

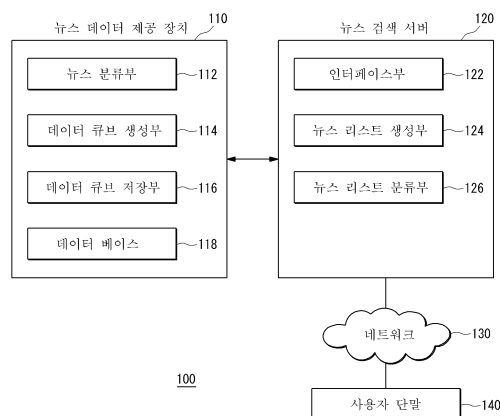
	(19) 대한민국특허청(KR) (12) 공개특허공보(A)	(11) 공개번호 10-2011-0012815 (43) 공개일자 2011년02월09일
(51) Int. Cl. H04N 5/445 (2011.01) H04N 7/08 (2006.01) (21) 출원번호 10-2009-0070690 (22) 출원일자 2009년07월31일 심사청구일자 2009년07월31일	(71) 출원인 고려대학교 산학협력단 서울 성북구 안암동5가 1 (72) 발명자 박대희 서울특별시 강남구 도곡동 타워팰리스 A동 1601호 이한성 경기도 용인시 기흥구 동백동 동백 택지개발지구 C2-1블럭 백현마을 2607동 2501호 (뒷면에 계속) (74) 대리인 박준석	
전체 청구항 수 : 총 9 항		

(54) 뉴스 제공 방법 및 시스템

(57) 요약

뉴스 비디오를 뉴스 기사 단위로 분할하고, 분할된 뉴스 비디오들을 카테고리 축, 시간 축 및 방송사 축을 기준으로 분류하여 데이터 큐브를 생성하는 뉴스 데이터 제공 장치 및 상기 뉴스 데이터 제공 장치로부터 검색 조건에 해당하는 뉴스 비디오를 검색하여 사용자 단말에 제공하는 뉴스 검색 서버를 포함하되, 상기 뉴스 검색 서버는 상기 검색 조건에 해당하는 뉴스 비디오를 포함하는 뉴스 리스트를 생성하고, 상기 생성된 뉴스 리스트에 포함된 뉴스 비디오들간의 유사도를 나타내는 유사도 행렬을 산출하는 뉴스 리스트 생성부 및 상기 뉴스 비디오의 워드 데이터와 영상 프레임에 기초하여 산출된 뉴스 비디오들 간의 유사도 행렬에 따라 상기 뉴스 리스트에 포함된 상기 뉴스 비디오들을 신규 뉴스 또는 진화 뉴스로 분류하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 뉴스 리스트 분류부를 포함하는 뉴스 제공 시스템을 제공한다.

대표도 - 도1



(72) 발명자

임영희

경기도 수원시 영통구 매탄3동 우남퍼스트빌아파트
208동 702호

유재학

충청북도 옥천군 옥천읍 죽향리 옥향아파트 103동
803호

오승근

인천광역시 부평구 삼산동 삼산아파트 3동 501호

정민중

서울특별시 관악구 봉천9동 관악벽산블루밍아파트
106동 1306호

박승진

전라남도 목포시 옥암동 제일2차 하이빌마트

강봉수

충청남도 연기군 조치원읍 서창리 호연오피스텔
202호

특허청구의 범위

청구항 1

뉴스 제공 시스템에 있어서,

뉴스 비디오를 뉴스 기사 단위로 분할하고, 분할된 뉴스 비디오들을 카테고리 축, 시간 축 및 방송사 축을 기준으로 분류하여 데이터 큐브를 생성하는 뉴스 데이터 제공 장치 및

상기 뉴스 데이터 제공 장치로부터 검색 조건에 해당하는 뉴스 비디오를 검색하여 사용자 단말에 제공하는 뉴스 검색 서버를 포함하되,

상기 뉴스 검색 서버는 상기 검색 조건에 해당하는 뉴스 비디오를 포함하는 뉴스 리스트를 생성하고, 상기 생성된 뉴스 리스트에 포함된 뉴스 비디오들간의 유사도를 나타내는 유사도 행렬을 산출하는 뉴스 리스트 생성부 및

상기 뉴스 비디오의 워드 데이터와 영상 프레임에 기초하여 산출된 뉴스 비디오들 간의 유사도 행렬에 따라 상기 뉴스 리스트에 포함된 상기 뉴스 비디오들을 신규 뉴스 또는 진화 뉴스로 분류하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 뉴스 리스트 분류부

를 포함하는 뉴스 제공 시스템.

청구항 2

제 1 항에 있어서,

상기 검색 조건은 키워드, 뉴스 비디오의 카테고리, 뉴스 비디오가 방송된 시간 또는 뉴스 비디오가 방송된 방송사 정보를 포함하는 것인 뉴스 제공 시스템.

청구항 3

제 1 항에 있어서,

상기 뉴스 검색 서버는 사용자 단말로부터 상기 검색 조건을 수신하고, 상기 생성된 신규 뉴스 리스트 및 진화 뉴스 리스트를 상기 사용자 단말로 전송하는 인터페이스부를 더 포함하는 뉴스 제공 시스템.

청구항 4

제 1 항에 있어서,

상기 뉴스 리스트 분류부는 사용자 단말로부터 뉴스 리스트 분류 요청을 수신한 경우에 한하여 상기 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 뉴스 제공 시스템.

청구항 5

제 1 항에 있어서,

상기 뉴스 리스트 분류부는 제 1 뉴스 비디오와 제 2 뉴스 비디오간의 유사도가 임계값 이상인 경우에는 상기 제 2 뉴스 비디오를 상기 제 1 뉴스 비디오의 진화 뉴스로 분류하고,

상기 제 1 뉴스 비디오와 상기 제 2 뉴스 비디오간의 유사도가 임계값 이하인 경우에는 상기 제 2 뉴스 비디오를 신규 뉴스로 분류하는 뉴스 제공 시스템.

청구항 6

제 5 항에 있어서,

상기 뉴스 리스트 분류부는 상기 제 1 뉴스 비디오를 제 1 신규 뉴스 리스트에 포함시키고,

상기 제 2 뉴스 비디오가 상기 제 1 뉴스 비디오의 진화 뉴스인 경우, 상기 제 1 신규 뉴스 리스트와 연계된 제 1 진화 뉴스 리스트에 상기 제 2 뉴스 비디오를 포함시키고,

상기 제 2 뉴스 비디오가 신규 뉴스인 경우, 상기 제 2 뉴스 비디오를 제 2 신규 뉴스 리스트에 포함시키는 뉴스 제공 시스템.

청구항 7

뉴스 제공 방법에 있어서,

뉴스 데이터베이스로부터 검색 조건에 해당하는 뉴스 비디오를 검색하여 뉴스 리스트를 생성하는 단계,

사용자 단말로부터 상기 뉴스 리스트의 분류 요청을 수신한 경우, 상기 뉴스 리스트에 포함된 각 뉴스 비디오의 워드 데이터와 영상 프레임에 기초하여 뉴스 비디오들 간의 유사도를 나타내는 유사도 행렬을 산출하는 단계,

상기 산출된 유사도 행렬에 따라 상기 뉴스 리스트에 포함된 상기 뉴스 비디오들을 신규 뉴스 또는 진화 뉴스로 분류하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 단계,

상기 생성된 신규 뉴스 리스트 및 진화 뉴스 리스트를 상기 사용자 단말에 전송하는 단계를 포함하되,

상기 뉴스 데이터베이스는 뉴스 비디오를 뉴스 기사 단위로 분할하고, 상기 분할된 뉴스 비디오들을 카테고리 별, 시간대 별 및 방송사 별로 분류하여 저장하는 것인 뉴스 제공 방법.

청구항 8

제 7 항에 있어서,

상기 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 단계는

제 1 뉴스 비디오와 제 2 뉴스 비디오간의 유사도가 임계값 이상인 경우에는 상기 제 2 뉴스 비디오를 상기 제 1 뉴스 비디오의 진화 뉴스로 분류하는 단계 및

상기 제 1 뉴스 비디오와 상기 제 2 뉴스 비디오간의 유사도가 임계값 이하인 경우에는 상기 제 2 뉴스 비디오를 신규 뉴스로 분류하는 단계

를 포함하는 뉴스 제공 방법.

