(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0097320
(43) 공개일자 2011년08월31일

(51) Int. Cl.
G06F 17/00 (2006.01) *G06F 19/00* (2011.01)
(21) 출원번호 10-2010-0017111
(22) 출원일자 2010년02월25일
심사청구일자 2010년02월25일

(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
특허법인 엠에이피에스

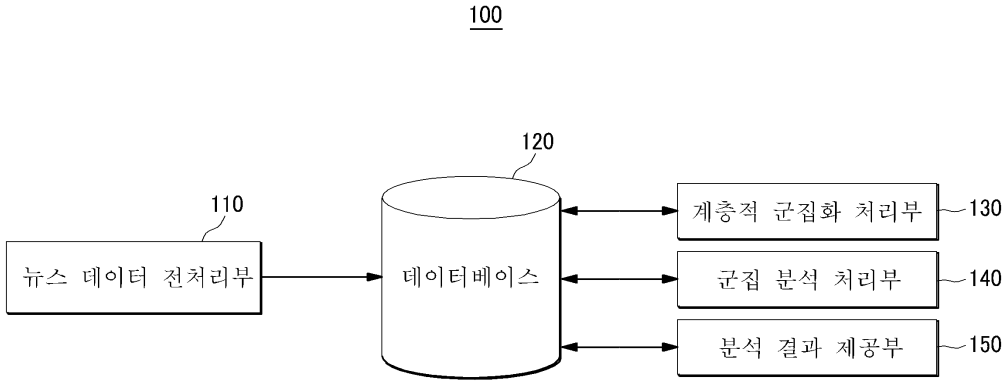
전체 청구항 수 : 총 12 항

(54) 뉴스 데이터 분석 장치 및 방법

(57) 요 약

뉴스 데이터 분석 방법은 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고, 뉴스 기사 클립들을 카테고리별로 군집화하고, 동일 카테고리에 포함된 뉴스 기사 클립들을 유사도에 기초하여 복수의 계층의 뉴스 기사 군집들로 군집화하고, 기설정된 군집 분석 처리 기준에 따라 뉴스 기사 군집들에 대해 기간별 성향 분석, 성장도 분석 및 패스파인더 네트워크 분석 중 적어도 하나에 대한 지적 구조 분석을 수행한다.

대 표 도 - 도1



(72) 발명자

유재학

충청북도 옥천군 옥천읍 죽향리 옥향아파트 103동 803호

임영희

경기도 수원시 영통구 매탄3동 우남퍼스트빌아파트 208동 702호

오승근

인천광역시 부평구 삼산동 삼산아파트 3동 501호

강봉수

경기도 용인시 기흥구 상하동 강남마을주공아파트 801동 103호

박승진

전라남도 목포시 옥암동 제일2차 하이빌마트 202동 105호

이 발명을 지원한 국가연구개발사업

과제고유번호 20090302185256

부처명 한국산업기술재단

연구관리전문기관

연구사업명 지역혁신인력양성사업

연구과제명 (1차년도)베터리로 동작하는 감시카메라용 비디오 압축 특성을 이용한 정보보호 및 상황 인지 연구를 통한 최적화

기여율

주관기관 고려대학교

연구기간 2009년 05월 01일 ~ 2010년 04월 30일

특허청구의 범위

청구항 1

뉴스 데이터 분석 장치에 있어서,

뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고, 상기 뉴스 기사 클립들을 카테고리별로 클러스터링하는 뉴스 데이터 전처리부;

상기 뉴스 기사 클립들을 유사도에 기초하여 복수의 계층의 뉴스 기사 군집으로 생성하는 계층적 군집화 처리부; 및

기설정된 군집 분석 처리 기준에 따라 상기 뉴스 기사 군집들에 대해 기간별 성향 분석, 성장도 분석 및 패스파인더 네트워크 분석 중 적어도 하나에 대한 지적 구조 분석을 수행하는 군집 분석 처리부를 포함하는 뉴스 데이터 분석 장치.

청구항 2

제 1 항에 있어서,

상기 지적 구조 분석 처리 결과에 따라 생성된 지적 정보를 식별 가능한 문자 및 그래프 정보 등으로 출력하는 분석 결과 제공부를 더 포함하는 뉴스 데이터 분석 장치.

청구항 3

제 1 항 또는 제 2 항에 있어서,

상기 군집 분석 처리부는,

뉴스 비디오 방송 시기 및 기간, 뉴스 기사의 주제 및 카테고리, 처리할 군집 분석의 종류 중 적어도 하나를 상기 군집 분석 처리 기준으로 설정하는 뉴스 데이터 분석 장치.

