018년03월08일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
(52) CPC특허분류
G06F 17/30616 (2013.01)
G06F 17/30038 (2013.01)
(21) 출원번호 10-2016-0111758
(22) 출원일자 2016년08월31일
심사청구일자 없음

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
이홍철
서울특별시 서초구 동광로19길 51, 701호(방배동, 래미안방배에버뉴)
김기현
서울특별시 중랑구 동일로 932, 201동2604호(묵동, 묵동자이아파트)
(뒷면에 계속)
(74) 대리인
특허법인충정

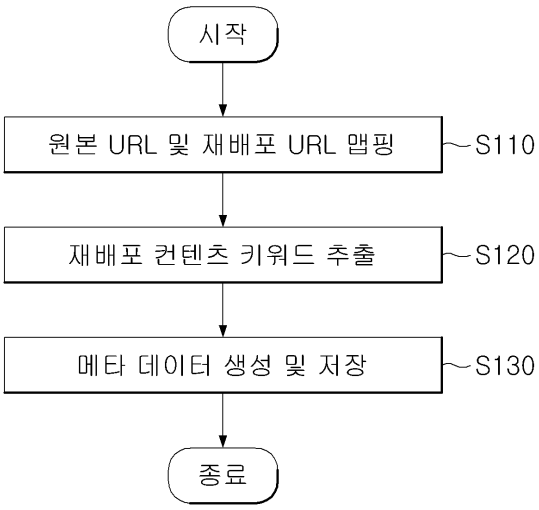
전체 청구항 수 : 총 1 항

(54) 발명의 명칭 **백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법, 그 방법을 수행하는 장치 및 컴퓨터 프로그램**

(57) 요 약

본 발명은 온라인 콘텐츠의 키워드를 자동으로 추출하는 방법으로서, 원본 콘텐츠 URL 정보에 상기 원본 콘텐츠 URL 정보가 재배포된 재배포 URL 정보를 맵핑하는 단계; 재배포된 콘텐츠에 추가된 코멘트로부터 키워드를 추출하는 단계; 및 추출된 키워드를 이용하여 상기 콘텐츠에 대한 메타데이터를 생성하고 저장하는 단계를 포함하는 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법을 개시한다.

대 표 도 - 도1



(52) CPC특허분류
G06F 17/30887 (2013.01)
(72) 발명자
이민영
서울특별시 성북구 안암로9가길 87-6, 605호 (안암동5가)
최영진
경상북도 구미시 백산로 153, 102동 601호(송정동, 동양한신아파트)

명 세 서

청구범위

청구항 1

백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법에 있어서,

원본 콘텐츠 URL 정보에 상기 원본 콘텐츠 URL 정보가 재배포된 재배포 URL 정보를 맵핑하는 단계;

재배포된 콘텐츠에 추가된 코멘트로부터 키워드를 추출하는 단계; 및

추출된 키워드를 이용하여 상기 콘텐츠에 대한 메타데이터를 생성하고 저장하는 단계를 포함하는 것을 특징으로 하는 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 온라인 콘텐츠의 키워드를 자동으로 추출하는 방법에 관한 것으로서, 보다 구체적으로는 온라인 콘텐츠가 하이퍼링크를 통해 재배포된 경우 재배포된 온라인 콘텐츠를 역으로 접근하여 재배포된 온라인 콘텐츠의 키워드를 자동으로 추출하는 방법, 그 방법을 수행하는 장치 및 컴퓨터 프로그램에 관한 것이다.

[0002]

배 경 기 술

[0003] 최근 인터넷상에서 텍스트뿐만 아니라 이미지, 동영상 등의 콘텐츠를 제공하는 업체들이 많아지면서 그와 더불어 온라인 콘텐츠 검색 서비스에 대한 필요성도 높아지고 있다.

[0004] 일반적으로, 콘텐츠 검색 방식은 해당 콘텐츠의 내용을 대표하는 키워드를 추출하여 색인하고, 이렇게 색인된 키워드를 이용하여 콘텐츠를 검색한다. 그에 따라 콘텐츠의 내용을 대표하는 적절한 키워드를 추출하는 과정은 매우 중요하다. 특히, 텍스트의 검색 시간보다 이미지, 동영상의 검색 시간이 더 오래 걸린다는 점에서 정확한 키워드를 추출하여 메타데이터로 설정하는 것은 매우 중요하다.