청구항 9

제 7 항에 있어서,

상기 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 단계는

상기 제 1 뉴스 비디오를 제 1 신규 뉴스 리스트에 포함시키는 단계,

상기 제 1 뉴스 비디오와 제 2 뉴스 비디오간의 유사도가 임계값 이상인 경우 상기 제 2 뉴스 비디오를 상기 제 1 신규 뉴스 리스트와 연계된 제 1 진화 뉴스 리스트에 포함시키는 단계 및

상기 제 1 뉴스 비디오와 제 2 뉴스 비디오간의 유사도가 임계값 이하인 경우 상기 제 2 뉴스 비디오를 상기 제 2 신규 뉴스 리스트에 포함시키는 단계

를 포함하는 뉴스 제공 방법.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 뉴스 제공 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 최근 들어 멀티미디어 및 컴퓨터 통신 기술이 발달함에 따라 비디오 정보가 점차 중요한 정보 자원으로 여겨지고 있으며, 그 활용 역시 폭발적으로 증가하고 있다. 특히 뉴스 비디오는 사회, 문화, 정치, 경제 등의 다양하고 풍부한 정보를 내포하고 있으므로, 이를 효과적으로 관리하고 검색하기 위한 뉴스 비디오 데이터베이스의 필요성이 대두되고 있다.

[0003] 이러한 필요성과 관련하여, 최근의 연구 문헌 조사에 따르면, 대부분의 연구들이 뉴스 비디오의 내부 구조 분석을 통한 뉴스 비디오 파싱(parsing), 색인, 브라우징 및 검색에 집중되고 있다. 그러나 해당 연구들은 단순한 검색어 기반의 뉴스 비디오 검색 및 뉴스 비디오 브라우징에 한정되어 있으므로, 뉴스 비디오의 주제별 분석 및 데이터마이닝을 지원하는 것은 매우 어렵다.

[0004] 뉴스 비디오의 내부 구조 분석 연구의 문제점들을 극복하기 위한 대안으로써, 토픽 쓰레딩(Topic-threading)과 오토 다큐멘팅(Auto-documenting)과 같은 뉴스 비디오의 외부 구조 분석을 시도하려는 일련의 연구들이 진행 중이다. 토픽 쓰레딩은 뉴스 비디오를 주제별로 분류함으로써 사용자로 하여금 관련 있는 뉴스 비디오들을 빠르게 브라우징 할 수 있도록 지원하는 시스템이며, 오토 다큐멘팅은 특정 주제의 뉴스 비디오들을 각 뉴스 비디오들의 의미를 고려하여 요약/정리하는 시스템으로써, 사용자에게 특정 주제에 대한 뉴스 비디오들의 간략한 요약을 제공하기 위한 시스템이다. 해당 연구들은 뉴스 비디오에 대한 내부 분석뿐만 아니라 뉴스 비디오의 기사 간 상관관계 분석을 통하여 부분적인 뉴스 비디오의 분석 및 요약을 제공하고 있으나, 여전히 브라우징과 검색에 초점을 맞추고 있어 뉴스 비디오의 주제별 의미 분석과 데이터마이닝과는 거리가 멀다.

[0005] 한편, 본 출원의 발명자가 연구 논문[이한성, “A Data Cube System for the Semantic Analysis of News Video”, 박사학위 논문, 고려대학교 전산학과, 2007]을 통해 제시한 뉴스 비디오 관리 시스템은 데이터 큐브를 기반으로 하고 있다. 이때, 데이터 큐브를 구축하는 과정은 1) 뉴스 비디오를 샷 단위로 나누기 위한 샷 경계 탐지(shot boundary detection) 단계, 2) 뉴스 앵커를 포함하고 있는 샷 들을 탐지하기 위한 앵커 샷 탐지(anchor shot detection) 단계, 3) 앵커 샷 들을 기준으로 뉴스 비디오를 뉴스 기사 단위로 나누는 뉴스 비디오 경계 탐지(story boundary detection) 단계, 4) 분할된 뉴스 비디오들을 주제별로 분류하는 뉴스 비디오 군집화 단계를 포함한다.

[0006] 한편, 이와 같이 생성된 데이터 큐브에서 특정 조건에 해당하는 뉴스 비디오를 검색하는 경우, 뉴스 비디오간의 유사성 또는 의존도를 고려하여 트리 구조 형태로 검색 결과를 생성하는 과정을 거쳤으나, 알고리즘 및 자료 구조의 복잡성으로 인하여 처리 속도가 느려지는 문제점이 있었다.

발명의 내용

해결 하고자하는 과제

[0007] 본 발명의 일부 실시예는 뉴스 비디오들 간의 유사도에 기초하여 신규 뉴스와 진화 뉴스를 분류하여 제공할 수 있는 뉴스 제공 시스템 및 방법을 제공한다.

[0008] 또한, 본 발명의 일부 실시예는 신규 뉴스와 진화 뉴스를 분류하는 최적의 알고리즘을 사용하는 뉴스 제공 시스템 및 방법을 제공한다.

과제 해결수단

[0009] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면은 뉴스 비디오를 뉴스 기사 단위로 분할하고, 분할된 뉴스 비디오들을 카테고리 축, 시간 축 및 방송사 축을 기준으로 분류하여 데이터 큐브를 생성하는 뉴스 데이터 제공 장치 및 상기 뉴스 데이터 제공 장치로부터 검색 조건에 해당하는 뉴스 비디오를 검

색하여 사용자 단말에 제공하는 뉴스 검색 서버를 포함하되, 상기 뉴스 검색 서버는 상기 검색 조건에 해당하는 뉴스 비디오를 포함하는 뉴스 리스트를 생성하고, 상기 생성된 뉴스 리스트에 포함된 뉴스 비디오들간의 유사도를 나타내는 유사도 행렬을 산출하는 뉴스 리스트 생성부 및 상기 뉴스 비디오의 워드 데이터와 영상 프레임에 기초하여 산출된 뉴스 비디오들 간의 유사도 행렬에 따라 상기 뉴스 리스트에 포함된 상기 뉴스 비디오들을 신규 뉴스 또는 진화 뉴스로 분류하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 뉴스 리스트 분류부를 포함하는 뉴스 제공 시스템을 제공한다.

[0010] 또한, 본 발명의 제 2 측면은 뉴스 데이터베이스로부터 검색 조건에 해당하는 뉴스 비디오를 검색하여 뉴스 리스트를 생성하는 단계, 사용자 단말로부터 상기 뉴스 리스트의 분류 요청을 수신한 경우, 상기 뉴스 리스트에 포함된 각 뉴스 비디오의 워드 데이터와 영상 프레임에 기초하여 뉴스 비디오들 간의 유사도를 나타내는 유사도 행렬을 산출하는 단계, 상기 산출된 유사도 행렬에 따라 상기 뉴스 리스트에 포함된 상기 뉴스 비디오들을 신규 뉴스 또는 진화 뉴스로 분류하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하는 단계, 상기 생성된 신규 뉴스 리스트 및 진화 뉴스 리스트를 상기 사용자 단말에 전송하는 단계를 포함하되, 상기 뉴스 데이터베이스는 뉴스 비디오를 뉴스 기사 단위로 분할하고, 상기 분할된 뉴스 비디오들을 카테고리 별, 시간대 별 및 방송사 별로 분류하여 저장하는 것인 뉴스 제공 방법을 제공한다.