청구항 4

제 1 항 또는 제 2 항에 있어서,

상기 군집 분석 처리부는,

설정된 기간 내에 상기 뉴스 기사 군집들이 각각 방송된 횟수 및 방송된 시기에 기초하여 방송 성향을 분석하는 기간별 군집 성향 분석 모듈;

설정된 기간을 복수의 기간으로 분할하고, 분할된 각 기간 내에 상기 뉴스 기사 군집들이 각각 방송된 횟수에 기초하여 뉴스 기사 군집 별 성장도를 분석하는 군집 성장도 분석 모듈; 및

상기 뉴스 기사 군집들을 각각 가중치를 갖는 링크로 연결하고, 각 뉴스 기사군집들의 경로 거리에 기초하여 상기 뉴스 기사 군집들 간의 연관도를 분석하는 패스파인더 네트워크 분석 모듈 중 적어도 하나의 모듈을 포함하는 뉴스 데이터 분석 장치.

청구항 5

제 1 항 또는 제 2 항에 있어서,

상기 계층적 군집화 처리부는,

상기 뉴스 기사 군집들 간의 유사도에 기초하여 상위 뉴스 기사 군집을 생성하되,

하나의 최상위 군집이 생성될 때까지 상기 상위 뉴스 기사 군집을 생성하는 과정을 반복하는 뉴스 데이터 분석 장치.

청구항 6

제 1 항 또는 제 2 항에 있어서,

상기 계층적 군집화 처리부는,

상기 뉴스 기사 클립으로부터 추출한 단어 및 대표 영상의 가중치에 기초하여 상기 유사도를 산출하는 뉴스 데이터 분석 장치.

청구항 7

뉴스 데이터 분석 방법에 있어서,

뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고, 상기 뉴스 기사 클립들을 카테고리별로 군집화하는 단계;

동일 카테고리에 포함된 뉴스 기사 클립들을 유사도에 기초하여 복수의 계층의 뉴스 기사 군집들로 군집화하는 단계; 및

기설정된 군집 분석 처리 기준에 따라 상기 뉴스 기사 군집들에 대해 기간별 성향 분석, 성장도 분석 및 패스파인더 네트워크 분석 중 적어도 하나에 대한 지적 구조 분석을 수행하는 단계를 포함하는 뉴스 데이터 분석 방법.

청구항 8

제 7 항에 있어서,

상기 지적 구조 분석을 수행하는 단계 이후에,

상기 지적 구조 분석 처리 결과에 따라 생성된 지적 정보를 제공하는 단계를 더 포함하는 뉴스 데이터 분석 방법.

청구항 9

제 7 항 또는 제 8 항에 있어서,

상기 군집 분석 처리 기준은,

적어도 하나의 뉴스 비디오에 대한 방송 시기 및 기간, 분석 처리할 뉴스 기사의 주제 및 카테고리, 분석 처리할 군집 분석의 종류 중 적어도 하나를 포함하는 뉴스 데이터 분석 방법.

청구항 10

제 7 항 또는 제 8 항에 있어서,

상기 지적 구조 분석을 수행하는 단계는,

설정된 기간 내에 상기 뉴스 기사 군집들이 각각 방송된 횟수 및 방송된 시기에 기초하여 방송 성향을 분석하는 단계, 설정된 기간을 복수의 기간으로 분할하고 분할된 각 기간 내에 상기 뉴스 기사 군집들이 각각 방송된 횟수에 기초하여 뉴스 기사 군집 별 성장도를 분석하는 단계 및 상기 뉴스 기사 군집들을 각각 가중치를 갖는 링크로 연결하고 각 뉴스 기사군집들의 경로 거리에 기초하여 상기 뉴스 기사 군집들 간의 연관도를 분석하는 단

계 중 적어도 하나의 단계를 수행하는 뉴스 데이터 분석 방법.

청구항 11

제 7 항 또는 제 8 항에 있어서,

상기 복수의 계층의 뉴스 기사 군집들로 군집화하는 단계에서,

하나의 최상위 군집이 생성될 때까지 상기 뉴스 기사 군집들 간의 유사도에 기초하여 상위 뉴스 기사 군집을 생성하는 과정을 반복하는 뉴스 데이터 분석 방법.