[0005] 일반적으로, 동영상의 키워드를 추출하기 위해서는 주로 메타데이터가 많이 사용되고 있으며, 그 외에도 시납시스, 리뷰, 웹에서 크롤링한 데이터 등도 사용되고 있다. 기존의 검색 서비스는 사전과 형태소 분석을 통해서 명사를 추출하고, 추출된 명사의 문헌 빈도와 역문헌 빈도를 이용하여 가중치를 부여하여 색인을 수행하고, 이러한 색인을 이용하여 검색을 수행한다.

[0006] 그러나, 원본 콘텐츠에 최초로 설정된 키워드, 메타데이터는 원본 제작자에 의해 설정되게 되므로, 콘텐츠의 재배포에 따라 동적으로 변하는 콘텐츠에 대한 키워드 및 메타데이터를 제대로 반영하지 못하는 문제점이 있었다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 상기와 같은 종래 기술의 문제점을 해결하기 위해 창안된 것으로, 원본 콘텐츠가 소비자에 의해 재배포됨에 따라 새롭게 생성되는 키워드를 반영하여 지속적인 변화에 동적으로 대응하도록 하는 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법, 그 방법을 수행하는 장치 및 컴퓨터 프로그램을 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0010] 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법은,

[0011] 원본 콘텐츠 URL 정보에 상기 원본 콘텐츠 URL 정보가 재배포된 재배포 URL 정보를 맵핑하는 단계; 재배포된 콘텐츠에 추가된 코멘트로부터 키워드를 추출하는 단계; 및 추출된 키워드를 이용하여 상기 콘텐츠에 대한 메타데이터를 생성하고 저장하는 단계를 포함할 수 있다.

발명의 효과

[0013] 본 발명의 실시예에 따르면, 재배포된 콘텐츠의 키워드를 추출하여 원본 콘텐츠의 주요 키워드를 콘텐츠 소비자에게 소비되는 형태의 키워드로 재정립하는 과정을 통해 온라인 콘텐츠가 지속적으로 소비되는 환경을 제공할 수 있다.

[0014] 또한, 콘텐츠 제작자가 별도로 메타데이터를 설정하지 않더라도 자동으로 키워드 관련 메타데이터를 설정하 수 있으며, 재배포되는 과정에서 새롭게 생성되는 키워드 변화에 동적으로 대응할 수 있다.

도면의 간단한 설명

[0016] 본 발명에 관한 이해를 돕기 위해 상세한 설명의 일부로 포함되는, 첨부도면은 본 발명에 대한 실시예를 제공하고, 상세한 설명과 함께 본 발명의 기술적 사상을 설명한다.

도 1은 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법의 순서를 도시한 것이다.

도 2는 본 발명에 따른 재배포 콘텐츠 키워드 자동 추출 방법의 순서를 도시한 것이다.

도 3은 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법의 일례를 도시한 것이다.

도 4는 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 장치를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0017] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 이하에서는 특정 실시예들을 첨부된 도면을 기초로 상세히 설명하고자 한다.

[0018] 이하의 실시예는 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

[0019] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시 예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.

[0020] 또한, 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되는 것은 아니며, 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[0022] 도 1은 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법의 순서를 도시한 것이다.

[0023] 도 1을 참조하면, 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법은 원본 URL 및 재배포 URL 맵핑 단계(S110), 재배포 콘텐츠 키워 추출 단계(S120) 및 메타데이터 생성 및 저장 단계(S130)를 포함하여 구성될 수 있다.

[0024] 우선, S110 단계에서 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 장치(100)는 원본 콘텐츠 URL에 재배포 콘텐츠 URL을 맵핑하여 하나의 연관된 그룹을 형성하도록 할 수 있다.

[0025] S120 단계에서는 재배포된 콘텐츠에 대한 키워드를 추출할 수 있다. 콘텐츠 재배포자들은 특정 콘텐츠를 재배포하면서 해당 콘텐츠에 대하여 자기 나름의 코멘트를 입력하여 게시하게 되고, 일반적으로 이 코멘트는 해당 콘텐츠와 상당한 관련성이 있다고 볼 수 있다. 그러므로 해당 코멘트로부터 키워드를 추출할 수 있다. 또한, 재배포된 콘텐츠뿐만 아니라 재재배포된 콘텐츠에 입력된 코멘트로부터도 키워드를 추출할 수도 있다.