효 과

[0011] 전술한 본 발명의 과제 해결 수단에 의하면, 검색 조건에 해당하는 뉴스 리스트를 최소의 비용으로 생성할 수 있음과 동시에, 생성된 뉴스 리스트에서 신규 뉴스와 진화 뉴스를 용이하게 분류할 수 있다. 한편, 본 발명에서 제공하는 데이터 큐브 시스템은 멀티미디어 데이터인 뉴스 비디오에 대하여 전통적인 키워드 기반의 뉴스 비디오 기사의 검색을 제공할 뿐만 아니라 다차원의 의미 수준 데이터 분석과 데이터 마이닝을 위한 프레임워크를 제공한다. 따라서 정부기관, 언론사 및 연구소의 정책 결정 과정과 연구자료 확보에 많은 도움을 줄 수 있으며, 뉴스 비디오의 단순한 색인 및 검색에만 머물러 있던 기존 시스템에서 뉴스 비디오간의 관계 및 주제 분류를 통한 다각적인 분석 작업을 수행할 수 있는 차세대 뉴스 비디오 분석 및 비디오 마이닝의 프레임워크로서의 역할을 수행할 수 있다.

발명의 실시를 위한 구체적인 내용

[0012] 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.

[0013] 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.

[0014] 도 1은 본 발명의 일 실시예에 따른 뉴스 제공 시스템을 도시한 도면이다.

[0015] 뉴스 제공 시스템(100)은 뉴스 데이터 제공 장치(110), 뉴스 검색 서버(120) 및 사용자 단말(140)을 포함한다. 이때, 사용자 단말(140)은 뉴스 검색 서버(120)와 네트워크(130)를 통해 접속되고, 네트워크(130)는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN) 또는 부가가치 통신망(Value Added Network; VAN) 등과 같은 유선 네트워크나 이동 통신망(mobile radio communication network) 또는 위성 통신망 등과 같은 모든 종류의 무선 네트워크로 구현될 수 있다. 상기 뉴스 제공 시스템(100)은 데이터베이스 시스템에서 기본으로 제공되는 정보검색, SQL 및 기본적인 통계 정보를 제공할 뿐만 아니라, 유사도 행렬과 표준 OLAP(Online analytical processing) 기능을 제공함으로써 다음과 같은 추가적인 분석 기능을 제공한다. 즉, 1) 특정 주제에 대해서 새롭게 발생하는 뉴스 비디오와 시간에 따른 뉴스 비디오의 진화/전개 (즉, 특정 주제에 대해서 새롭게 나타나는 부 주제의 뉴스 비디오, 해당 뉴스 비디오의 시간상 전개, 해당 기사의 소멸); 2) 시간상 특정 시기의 뉴스 비디오의 전개 상황 분석; 3) 방송국 및 방송 프로그램에 따른 주제별 뉴스 비디오의 요약 및 전개 상황; 4) 유사도 행렬을 통한 뉴스 비디오의 군집 분석, 군집 간 관계를 포함하는 네트워크 분석 등의 데

이터마이닝 분석 지원 등이 이에 해당한다.

- [0016] 뉴스 데이터 제공 장치(110)는 뉴스 분류부(112), 데이터 큐브 생성부(114), 데이터 큐브 저장부(116) 및 데이터베이스(118)를 포함한다.
- [0017] 뉴스 분류부(112)는 데이터베이스(118)에 저장된 뉴스 비디오를 기사 단위로 분할하고, 분할된 뉴스 비디오를 카테고리 별로 분류한다.
- [0018] 먼저 뉴스 비디오를 기사 단위로 분할하는 과정은 샷 경계 탐지(shot boundary detection) 단계, 앵커 샷 탐지(anchor shot detection) 단계 및 뉴스 비디오 경계 탐지(news story boundary detection) 단계로 구성 된다.
- [0019] 본 발명에서 사용하는 샷 경계 탐지 알고리즘은 특이 값 분해(SVD)를 기반으로 점증적 클러스터링 알고리즘인 ART(adaptive resonance theory)와 머서 커널(mercer kernel)을 결합한 구조를 갖는다.
- [0020] 도 2는 본 발명에 적용되는 뉴스 비디오 샷 경계 탐지 알고리즘을 도시한 도면이다.
- [0021] 본 발명에서 사용하는 샷 경계 탐지 알고리즘은 특이 값 분해(SVD)를 기반으로 점증적 클러스터링 알고리즘인 ART(adaptive resonance theory)와 머서 커널(mercer kernel)을 결합한 구조를 갖는 것으로 상세한 내용은 앞서 언급한 발명자의 연구 논문에 개시된 내용을 참조하기로 한다.
- [0022] 상기 알고리즘에 따르면 뉴스 비디오 샷 경계 탐지의 재현율을 높임으로써 앵커 샷 탐지 단계의 입력으로 사용되는 데이터의 오류를 최소화 할 수 있다. 또한, 급격한 장면 변환과 점증적 장면 변환을 하나의 알고리즘으로 탐지하여 한 번의 데이터 탐색으로 샷 분할을 수행할 수 있다. 또한, 분할된 샷들을 정적 샷과 동적 샷으로 분류함으로써 앵커 샷 탐지 단계의 탐색 공간을 축소할 수 있다.
- [0023] 다음으로, 앵커 샷 탐지는 피부색 탐지 모듈(skin color detection module), 얼굴 탐지 모듈(face detection module), SVM 모듈을 통해 이루어진다. 즉, 피부색 탐지 모듈과 얼굴 탐지 모듈을 통하여 피부색이 존재하지 않는 프레임들과 얼굴이 존재하지 않는 프레임들을 제거함으로써, 앵커 샷 탐지를 위한 탐색 공간을 대폭 축소시킨다. 그리고, SVM 모듈을 통해 뉴스 비디오 샷들을 탐지된 얼굴의 개수와 위치에 따라서 앵커 샷과 일반 샷으로 분류한다. 즉, 탐지된 얼굴이 하나인지 두 개인지 여부 또는 얼굴의 위치가 화면에서 왼쪽에 위치하는지, 오른쪽에 위치하는지 중앙에 위치하는지 여부로 분류를 한다.
- [0024] 한편, 뉴스 비디오는 일반적으로 앵커 샷과 다음 앵커 샷이 나오기까지 하나의 뉴스 비디오를 이룬다. 따라서, 이에 기초하여 스테이트 머신을 구성하여 뉴스 비디오의 기사 분할을 수행한다.
- [0025] 뉴스 분류부(112)는 상기 분할된 뉴스 비디오의 텍스트 정보(textual information)와 비주얼 정보(visual information)를 이용하여 뉴스 비디오들을 카테고리 별로 분류한다. 뉴스 비디오의 카테고리 별 분류는 코 클러스터링(Co-clustering) 알고리즘[X. Wu 등의 방법 도입, Threading and Auto-documenting News Videos: a promising solution to rapidly browse news topics, IEEE Signal Processing Magazine, Vol. 23, Issue. 2, pp. 59-68, 2006]을 사용한다.
- [0026] 도 3은 본 발명에 적용되는 뉴스 비디오의 주제별 분류 과정을 설명하기 위한 도면이고, 도 4는 상기 주제별 분류 과정에 사용되는 코 클러스터링 알고리즘을 도시한 도면이다.
- [0027] 먼저 뉴스 분류부(112)를 통해 분할된 뉴스 비디오들이 제공된다(S310).
- [0028] 다음으로, 코 클러스터링 알고리즘을 통해, 뉴스 비디오를 카테고리 단위로 분류 한다. 뉴스 비디오는 텍스트와 비디오로 이루어지므로, 텍스트 단위로 클러스터링을 수행하고, 비디오 단위로 클러스터링을 각각 수행한 후, 각 클러스터에 공통으로 포함되는 뉴스 비디오가 동일한 카테고리를 갖는 것으로 판단한다.
- [0029] 도 4에 도시된 바와 같이, 텍스트를 구성하는 워드 데이터 단위로 클러스터링을 수행하고, 비디오를 구성하는 영상 프레임 단위로 클러스터링을 수행한다.
- [0030] 그리고 서로 다른 두 개의 뉴스 비디오가 있는 경우, 두 뉴스 비디오가 동일한 워드 클러스터 및 프레임 클러스터에 포함되는 경우라면 동일한 카테고리 클러스터에 포함시킨다. 반면, 포함되는 워드 클러스터가 상이하거나 프레임 클러스터가 상이한 경우에는 상이한 카테고리 클러스터에 포함시킨다.
- [0031] 한편, 뉴스 비디오에 대하여 클러스터링을 완료한 경우, 뉴스 비디오가 속한 워드 클러스터, 프레임 클러스터 및 카테고리 클러스터에 대한 정보를 각각 인덱싱하여 저장한다(S360). 이러한 정보는 추후 데이터 큐브를 구축하는 과정에 사용된다.