청구항 12

제 7 항 또는 제 8 항에 있어서,

상기 복수의 계층의 뉴스 기사 군집들로 군집화하는 단계는,

기설정된 뉴스 기사의 주제 또는 카테고리에 상응하는 뉴스 기사 클립으로부터 추출된 단어 및 대표 영상의 가중치에 기초하여 상기 유사도를 산출하는 단계;

상기 뉴스 기사 클립들 간에 산출된 상기 유사도가 기설정된 임계 값 이상인 뉴스 기사 클립들을 동일 뉴스 기사 군집으로 군집화하는 단계를 포함하는 뉴스 데이터 분석 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 뉴스 데이터에 대한 분석 장치 및 방법에 관한 것으로, 더욱 상세하게는 뉴스 비디오를 분류하여 지적 구조 분석(Intellectual Structure Analysis)을 수행하는 뉴스 데이터 분석 장치 및 그 방법에 관한 것이다.

배 경 기 술

[0002] 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 대두 되었다.

[0003] 이러한 필요성과 관련하여, 최근에는 구조화된 뉴스 비디오에 대해 내/외부 구조 분석을 통해 뉴스 비디오 파싱(parsing), 색인, 브라우징 및 검색을 제공하는 방식에 대한 연구들이 활발히 진행되고 있다. 예를 들어, 뉴스 비디오의 구조를 분석하는 방식에는 토픽 쓰레딩(Topic-threading)과 오토 다큐멘팅(Auto-documenting) 등이 포함된다.

[0004] 구체적으로, 토픽 쓰레딩은 뉴스 비디오의 자막 및 스크립트로부터 단어들을 추출하여 미리 정의되어 있는 주제별로 분류함으로써 사용자로 하여금 관심 있는 뉴스 비디오들을 빠르게 브라우징할 수 있도록 지원하는 시스템이다.

[0005] 또한, 오토 다큐멘팅은 특정 주제의 뉴스 비디오들에 대해 토픽 트리(topic tree)를 생성하여 각 뉴스 비디오들을 의미를 고려하여 요약 및 정리함으로써 사용자에게 특정 주제에 대한 뉴스 비디오들의 간략한 요약을 제공하기 위한 시스템이다.

[0006] 그러나, 이와 같은 종래의 뉴스 비디오의 구조 분석 방식들은 단순히 검색어 기반의 뉴스 비디오 검색 및 브라우징에 한정되거나, 뉴스 비디오에 대한 부분적인 의미 분석 및 요약 등을 제공하는 수준에 그치고 있다.

[0007] 따라서, 멀티미디어 정보로서의 뉴스 비디오에 대한 지적 구조 분석을 통해 주제별 분석 및 데이터마이닝을 지원하여, 뉴스 비디오에 대한 검색 및 브라우징뿐만 아니라 뉴스 비디오에 관련된 새로운 지적 정보를 생성 및

제공할 수 있는 방법이 필요한 실정이다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 상술한 문제점을 해결하기 위한 것으로서, 본 발명의 실시예는 뉴스 비디오에 대한 계층적 군집화를 통해 체계적인 지적 구조 분석을 수행하고, 그에 따른 새로운 지적 정보를 생성할 수 있는 장치 및 방법을 제공하고자 한다.

과제의 해결 수단

[0009] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로써, 본 발명의 일 측면에 따른 뉴스 데이터 분석 장치는, 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고 상기 뉴스 기사 클립들을 카테고리별로 클러스터링하는 뉴스 데이터 전처리부, 상기 뉴스 기사 클립들을 유사도에 기초하여 복수의 계층의 뉴스 기사 군집으로 생성하는 계층적 군집화 처리부 및 기설정된 군집 분석 처리 기준에 따라 상기 뉴스 기사 군집들에 대해 기간별 성향 분석, 성장도 분석 및 패스파인더 네트워크 분석 중 적어도 하나에 대한 지적 구조 분석을 수행하는 군집 분석 처리부를 포함한다.

[0010] 또한, 본 발명의 다른 측면에 따른 뉴스 데이터 분석 방법은, 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고, 상기 뉴스 기사 클립들을 카테고리별로 군집화하는 단계, 동일 카테고리에 포함된 뉴스 기사 클립들을 유사도에 기초하여 복수의 계층의 뉴스 기사 군집들로 군집화하는 단계 및 기설정된 군집 분석 처리 기준에 따라 상기 뉴스 기사 군집들에 대해 기간별 성향 분석, 성장도 분석 및 패스파인더 네트워크 분석 중 적어도 하나에 대한 지적 구조 분석을 수행하는 단계를 포함한다.

발명의 효과

[0011] 전술한 본 발명의 과제 해결 수단에 의하면, 계층적 군집화된 뉴스 기사들에 대해 기간별 성향, 성장도 및 뉴스 기사 군집 간의 연관도 등의 지적 구조 분석이 가능하다.

[0012] 따라서, 정부기관, 언론사 및 연구소 등에서의 정책 결정 과정과 연구 자료 확보에 많은 도움을 줄 수 있으며, 뉴스 비디오에 대한 주제 분류 및 뉴스 기사 간의 관계에 대한 다각적인 지적 구조 분석을 수행할 수 있는 뉴스 비디오 분석 및 비디오 마이닝의 프레임워크로써의 역할을 수행할 수 있다.