[0026] 이하, 도 2를 참조하여 재배포 콘텐츠의 키워드를 자동으로 추출하는 방법을 보다 상세히 설명한다.

[0027] 도 2를 참조하면, 본 발명에 따른 재배포 콘텐츠 키워드 추출 방법은 데이터를 수집하고 전처리 하는 단계(S121), 형태소 분석 및 해시태그를 추출하는 단계(S122), 키워드를 선정하는 단계(S123) 및 메타 태그로 변환하는 단계(S124)를 포함하여 구성될 수 있다.

[0029] 도 3은 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 방법의 일례를 도시한 것이다.

[0030] 온라인 콘텐츠 제작자 A가 "라면을 맛있게 끓이는 법"이라는 제목으로 동영상을 제작하여 게시하였고, 재배포자 B는 A가 업로드한 동영상을 보고 "이렇게 라면을 끓이면 해장하는데 너무 좋겠네"라는 코멘트를 더하여 A의 원본 동영상 URL을 링크하였다. 이 후, 재배포자 C가 다시 "해장라면"이라는 코멘트를 더하여 원본 동영상 URL을 링크하였고, 재배포자 D는 영어로 "I will teach you how to cook a delicious ramen"이라는 코멘트를 더하여 원본 동영상 URL을 링크하였다.

[0031] 이 때, 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 장치(100)는 재배포자(B, C, D)의 재배포 URL을 원본 URL에 맵핑하여 별도의 저장하고, 재배포 콘텐츠의 키워드를 추출한다. 구체적으로, 재배포자 B의 코멘트로부터 "라면", "해장"의 키워드를 추출할 수 있고, 재배포자 C의 코멘트로부터는 "해장라면"을 추출할 있고, 재배포자 D의 코멘트로부터는 "delicious ramen"이 추출될 수 있다. 이 후, 추출된 키워드를 이용하여 메타데이터를 생성하고 상기 별도의 저장장치에 저장할 수 있다.

[0032] 최초에 온라인 콘텐츠 제작자는 "해장" 또는 "해장라면" 등의 키워드를 설정하지 않았으나 재배포되는 과정에서 백링크를 이용하여 키워드가 추가로 설정될 수 있다. 이로써, 온라인 콘텐츠가 재배포되면서 발생하는 키워드의 추가 또는 변화에 동적으로 대응할 수 있다.

[0034] 도 4는 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 장치를 도시한 것이다.

[0035] 도 4를 참조하면, 본 발명에 따른 백링크를 이용한 온라인 콘텐츠 키워드 자동 추출 장치(100)는 URL 맵핑부(110), 키워드 추출부(120) 및 메타데이터 생성부(130)를 포함하여 구성될 수 있다. 나아가, 생성된 메타데이터를 저장하는 저장장치(DB)(140)를 더 포함할 수도 있다.

[0036] 상기 URL 맵핑부(110)는 원본 콘텐츠 URL에 재배포 콘텐츠 URL을 맵핑하여 하나의 연관된 그룹을 형성하도록 할 수 있다.

[0037] 상기 키워드 추출부(120)는 재배포자에 의하여 재배포된 콘텐츠에 추가로 입력된 코멘트로부터 키워드를 추출할 수 있다.

[0038] 상기 메타데이터 생성부(130)는 추출된 키워드를 이용하여 메타데이터를 생성하고 상기 저장장치(140)에 해당 메타데이터를 저장할 수 있다.

[0039] 상기 저장장치(DB)(140)는 맵핑된 재배포 콘텐츠 URL 정보, 생성된 메타데이터 정보 등을 저장할 수 있다.

[0041] 한편, 상술한 본 발명의 실시예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성가능하고, 작성된 프로그램은 매체에 저장될 수 있다.

[0042] 상기 매체는 마그네틱 저장매체(예를 들면, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장매체를 포함할 수 있으나, 이에 한정되는 것은 아니다.