- [0032] 다시 도 1을 참조하면, 데이터 큐브 생성부(114)는 뉴스 분류부(112)를 통해 토픽 단위로 분류된 뉴스 비디오를 뉴스가 생성된 시간, 뉴스를 제작한 방송사 및 뉴스의 카테고리 단위로 재분류한 데이터 큐브를 생성한다.
- [0033] 도 5과 도 6은 본 발명에 적용되는 데이터 큐브를 설명하기 위한 도면이다.
- [0034] 먼저, 도 5에 도시된 바와 같이, 뉴스 분류부(112)를 통해 카테고리 단위로 분류된 뉴스 비디오를 해당 뉴스 비디오가 생성 또는 방송된 시간, 해당 뉴스를 제작한 방송사 단위로 재분류한다. 즉, 카테고리 축, 시간 축, 방송사 축을 기준으로 각 뉴스 비디오를 재배치하고, 각 좌표에 배치된 기사에 대하여 별도의 식별 번호를 부여하여 분류한다. 이때, 방송사 단위는 다시 각 방송사별 뉴스 프로그램의 단위로 재분류될 수 있다.
- [0035] 데이터 큐브를 뉴스 비디오 단위로 도시하면, 도 6과 같이 도시할 수 있다. 즉, 하나의 뉴스 비디오는 시간 정보(time_key), 카테고리 정보(category_key), 프로그램 정보(program_key) 및 고유 식별 번호(story_unit_no)를 포함한다. 그리고 시간 정보, 카테고리 정보 및 프로그램 정보는 각각 하위 체계의 정보를 더 포함할 수 있다. 예를 들어, 시간 정보는 뉴스가 생성 또는 방송된 년, 월, 일, 시간 등에 대한 정보를 포함할 수 있다. 또한, 프로그램 정보는 방송 프로그램 명칭 외에 방송사 정보를 포함할 수 있다.
- [0036] 이와 같은 과정에 따라, 뉴스 비디오를 체계적으로 분류할 수 있고, 뉴스 비디오의 주제별 분석 및 데이터마이닝을 수행할 수 있다. 예를 들어, 특정 기간 동안, 특정 방송사에서 제작한 특정 카테고리의 기사를 용이하게 검색할 수 있으며, 추후 설명할 뉴스 검색 서버(120)를 통해, 검색된 뉴스 비디오를 재가공 할 수 있다.
- [0037] 데이터 큐브 저장부(116)는 상기 데이터 큐브 생성부(114)를 통해 생성된 데이터 큐브를 저장한다. 이때, 저장되는 값은 각 뉴스 비디오의 분류 기준 및 데이터 큐브에서의 좌표 또는 식별 번호, 뉴스 비디오가 데이터베이스(118)에서 위치하는 주소값이 될 수 있으며, 뉴스 비디오 자체는 데이터베이스(118)에 저장된다.
- [0038] 데이터베이스(118)는 뉴스 분류부(112)를 통해 분류된 뉴스 비디오를 저장한다. 이때, 뉴스 비디오의 텍스트 데이터 및 비디오 데이터를 같이 저장된다. 또한, 각 뉴스 비디오는 뉴스가 생성 또는 방송된 시간, 해당 뉴스 비디오를 제작한 방송사, 방송 프로그램 정보 등이 저장된 메타 데이터를 포함할 수 있다.
- [0039] 도 7은 본 발명의 데이터베이스에 저장되는 뉴스 비디오의 메타 데이터 정보를 도시한 도면이다.
- [0040] 도시된 바와 같이, 각각의 뉴스 비디오는 고유 식별 번호(story_unit_no), 날짜, 기사 제목, 키워드, 키프레임, 뉴스 클립 이미지 등을 메타 데이터로 포함할 수 있다. 다만, 이에 한정되는 것은 아니며, 실시자의 선택에 따라 변경될 수 있다.
- [0041] 다시 도 1을 참조하면, 뉴스 검색 서버(120)는 인터페이스부(122), 뉴스 리스트 생성부(124) 및 뉴스 리스트 분류부(126)를 포함한다.
- [0042] 인터페이스부(122)는 사용자 단말(140)로부터 뉴스의 검색 조건을 수신하고, 이를 뉴스 리스트 생성부(124)에 전달하여 뉴스 리스트가 생성되도록 한다. 예를 들어, 뉴스가 생성 또는 방송된 시간, 방송사 또는 방송 프로그램 정보, 뉴스 비디오의 키워드 또는 비디오 프레임 등과 같은 검색 조건을 수신한다. 또한, 인터페이스부(122)는 사용자 단말(140)로부터 신규 뉴스 리스트 생성 요청 또는 진화 뉴스 리스트 생성 요청을 수신하고, 이를 뉴스 리스트 분류부(126)에 전달하여 신규 뉴스 리스트 또는 진화 뉴스 리스트가 생성되도록 한다. 또한, 생성된 신규 뉴스 리스트 또는 진화 뉴스 리스트를 사용자 단말(140)로 전송하여, 사용자 단말(140)에 표시시킨다.
- [0043] 뉴스 리스트 생성부(124)는 인터페이스부(122)가 수신한 검색 조건에 해당하는 뉴스 비디오를 데이터 큐브에서 검색하여 뉴스 리스트를 생성한다. 또한 상기 생성된 뉴스 리스트에 포함된 뉴스 비디오들간의 유사도를 나타내는 유사도 행렬을 산출한다.
- [0044] 전통적인 데이터 큐브 모델의 경우 각 큐브를 구성하는 각 셀은 수치 데이터를 저장하기 때문에, 합산 함수(aggregation function)와 같은 수치 연산을 통해, 특정 검색 조건에 해당하는 셀들의 데이터를 합산 할 수 있었다. 그러나 본 발명에 따른 데이터 큐브 모델의 경우에는 각 셀에 저장된 수치 데이터를 합산하더라도 의미 있는 데이터를 얻을 수 없기 때문에, 이러한 합산 함수를 단순히 적용할 수 없다.
- [0045] 한편, 종래의 연구에서는 상기 검색 조건에 해당하는 뉴스 비디오를 검색하여, 각 뉴스 비디오 별 의존도 또는 관련성에 기초하여 이진 트리 형태로 뉴스 비디오를 합산하는 과정을 수행하였다. 그러나, 이진 트리 형태의 자료 구조를 생성하는 과정에서 리소스 낭비가 상당하여, 검색 결과를 사용자에게 보여주는데 상당한 시간이 소요되는 문제점이 있었다.

[0046] 본 발명에서는 리스트 형태의 자료 구조를 이용하여 뉴스 리스트를 생성함으로써, 종래 기술의 문제점을 해결하고자 한다. 즉, 검색 조건에 해당하는 뉴스 비디오를 데이터 큐브에서 검색하여 하나의 전체 뉴스 리스트를 생성한다.

[0047] 한편, 유사도 행렬은 다음 수학적식을 통해 산출한다.

수학적식 1

$$SM(S_i, S_j) = \frac{\sum_{k=1}^m w_k(S_i)w_k(S_j) + \sum_{l=1}^n v_l(S_i)v_l(S_j)}{\sqrt{\sum_{k=1}^m w_k(S_i)^2 \sum_{k=1}^m w_k(S_j)^2} \sqrt{\sum_{l=1}^n v_l(S_i)^2 v_l(S_j)^2}}$$

where, $w_k(S_i)$ is the weight of word w_k in news story S_i ,
 $v_l(S_i)$ is the weight of key-frame v_l in news story S_i .

[0048]

[0049] 또한, 수학적식 1에서 $w_k(S_i)$ 와 $v_l(S_i)$ 는 다음 수학적식을 통해 산출한다.