[0013] 또한, 본 발명의 과제 해결 수단에 의하면, 뉴스 기사와 더불어 상기 뉴스 기사에 상응하는 뉴스 기사 군집들에 대해 생성된 지적 정보를 함께 제공할 수 있는 효과가 있다.

도면의 간단한 설명

[0014] 도 1은 본 발명의 실시예에 따른 뉴스 데이터 분석 장치의 구성을 나타내는 블록도이다.

도 2는 본 발명의 실시예에 따른 계층적 군집화 과정을 설명하기 위한 도면이다.

도 3은 본 발명의 실시예에 따른 군집 분석 처리부의 구성을 나타내는 블록도이다.

도 4는 본 발명의 실시예에 따른 기간별 군집 성향 분석 과정을 설명하기 위한 도면이다.

도 5는 본 발명의 실시예에 따른 군집 성장도 분석 과정을 설명하기 위한 도면이다.

도 6은 본 발명의 실시예에 따른 패스파인더 네트워크 분석 과정을 설명하기 위한 도면이다.

도 7은 본 발명의 실시예에 따른 뉴스 데이터 분석 방법을 설명하기 위한 순서도이다.

도 8은 본 발명의 실시예에 따른 군집 분석 처리 과정을 설명하기 위한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0015] 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.

[0016] 명세서 전체에서, 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.

[0017] 도 1은 본 발명의 실시예에 따른 뉴스 데이터 분석 장치의 구성을 나타내는 블록도이다.

[0018] 그리고, 도 2는 본 발명의 실시예에 따른 계층적 군집화 과정을 설명하기 위한 도면이다.

[0019] 먼저, 도 1에서와 같이, 본 발명의 실시예에 따른 뉴스 데이터 분석 장치(100)는 뉴스 데이터 전처리부(110), 데이터베이스(120), 계층적 군집화 처리부(130), 군집 분석 처리부(140) 및 분석 결과 제공부(150)를 포함한다.

[0020] 이때, 뉴스 데이터 전처리부(110)는 입력되는 뉴스 비디오를 뉴스 기사 클립 단위로 분할하고, 분할된 뉴스 기사 클립들에 대하여 코-클러스터링(co-clustering)을 수행하여 뉴스 기사 클립들을 카테고리별로 분류 및 색인(indexing)한다.

[0021] 구체적으로, 뉴스 데이터 전처리부(110)는 샷 경계 탐지(shot boundary detection) 단계, 앵커 샷 탐지(anchor shot detection) 단계 및 뉴스 비디오 경계 탐지(news story boundary detection) 단계 등의 과정을 통해 하나의 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할한다.

[0022] 본 발명의 실시예에서는 샷 경계 탐지 알고리즘으로써, 하기 알고리즘 1에서 나타낸 바와 같이 특이 값 분해(Singular Value Decomposition, SVD)를 기반으로 점증적 클러스터링 알고리즘인 ART(adaptive resonance theory)와 머서 커널(mercer kernel)을 결합한 구조 이용하는 것을 예로서 나타내었다.

[0023] 참고로, ART(adaptive resonance theory)와 머서 커널(mercer kernel)을 결합한 알고리즘에 대한 상세한 내용은, 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 내용을 참조하기로 한다.

[0024] [알고리즘 1]

1. Initialize weights: $w_1 = \psi_1$
2. For each input data,
   2.1. Compute activation function:
   $$AF(\psi_i, w_j) = \exp\left\{-\frac{1}{c} \parallel \psi_i - w_j \parallel^2\right\}$$
   2.2. Test the shot boundary condition:
   If $AF(\psi_i, w_j) \geq \rho$ then
      2.2.1. Assign frame ψ_i to the cluster j
      and update cluster median w_j
   else
      2.2.2. Classify shot S_i into one of
      static and dynamic shot:
      $$D(S_i) = \frac{1}{n_i} \sum_{\psi_i \in S_i} \|\psi_i\|^2$$
      2.2.3. Create a new cluster: $j = j + 1$
      2.2.4. Initialize the new cluster: $w_j = \psi_i$

[0025]

[0026] 상기 알고리즘 1에 따르면 뉴스 비디오 샷 경계 탐지의 재현율을 높임으로써 앵커 샷 탐지 단계의 입력으로 사용되는 데이터의 오류를 최소화할 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.

[0027] 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.