[0043] 이상의 설명은 본 발명의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서 본 발명에 기재된 실시예는 본 발명의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 한정되는 것은 아니다. 본 발명의 보호 범위는 아래의 청구범위에 의해서 해석되어야 하며, 그와 균등한 범위 내에 있는 모든 기술 사상은 본 발명의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0045] 100: 키워드 자동 추출 장치

110: URL 맵핑부

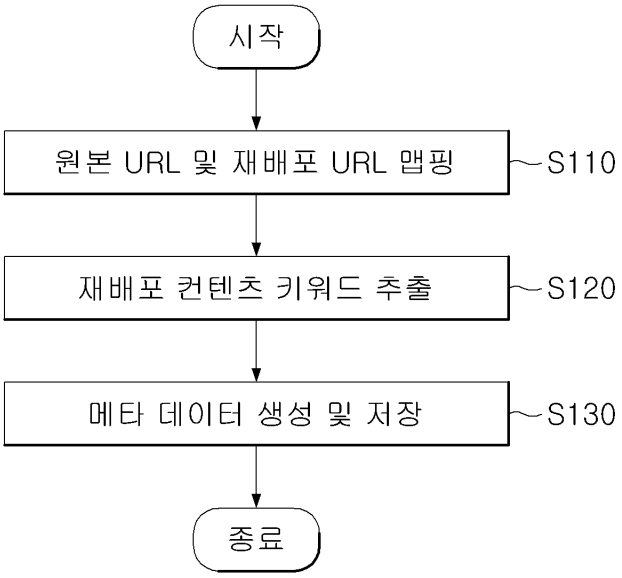
120: 키워드 추출부

130: 메타데이터 생성부

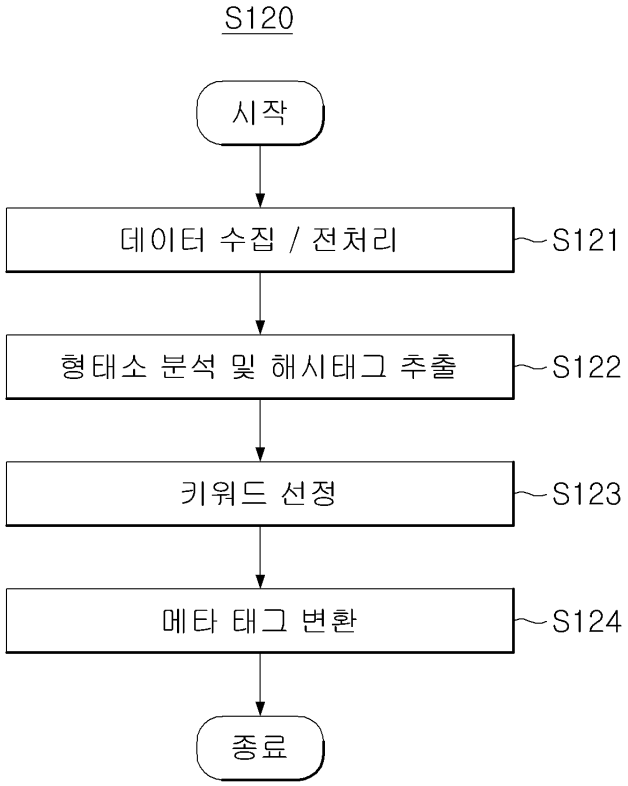
140: 저장장치

도면