수학적식 2

$$w_k(S_i) = \frac{tf_{w_k, S_i}}{tf_{w_k, S_i} + 0.5 + (1.5 * \frac{len(S_i)}{awl})} * \frac{\log \frac{g+0.5}{sf_{wk}}}{\log(g+1.0)}$$

$$v_l(S_i) = \begin{cases} \frac{tf_{v_l, S_i}}{tf_{v_l, S_i} + 0.5 + (1.5 * \frac{kflen(S_i)}{akfl})} * \frac{\log \frac{g+0.5}{sf_{vl}}}{\log(g+1.0)} & \text{if } v_l \notin NDK \\ \alpha \frac{tf_{v_l, S_i}}{tf_{v_l, S_i} + 0.5 + (1.5 * \frac{kflen(S_i)}{akfl})} * \frac{\log \frac{g+0.5}{sf_{vl}}}{\log(g+1.0)} & \text{if } v_l \in NDK \end{cases}$$

[0050]

[0051] 이때, $tf_{wk, si}$ 는 S_i 내에서 워드의 빈도수, $tf_{vl, si}$ 는 S_i 내에서 키 프레임의 빈도수, awl 은 뉴스 비디오에서의 평균 워드 개수, $akfl$ 은 뉴스 비디오에서의 평균 키 프레임의 개수, sf_{wk} 는 워드 w_k 를 포함하는 뉴스 비디오의 개수, sf_{vl} 는 키 프레임 v_l 을 포함하는 뉴스 비디오의 개수, $len(S_i)$ 는 S_i 내에서 워드의 개수, $kflen(S_i)$ 는 S_i 내에서 키 프레임의 개수, g 는 카테고리에 대한 뉴스 비디오의 개수, α 는 가중치를 나타낸다.

[0052] 뉴스 리스트 분류부(126)는 인터페이스부(122)를 통해 사용자의 요청이 전달된 경우, 뉴스 리스트 생성부(124)에서 생성된 전체 뉴스 리스트를 유사도 행렬에 기초하여 신규 뉴스 리스트와 진화 뉴스 리스트로 분류한다.

[0053] 도 8은 본 발명의 일 실시예에 따른 뉴스 리스트 분류에 사용되는 알고리즘을 도시한 도면이고, 도 9는 본 발명의 일 실시예에 따른 뉴스 리스트 분류 과정을 설명하기 위한 도면이다.

- [0054] 먼저, 뉴스 리스트 생성부(124)에서 생성된 전체 뉴스 리스트(AL)에 포함된 제 1 뉴스 비디오(뉴스 비디오_1)를 제 1 신규 뉴스 리스트(NL(1))에 포함시킨다. 그리고, 제 1 신규 뉴스 리스트(NL(1))와 유사한 뉴스를 포함하는 제 1 진화 뉴스 리스트(EL(1))를 생성하여 제 1 신규 뉴스 리스트(NL(1))와 연계시킨다.
- [0055] 한편, 상기 뉴스 리스트 생성부(124)를 통해 산출된 유사도 행렬에 기초하여, 유사도가 임계값 이상인 경우에는 해당 기사에서 진화된 진화 기사인 것으로 판단하고, 제 1 신규 뉴스 리스트(NL(1))와 연계된 제 1 진화 뉴스 리스트(EL(1))에 포함시킨다.
- [0056] 그러나, 유사도가 임계값 이하인 경우에는 해당 기사와 다른 신규 기사인 것으로 판단하고, 제 2 신규 뉴스 리스트(NL(2))를 생성하여, 해당 리스트에 포함시킨다. 또한, 제 2 신규 뉴스 리스트(NL(2))와 유사한 뉴스를 포함하는 제 2 진화 뉴스 리스트(EL(2))를 생성하여 제 2 신규 뉴스 리스트(NL(2))와 연계시킨다.
- [0057] 도 9의 예에서는, 제 1 뉴스 비디오(뉴스 비디오_1)가 제 1 신규 뉴스 리스트(NL(1))에 포함된다. 또한, 제 2 뉴스 비디오(뉴스 비디오_2)는 제 1 뉴스 비디오(뉴스 비디오_1)와 유사도가 낮아 제 2 신규 뉴스 리스트(NL(2))에 포함된다. 제 3 뉴스 비디오(뉴스 비디오_3)는 제 1 뉴스 비디오(뉴스 비디오_1)와 유사도가 높아 제 1 신규 뉴스 리스트(NL(1))와 연계된 제 1 진화 뉴스 리스트(EL(1))에 포함된다. 또한, 제 4 뉴스 비디오(뉴스 비디오_4)는 제 2 뉴스 비디오(뉴스 비디오_2)와 유사도가 높아 제 2 신규 뉴스 리스트(NL(2))와 연계된 제 2 진화 뉴스 리스트(EL(2))에 포함된다.
- [0058] 다시 도 1을 참조하면, 사용자 단말(140)은 사용자 인터페이스를 실행하여 사용자로부터 뉴스 비디오의 검색 조건 또는 신규 뉴스 리스트 및 진화 뉴스 리스트의 생성 요청을 입력 받는다.
- [0059] 이때, 사용자 단말(140)은 네트워크(130)를 통해 뉴스 검색 서버(120)에 접속할 수 있는 컴퓨터나 휴대용 단말기로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop) 등을 포함하고, 휴대용 단말기는 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, PCS(Personal Communication System), GSM(Global System for Mobile communications), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(W-Code Division Multiple Access), Wibro(Wireless Broadband Internet) 단말 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.
- [0060] 또한, 본 발명의 실시예에 따른 도 1에 도시된 구성 요소들은 소프트웨어 또는 FPGA(Field Programmable Gate Array) 또는 ASIC(Application Specific Integrated Circuit)와 같은 하드웨어 구성 요소를 의미하며, 소정의 역할들을 수행한다.
- [0061] 그렇지만 '구성 요소들'은 소프트웨어 또는 하드웨어에 한정되는 의미는 아니며, 각 구성 요소는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다.
- [0062] 따라서, 일 예로서 구성 요소는 소프트웨어 구성 요소들, 객체지향 소프트웨어 구성 요소들, 클래스 구성 요소들 및 태스크 구성 요소들과 같은 구성 요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다.
- [0063] 구성 요소들과 해당 구성 요소들 안에서 제공되는 기능은 더 작은 수의 구성 요소들로 결합되거나 추가적인 구성 요소들로 더 분리될 수 있다.
- [0064] 도면을 통해 사용자 단말(140)에서 실행되는 사용자 인터페이스를 살펴보기로 한다.
- [0065] 도 10은 본 발명의 일 실시예에 따라 사용자 단말에서 실행되는 사용자 인터페이스를 도시한 도면이고, 도 11a~11c는 상기 사용자 인터페이스를 통한 검색 조건 설정 방법을 설명하기 위한 도면이다.
- [0066] 사용자 인터페이스(1100)에는 방송사 선택부(1110), 시간 선택부(1120), 카테고리 선택부(1130), 키워드 입력부(1140), 신규 뉴스 리스트 표시부(1150), 진화 뉴스 리스트 표시부(1160), 신규 뉴스 리스트 생성 요청부(1170) 및 진화 뉴스 리스트 생성 요청부(1180)를 포함한다.
- [0067] 사용자는 방송사 선택부(1110), 시간 선택부(1120), 카테고리 선택부(1130), 키워드 입력부(1140)를 통해, 뉴스 데이터 제공 장치(110)의 데이터 큐브에서 검색하고자 하는 뉴스 비디오의 검색 조건을 입력할 수 있다. 특히, 도 11a, 11b, 11c에 도시된 바와 같이 OLAP의 롤업(roll up), 드릴 다운(drill down), 슬라이스(slice), 다이

스(dice) 기능을 이용하여 다양한 검색 조건을 설정할 수 있다. 슬라이스 기능을 통해 특정 기간에 해당하는 모든 방송사 및 모든 카테고리에 속하는 기사를 검색할 수 있다. 또한, 다이스 기능을 통해 특정 방송사, 특정 시간 및 특정 카테고리에 해당하는 기사를 검색할 수 있다.