[0028] 일반적으로, 뉴스 비디오는 앵커 샷과 다음 앵커 샷이 나오기까지 하나의 뉴스 기사 클립을 이룬다. 따라서, 이에 기초하여 뉴스 비디오 경계 탐지 단계에서 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할한다.

[0029] 또한, 뉴스 데이터 전처리부(110)는 분할된 뉴스 기사 클립들을 카테고리에 따라 인덱싱하여 데이터베이스(120)에 저장한다.

[0030] 이때, 뉴스 데이터 전처리부(110)는 상기 뉴스 기사 클립들의 각 텍스트 정보(textual information)와 비쥬얼 정보(visual information)를 코-클러스터링하여 뉴스 기사 클립들을 카테고리별로 분류한다.

[0031] 구체적으로, 뉴스 데이터 전처리부(110)는 코-클러스터링(Co-clustering) 알고리즘을 통해, 상기 분할된 뉴스 기사 클립에 대해 텍스트 단위로 클러스터링을 수행하고 비디오 단위로 클러스터링을 각각 수행한 후, 각 클러스터에 공통으로 포함되는 뉴스 비디오가 동일한 카테고리를 갖는 것으로 판단한다.

[0032] 이때, 본 발명의 실시예에서는 뉴스 데이터 전처리부(110)가 하기 알고리즘 2와 같은 코-클러스터링 알고리즘 [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 사용하여 뉴스 기사 클립들에 대한 카테고리별 분류를 수행하는 것을 예로써 나타내었다.

[0033] [알고리즘 2]

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, $u_2, u_3, \ldots, u_{l+1}$ and $v_2, v_3, \ldots, v_{l+1}$ 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.

[0034]

[0035] 상기 알고리즘 2에 따르면, 뉴스 데이터 전처리부(110)는 뉴스 비디오의 텍스트를 구성하는 워드 데이터 단위로 클러스터링을 수행하고, 비디오를 구성하는 영상 프레임 단위로 클러스터링을 수행한다.

[0036] 구체적으로, 뉴스 데이터 전처리부(110)는 각 뉴스 기사 클립 별 뉴스 비디오에 대해 스크립트로부터 형태소 분석을 통하여 단어(즉, 워드)들을 추출하고, 동영상 클립으로부터 대표 영상(Key-frame)을 추출한다. 그리고, 뉴스 데이터 전처리부(110)는 상기 추출된 단어 및 대표 영상을 이용하여 이분 그래프(bipartite graph) 모델을 생성하고, 상기 알고리즘 2를 이용하여 여러 개의 부분 그래프(sub-graph)로 분할한다.

[0037] 이때, 뉴스 데이터 전처리부(110)는 서로 다른 두 개의 뉴스 기사 클립이 동일한 워드 클러스터 및 프레임 클러스터에 포함되는 경우라면 동일한 카테고리 클러스터에 포함시킨다. 반면, 상기 두 뉴스 비디오가 각각 포함되

는 워드 클러스터가 상이하거나 프레임 클러스터가 상이한 경우에는 상기 두 뉴스 기사 클립을 상이한 카테고리 클러스터에 포함시킨다.

[0038] 그리고, 뉴스 데이터 전처리부(110)는 뉴스 비디오에 대하여 카테고리 클러스터링을 완료한 후, 각 뉴스 기사 클립이 속한 워드 클러스터, 프레임 클러스터 및 카테고리 클러스터에 대한 정보를 각각 인덱싱하여 데이터베이스(120)에 저장한다.

[0039] 한편, 본 발명의 실시예에 따른 뉴스 데이터 전처리부(110)는 뉴스 비디오에 대한 전처리 결과 즉, 워드 클러스터, 프레임 클러스터 및 카테고리 클러스터들에 대한 정보를 계층적 군집화 처리부(130)로 자동 전송하는 것도 가능하다.

[0040] 계층적 군집화 처리부(130)는 상기 전처리 과정에서 추출된 뉴스 기사 클립 별 텍스트 정보(즉, 단어) 및 영상 정보(즉, 대표 영상)를 이용하여 뉴스 기사 클립 간 유사도를 계산하여 유사 군집을 생성하고, 상기 생성된 유사 군집 간의 유사도를 계산하여 상위 개념의 상위 유사 군집을 생성하는 계층적 군집화를 수행한다.

[0041] 구체적으로, 계층적 군집화 처리부(130)는 데이터베이스(120)에 저장된 뉴스 기사 클립들에 대한 클러스터 정보를 획득한다. 그리고, 계층적 군집화 처리부(130)는 하기 수학식 1을 이용하여, 상기 획득한 뉴스 기사 클립 별 단어 및 대표 영상에 대해 tf-idf 함수를 적용하여 단어 별 가중치 및 대표 영상 별 가중치를 계산한다.