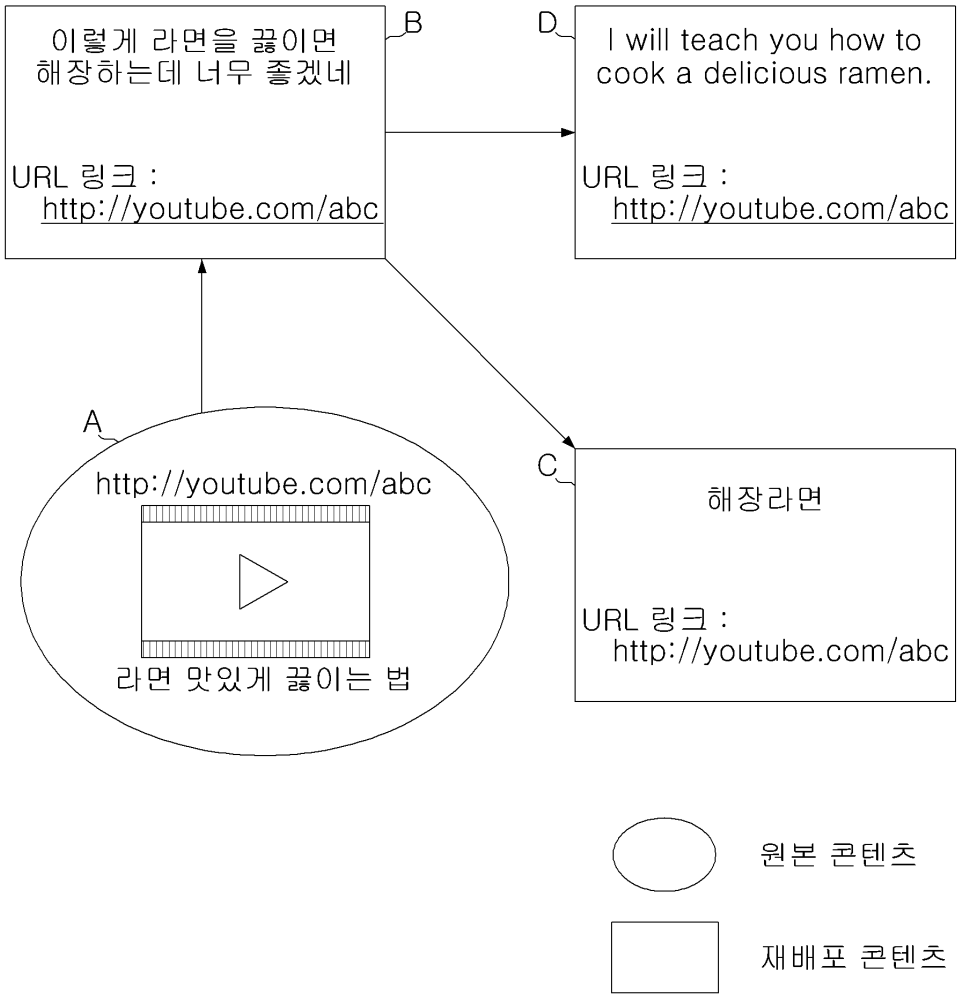
도면1



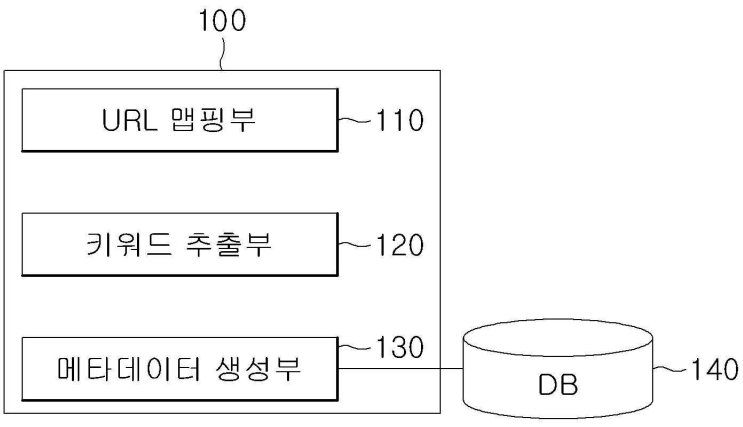
도면2


S120
시작
데이터 수집 / 전처리
S121
형태소 분석 및 해시태그 추출
S122
키워드 선정
S123
메타 태그 변환
S124
종료

도면3


이렇게 라면을 끓이면
해장하는데 너무 좋겠네
URL 링크 :
http://youtube.com/abc
B
D
I will teach you how to
cook a delicious ramen.
URL 링크 :
http://youtube.com/abc
A
http://youtube.com/abc
라면 맛있게 끓이는 법
C
해장라면
URL 링크 :
http://youtube.com/abc
원본 콘텐츠
재배포 콘텐츠

도면4


100
URL 맵핑부
110
키워드 추출부
120
메타데이터 생성부
130
DB
140