- [0068] 또한, 사용자는 신규 뉴스 리스트 생성 요청부(1170)와 진화 뉴스 리스트 생성 요청부(1180)를 통해, 뉴스 리스트를 분류하여 표시하도록 요청할 수 있다.
- [0069] 신규 뉴스 리스트 표시부(1150), 진화 뉴스 리스트 표시부(1160)에 표시된 바와 같이, 복수의 뉴스 비디오가 텍스트 정보 및 프레임 정보를 포함한 상태로 표시되고 있다.
- [0070] 도 12은 본원 발명의 일 실시예에 따른 뉴스 검색 제공 방법을 도시한 순서도이다.
- [0071] 먼저, 데이터 큐브 형태로 구성된 뉴스 데이터가 제공된다(S1210). 앞서 설명한 뉴스 데이터 제공 장치(110)를 통해, 뉴스 비디오 단위로 분류되고, 시간 축, 카테고리 축, 방송사 축으로 분류되어 생성된 데이터 큐브 형태의 뉴스 데이터가 제공된다.
- [0072] 다음으로, 사용자 단말(140)로부터 검색 조건을 수신한다(S1220). 뉴스 비디오에 포함되는 키워드뿐만 아니라, 데이터 큐브의 기준축이 되는 시간, 카테고리, 방송사 정보 등이 검색 조건이 될 수 있다.
- [0073] 다음으로, 상기 검색 조건에 해당하는 뉴스 비디오를 검색하여 뉴스 리스트를 생성한다(S1230). 또한, 생성된 뉴스 리스트에 포함된 뉴스 비디오들에 대하여 유사도 행렬을 산출한다.
- [0074] 다음으로, 상기 생성된 뉴스 리스트에 대하여 사용자로부터 뉴스 리스트 분류 요청이 수신되면(S1240), 상기 생성된 유사도 행렬에 따라 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성한다(S1250). 즉, 뉴스 비디오에 포함되는 워드와 프레임에 기초하여 산출된 유사도 행렬에 따라, 신규 뉴스인지 신규 뉴스에 대한 진화 뉴스인지를 판별하여, 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성한다.
- [0075] 다음으로, 사용자 단말에 신규 뉴스 리스트 및 진화 뉴스 리스트를 표시한다(S1260).
- [0076] 이와 같이, 검색 조건에 해당하는 뉴스 리스트의 자료 구조가 단순하고, 사용자의 요청이 있는 경우에 한하여 신규 뉴스 리스트 및 진화 뉴스 리스트를 생성하므로, 리소스의 낭비를 최소화할 수 있다.
- [0077] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 메커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.
- [0078] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.
- [0079] 진술한 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.
- [0080] 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.

도면의 간단한 설명

- [0081] 도 1은 본 발명의 일 실시예에 따른 뉴스 제공 시스템을 도시한 도면이다.
- [0082] 도 2는 본 발명에 적용되는 뉴스 비디오 샷 경계 탐지 알고리즘을 도시한 도면이다.

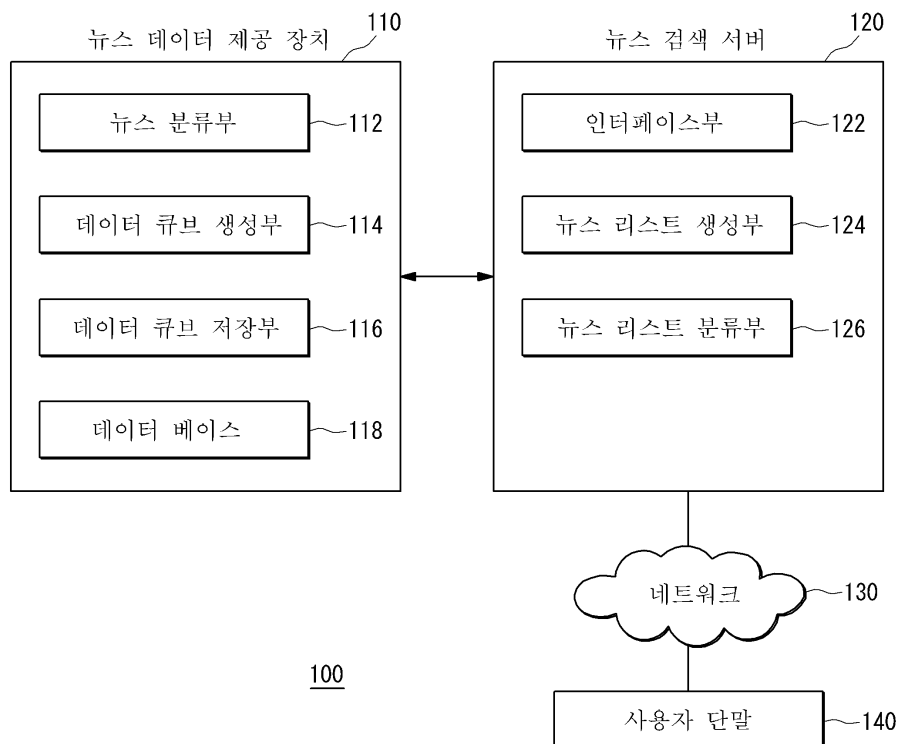
- [0083] 도 3은 본 발명에 적용되는 뉴스 비디오의 주제별 분류 과정을 설명하기 위한 도면이다.
- [0084] 도 4는 상기 주제별 분류 과정에 사용되는 코 클러스터링 알고리즘을 도시한 도면이다.
- [0085] 도 5과 도 6은 본 발명에 적용되는 데이터 큐브를 설명하기 위한 도면이다.
- [0086] 도 7은 본 발명의 데이터베이스에 저장되는 뉴스 비디오의 메타 데이터 정보를 도시한 도면이다.
- [0087] 도 8은 본 발명의 일 실시예에 따른 뉴스 리스트 분류에 사용되는 알고리즘을 도시한 도면이다.
- [0088] 도 9는 본 발명의 일 실시예에 따른 뉴스 리스트 분류 과정을 설명하기 위한 도면이다.
- [0089] 도 10은 본 발명의 일 실시예에 따라 사용자 단말에서 실행되는 사용자 인터페이스를 도시한 도면이다.
- [0090] 도 11a~11c는 상기 사용자 인터페이스를 통한 검색 조건 설정 방법을 설명하기 위한 도면이다.
- [0091] 도 12은 본원 발명의 일 실시예에 따른 뉴스 검색 제공 방법을 도시한 순서도이다.

[0092] <도면의 주요 부분에 대한 설명>

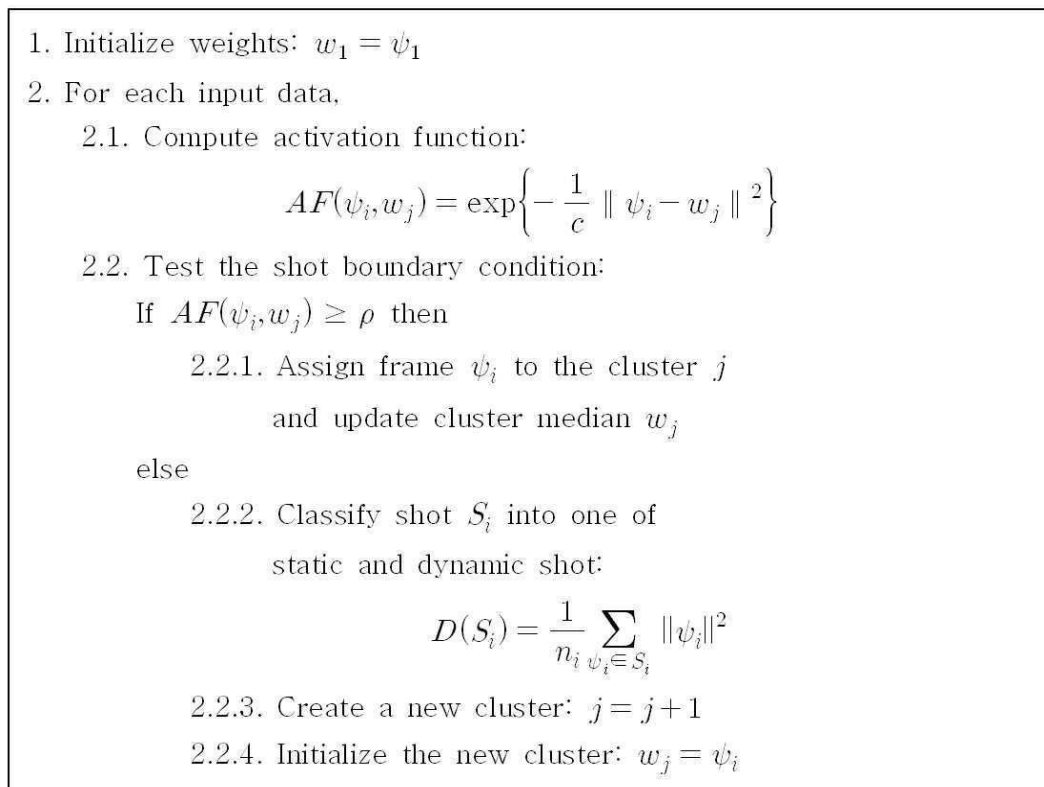
- | | |
|------------------------|-------------------|
| [0093] 100: 뉴스 제공 시스템 | 110: 뉴스 데이터 제공 장치 |
| [0094] 112: 뉴스 분류부 | 114: 데이터 큐브 생성부 |
| [0095] 116: 데이터 큐브 저장부 | 118: 데이터 베이스 |
| [0096] 120: 뉴스 검색 서버 | 122: 인터페이스부 |
| [0097] 124: 뉴스 리스트 생성부 | 126: 뉴스 리스트 분류부 |
| [0098] 130: 네트워크 | 140: 사용자 단말 |