[0042] [수학식 1]

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

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

[0043]

[0044] 이때, 상기 수학식 1에서 $w_k(S_i)$는 뉴스 기사 클립 S_i에 속해 있는 단어 w_k의 가중치를 의미하며, $v_l(S_i)$는 뉴스 기사 클립 S_i에 속해 있는 대표 영상 v_l의 가중치를 의미한다. 그리고, tf_{w_k,s_i}는 뉴스 기사 클립 S_i에 속해 있는 단어 w_k의 빈도수이며, tf_{v_l,s_i}는 뉴스 기사 클립 S_i에 속해있는 대표 영상 v_l의 빈도수 이다. 그리고, awl은 뉴스 기사 클립들에서 나타나는 단어 수의 평균이며, $akfl$은 뉴스 기사 클립들에 나타나는 대표 영상 수의 평균이다. 그리고, sf_{wk}는 단어 w_k를 포함하고 있는 뉴스 기사 클립의 수이며, sf_{vl}는 대표 영상 v_l을 포함하고 있는 뉴스 기사 클립의 수를 의미한다. 그리고, $len(S_i)$는 뉴스 기사 클립 S_i에 속해 있는 단어 수이며, $kflen(S_i)$은 뉴스 기사 클립 S_i에 속해 있는 대표 영상의 수이다. 또한, g는 전체 뉴스 기사 클립의 수를 의미하며, α는 중복되지 않는 대표 영상의 가중치 값을 의미한다.

[0045] 그리고, 계층적 군집화 처리부(130)는 상기 수학식 1을 통해 계산된 단어별 가중치 및 대표 영상별 가중치를 하기 수학식 2에 적용하여 뉴스 기사 클립 간 유사도 행렬을 생성한다.

[0046] [수학식 2]

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

[0047]

[0048] 이때, 상기 수학식 2에서 m 은 전체 단어의 개수이며, n 은 전체 대표 영상의 개수를 의미한다.

[0049] 다음으로, 계층적 군집화 처리부(130)는 생성된 유사도 행렬에 평균 연결법(average linkage)을 적용하여 복수의 뉴스 기사 클립 중 유사도가 가장 큰 뉴스 기사 클립들을 유사 군집으로 생성하고, 상기 생성된 유사 군집 간의 비교를 통하여 상위 개념의 새로운 상위 유사 군집을 생성하는 작업을 반복하여 최종적으로 하나의 최상위 군집을 생성하는 계층적 군집화를 수행한다. 이때, 유사 군집은 실제적으로 동일한 주제(topic)의 뉴스 기사 클립들의 군집을 의미한다. 또한, 평균 연결법을 이용하여 군집 간의 거리(즉, 유사도)를 계산할 때, 한 군집의 모든 구성원들과 다른 군집의 모든 구성원들 간의 거리의 평균을 군집 간 유사도로써 계산한다.

[0050] 이때, 계층적 군집화 처리부(130)는 하기 수학식 3을 이용하여 뉴스 기사의 유사 군집 간의 유사도를 계산한다.

[0051] [수학식 3]

$$S_{(U,V)W} = \frac{\Sigma\Sigma s_{ik}}{N_{(U,V)} N_W}$$

[0052]

[0053] 상기 수학식 3에서 s_{ik} 는 군집 (U , V)의 i 번째 개체와 군집 W의 k번째 개체 간의 유사도를 의미한다. 그리고, $N_{(U,V)}$ 는 군집 (U , V)의 개체 수이고, N_W 는 군집 W의 개체 수이다.

[0054] 그리고, 계층적 군집화 처리부(130)는 수학식 3을 통해 계산된 뉴스 기사 군집 간의 유사도가 기설정된 임계값 이상인 군집들을 상위 유사 군집으로 생성하고, 상기 상위 유사 군집과 또 다른 유사 군집의 유사도를 계산하는 과정을 반복하여 새로운 상위 유사 군집을 생성한다. 이때, 계층적 군집화 처리부(130)는 상위 유사 군집을 생성하는 과정을 반복한 결과 하나의 최상위 군집이 생성될 때까지 뉴스 기사 클립을 계층적 군집화한다.

[0055] 예를 들어, 도 2에서는 본 발명의 실시예에 따른 계층적 군집화 처리부(130)가 데이터베이스(120)로부터 임의의 뉴스 비디오의 복수의 카테고리 클러스터 중 임의의 카테고리 클러스터를 획득하고, 상기 임의의 카테고리 클러스터에 포함된 복수의 뉴스 기사 클립들에 대해 계층적 군집화 처리를 수행한 것을 예로써 나타내었다.