도면

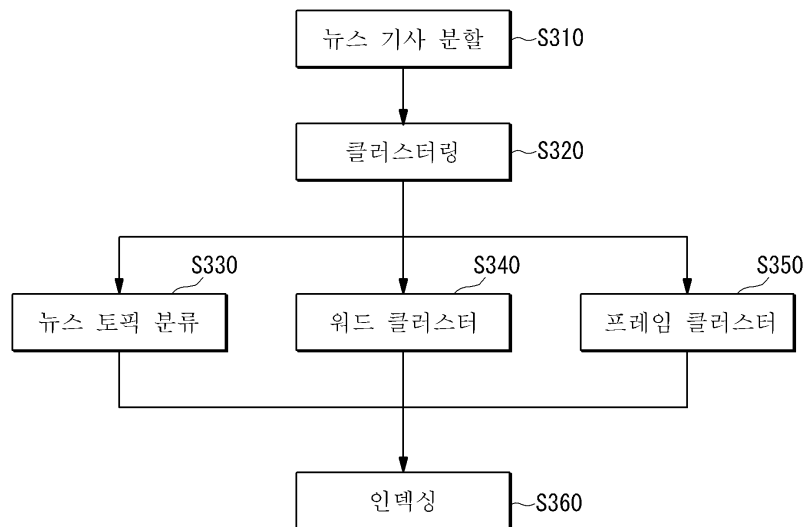
도면1



도면2



도면3



도면4

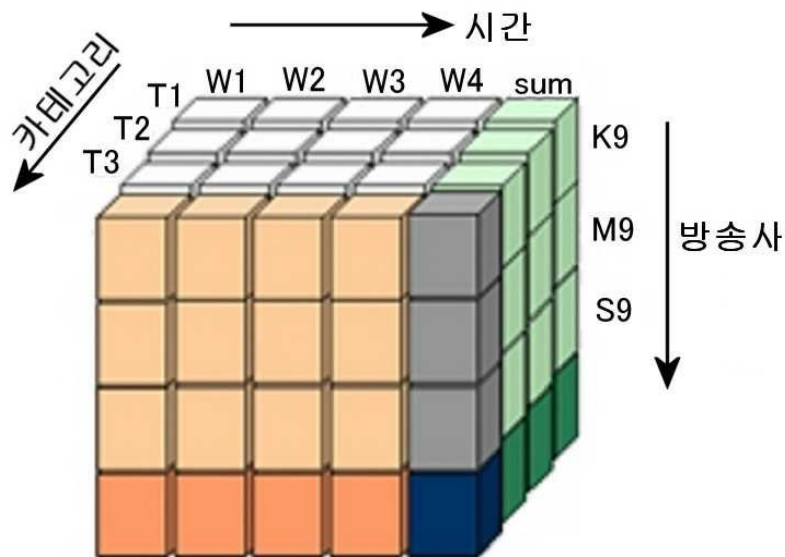
1. Set up word-by-story matrix A_1 and key-frame-by-story matrix A_2 .

$$A_1(i, j) = W_{ij} = tf_{ij} \times idf_i$$

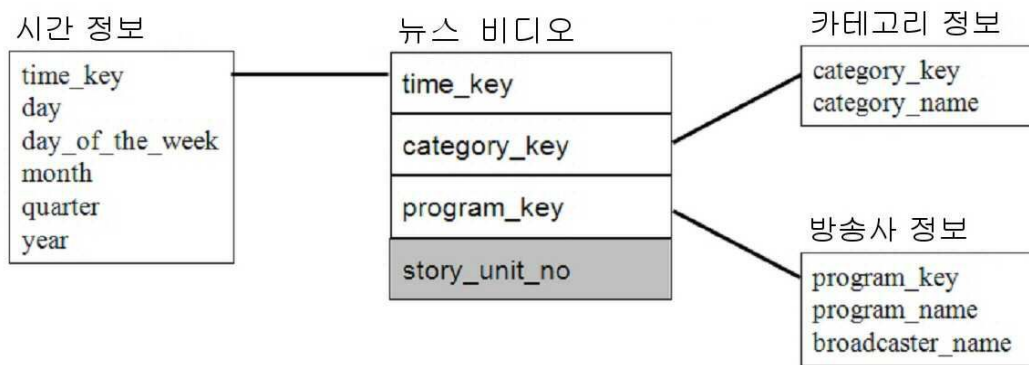
$$A_2(i, j) = W_{ij} = kf_{ij} \times \log_2(N/sf_i)$$
2. Construct the matrix $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$.
3. Calculate $A_n = D_1^{-1/2} A D_2^{-1/2}$.
4. Compute $l = \lceil \log_2 k \rceil$ singular vectors of A_n , $u_2, u_3, \dots, u_{l+1}$ and $v_2, v_3, \dots, v_{l+1}$ and form the matrix Z .

$$Z = \begin{bmatrix} D_1^{-1/2} U \\ D_2^{-1/2} V \end{bmatrix}$$
5. Run the k-means algorithm on the l -dimensional data Z to obtain the desired k -way multi-partition.