[0056] 도 2에서는, 계층적 군집화 처리부(130)가 임의의 카테고리 클러스터에 대해 계층적 군집화를 수행한 결과, '남북 정상 회담' 이라는 최상위 군집(Most Significant Cluster, 이하 'MSC' 라고 함)이 생성될 때까지 3 계층의 군집화가 수행된 것을 나타내었다.

[0057] 이때, 도 2에서는 상기 임의의 카테고리 클러스터에 포함된 복수의 뉴스 기사 클립들이 19개의 유사 군집(C1 ~ C19)으로 군집화되었을 때, 상기 19개의 유사 군집들이 최하위 군집(제3 계층)으로 정의되는 것을 나타내었다. 그리고, 최하위 군집들 간의 유사도를 계산한 결과 7개의 상위 군집(Superordinate Cluster, 이하 'SC' 라고 함)들이 생성된다. 또한, 7개의 SC들(SC 1 ~ SC 7) 간의 유사도를 계산한 결과 SC 4, SC 5, SC 6, SC 7에 대해 하나의 상위 군집 SC 8이 생성되고, 나머지 SC들(SC 1 ~ SC 4)과 SC 8 간의 유사도를 계산한 결과 최상위 군집(Most Significant Cluster, 이하 'MSC' 라고 함)이 생성된다. 참고로, 상기 최하위 군집에는 각각 적어도 하나의 뉴스 기사 클립이 포함된다.

[0058] 그리고, 계층적 군집화 처리부(130)는 뉴스 기사 클립들에 대한 계층적 군집화 처리를 수행한 결과를 데이터베이스(120)에 인덱싱하여 저장한다.

[0059] 참고로, 본 발명의 실시예에 따른 계층적 군집화 과정에서 최하위 군집들이 복수의 상위 군집으로 군집화되는 기준 즉, 주제는 데이터베이스(120) 또는 계층적 군집화 처리부(130)의 자체 저장 공간(미도시)에 기설정되어 있다.

[0060] 다시 도 1을 참조하면, 군집 분석 처리부(140)는 상기 계층적 군집화된 뉴스 기사 클립들에 대해 지적 구조 분석을 수행한다.

[0061] 이때, 군집 분석 처리부(140)는 군집 분석 처리 기준에 따라 뉴스 기사의 계층적 군집들에 대해 기간별 군집 성향, 군집 성장도, 뉴스 기사 군집 간 연관도 등을 분석하는 지적 구조 분석을 수행하여 지적 정보를 생성한다. 참고로, 상기 군집 분석 처리 기준은 뉴스 비디오의 카테고리별로 자동 설정되어 있거나, 외부 (즉, 뉴스 기사 및 지적 정보를 검색하고자 하는 사용자)로부터 입력되는 정보에 기초하여 설정될 수 있다.

[0062] 이와 같은, 군집 분석 처리부(140)가 뉴스 비디오에 대한 지적 구조 분석을 수행하는 과정에 대해서는 이하 도 3 내지 도 6을 참조하여 상세히 설명하도록 한다.

[0063] 다시 도 1을 참조하면, 데이터베이스(120)는 뉴스 비디오에 대한 전처리 결과로써 생성된 뉴스 기사 클립 별 클러스터 정보와, 상기 뉴스 기사 클립 별 클러스터 정보를 이용하여 계층적 군집화를 수행하여 생성된 적어도 하나의 계층의 뉴스 기사 군집들의 정보 및 상기 계층적 군집화된 뉴스 기사 군집들에 대해 지적 구조 분석을 수행한 결과 생성된 지적 정보들 즉, 뉴스 데이터 정보들을 인덱싱하여 저장한다.

[0064] 분석 결과 제공부(150)는 데이터베이스(120)에 저장된 뉴스 데이터 정보들을 이용하여, 계층적 군집화된 뉴스 기사들에 대한 지적 구조 분석 결과를 식별 가능한 정보로써 출력한다. 이때, 분석 결과 제공부(150)는 복수의 뉴스 비디오들을 뉴스 기사 클립 단위로 분할한 전처리 결과 정보, 뉴스 기사 클립들을 계층적 군집화한 결과 정보 및 계층적 군집화된 뉴스 기사 군집들을 지적 구조 분석한 결과 정보 중 적어도 하나의 정보를 사용자가 식별 가능한 정보로 생성하여 출력한다. 참고로, 분석 결과 제공부(150)는 웹 페이지 형식의 뉴스 데이터 및 문자 또는 도표 데이터로 작성된 뉴스 데이터 인쇄물 등을 출력할 수 있다.

[0065] 도 3은 본 발명의 실시예에 따른 군집 분석 처리부의 구성을 나타내는 블록도이다.