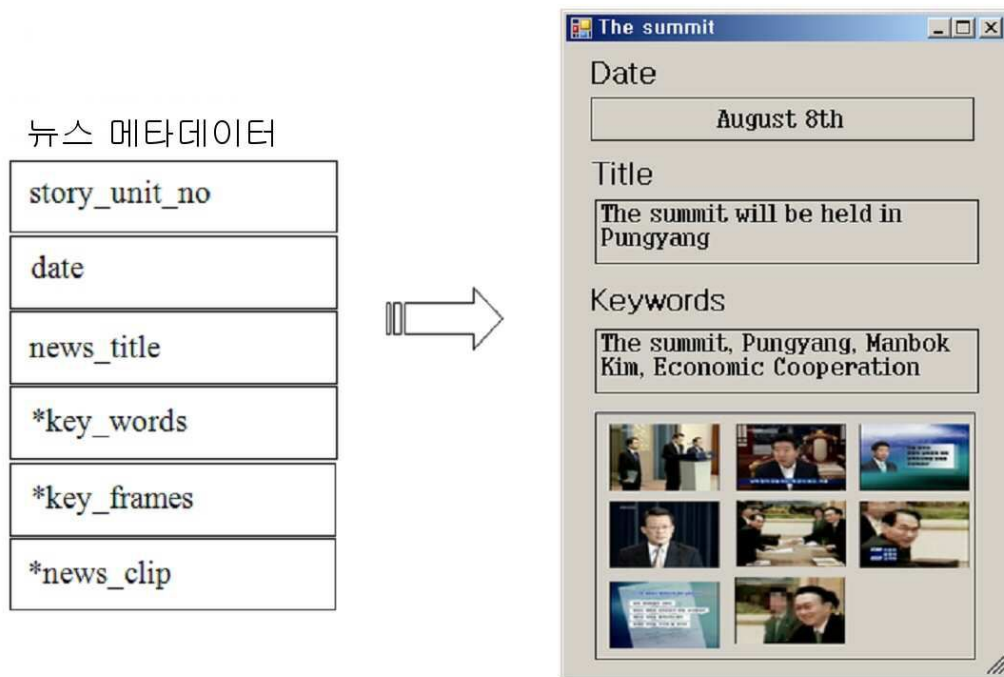
도면5



도면6



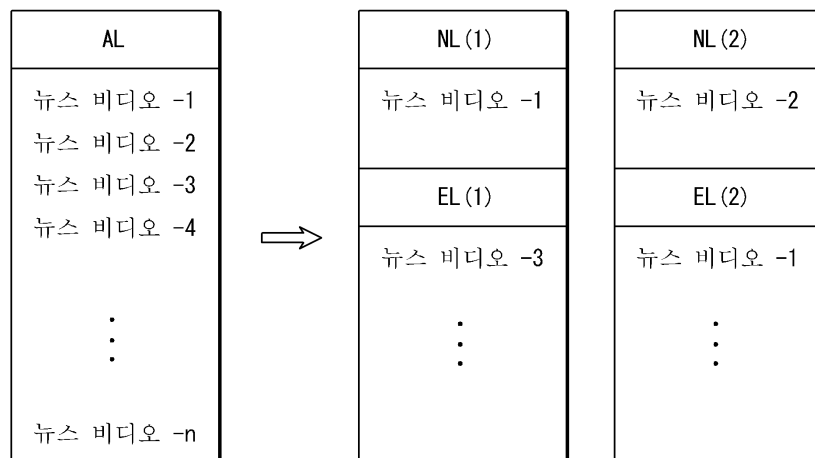
도면7



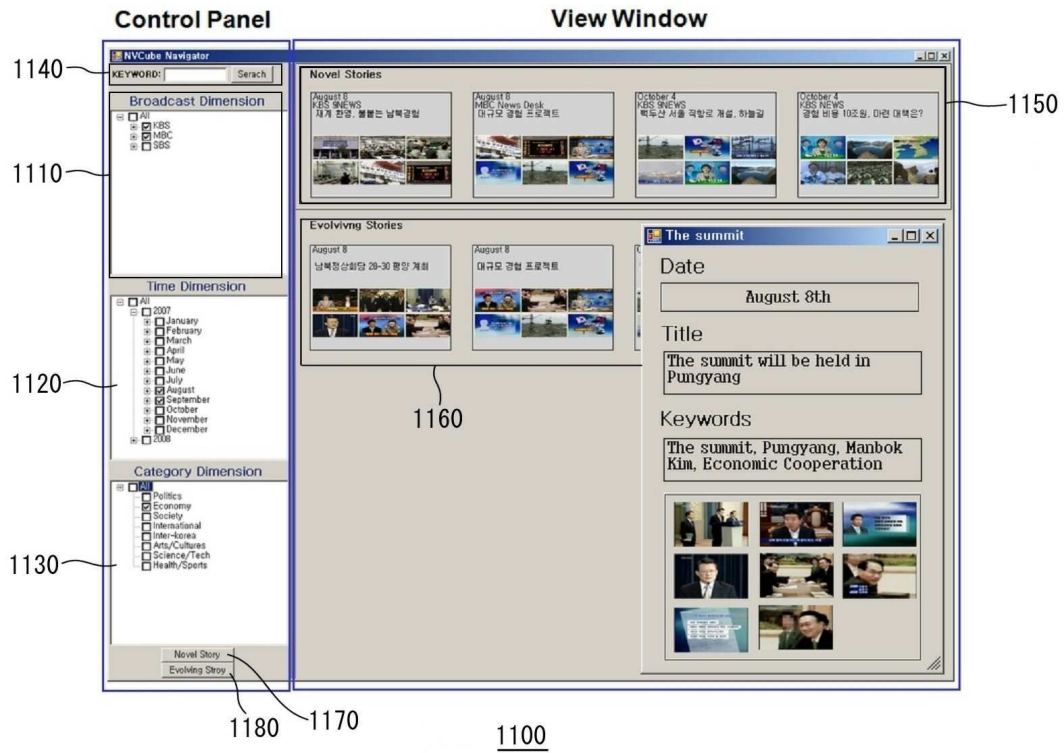
도면8

Input:	Similarity matrix SM , Aggregation list AL .
Output:	Novel story list NL , Evolving story list EL .
Method:	
1.	Set the first news story in AL to novel story and add to NL . Create a $EL(1)$, $k=1$, and link to first element in NL .
2.	Find $M_i = \arg \max_{1 \leq j \leq i} SM_{ij}$, for the i -th news story in AL .
3.	If the $M_i \leq \text{threshold}$, set the i -th news story to novel story and add to NL . Create a $EL(k++)$ and link to first element in NL . Else, set the i -th news story to evolving story and add to $EL(k)$ which include j -th news story.
4.	Go to the step 2, until i less than # of elements in AL .
5.	Output NL and EL .

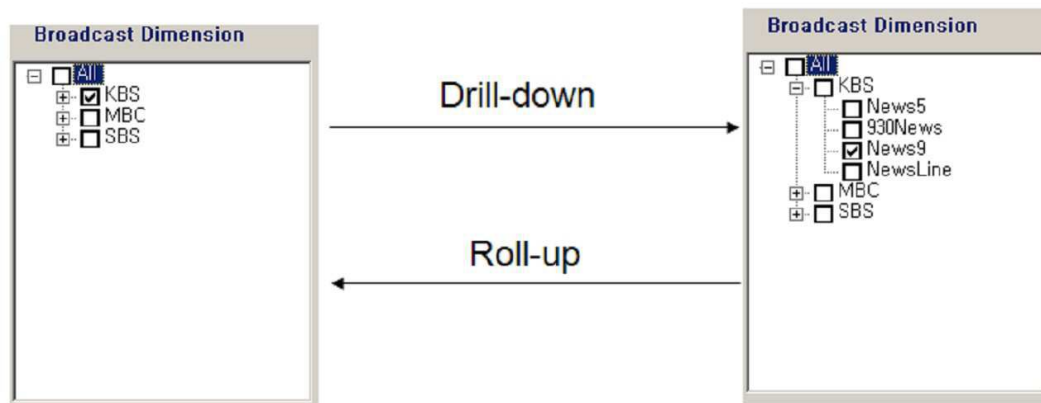
도면9



도면10



도면11a



도면11b

Broadcast Dimension	Time Dimension	Category Dimension
☒ All ☐ KBS ☐ MBC ☐ SBS	☐ All ☐ 2007 ☐ January ☐ February ☐ March ☐ April ☐ May ☐ June ☐ July ☒ August ☐ September ☐ October ☐ November ☐ December ☐ 2008	☒ All ☒ Politics ☒ Economy ☒ Society ☐ International ☐ Inter-korea ☒ Arts/Cultures ☒ Science/Tech ☒ Health/Sports

Slice for time = "August 2007"

도면11c

Broadcast Dimension	Time Dimension	Category Dimension
☐ All ☐ KBS ☐ News5 ☐ 930News ☒ News9 ☒ NewsLine ☐ MBC ☐ SBS	☐ All ☐ 2007 ☐ January ☐ February ☐ March ☐ April ☐ May ☐ June ☐ July ☒ August ☒ September ☐ October ☐ November ☐ December ☐ 2008	☐ All ☒ Politics ☒ Economy ☐ Society ☐ International ☐ Inter-korea ☐ Arts/Cultures ☐ Science/Tech ☐ Health/Sports

Dice for broadcast = "News9, KBS and NewsLine, KBS"
AND time = " August 2007 and September 2007"
AND category = "Politics and Economy"

도면12