[0066] 그리고, 도 4는 본 발명의 실시예에 따른 기간별 군집 성향 분석 과정을 설명하기 위한 도면이고, 도 5는 본 발명의 실시예에 따른 군집 성장도 분석 과정을 설명하기 위한 도면이다.

[0067] 또한, 도 6은 본 발명의 실시예에 따른 패스파인더 네트워크 분석 과정을 설명하기 위한 도면이다.

[0068] 도 3에서와 같이, 군집 분석 처리부(140)는 기간별 군집 성향 분석 모듈(141), 군집 성장도 분석 모듈(142) 및 패스파인더 네트워크 분석 모듈(143)을 포함한다.

[0069] 이때, 군집 분석 처리부(140)는 데이터베이스(120)로부터 계층적 군집화된 뉴스 기사 군집들을 획득하고, 군집 분석 처리 기준에 기초하여 상기 뉴스 기사 군집들에 대한 지적 구조 분석을 수행한다.

[0070] 본 발명의 실시예에 따른 군집 분석 처리부(140)는 군집 분석하고자 하는 뉴스 기사의 주제(topic) 또는 카테고리, 뉴스 비디오 방송 시기 또는 기간, 및 처리할 군집 분석 종류 등을 포함하는 군집 분석 처리 기준이 기설정되어 있거나, 외부로부터 상기 군집 분석 처리 기준들을 입력 받을 수 있다.

[0071] 구체적으로, 군집 분석 처리부(140)는 데이터베이스(120)로부터 상기 뉴스 기사의 주제 또는 카테고리 기준에 상응하는 카테고리 클러스터 중 상기 뉴스 비디오 방송 시기 기준에 상응하는 뉴스 기사 클립들에 대한 계층적 군집화 결과를 획득한다.

[0072] 그리고, 군집 분석 처리부(140)는 상기 획득한 계층적 군집화 결과에 대해 상기 군집 분석 종류 기준에 따른 지적 구조 분석을 수행한다. 이때, 군집 분석 처리부(140)는 상기 군집 분석 종류 기준이 특별하게 설정되지 않은 경우 기간별 군집 성향, 군집 성장도 및 패스파인더 네트워크 분석을 모두 수행한다.

[0073] 구체적으로, 기간별 군집 성향 분석 모듈(141)은 상기 뉴스 비디오 방송 시기 또는 기간 기준에 따른 일정 기간 동안 상기 뉴스 기사의 주제 또는 카테고리 기준에 포함되는 뉴스 기사 군집들이 방송된 성향을 분석한다. 이때, 기간별 군집 성향 분석 모듈(141)은 설정된 방송 시기 또는 기간 동안 뉴스 기사 군집 별로 방송된 횟수 및 각 뉴스 기사 군집들이 방송된 시기 등의 성향 정보를 분석할 수 있다.

[0074] 예를 들어, 도 4에서는 기간별 군집 성향 분석 모듈(141)이 설정된 뉴스 비디오 방송 시기 기준에 상응하는 기간 동안 도 2에서 나타낸 '남북 정상 회담'이라는 주제에 대한 19개의 뉴스 기사 군집(C1 ~ C19)들이 방송된 성향을 분석하여 기간별 군집 성향 그래프를 생성하는 것을 나타내었다.

[0075] 도 4에서는, 8월 8일부터 8월 10일까지의 방송 기간과 10월 1일부터 10월 5일까지의 방송 기간 동안 상기 선택된 주제에 대한 뉴스 기사 군집(C1 ~ C19)들 중 해당 일에 방송된 뉴스 기사 군집의 종류와 방송 빈도 등을 그

래프로 나타내었다. 참고로, 본 발명의 실시예에 따른 기간별 군집 성향 분석 모듈(141)은 동일 방송 시기에 대한 복수의 뉴스 비디오들(예를 들어, 복수의 방송사에서 방송된 뉴스 비디오들)에 대해 계층적 군집화 결과로부터 상기 기간별 군집 성향을 분석할 수 있다.

[0076] 군집 성장도 분석 모듈(142)은 상기 뉴스 기사 방송 시기 또는 기간 기준에 상응하는 일정 기간 동안 상기 뉴스 기사의 주제 또는 카테고리 기준에 상응하는 뉴스 기사 군집들의 성장도를 분석한다.

[0077] 구체적으로, 군집 성장도 분석 모듈(142)은 상기 뉴스 기사 방송 시기 또는 기간 기준에 따른 일정 기간을 복수의 기간으로 분할하고, 상기 분할된 각 기간 별로 임의의 뉴스 기사 군집의 개수를 이용하여 군집 성장 지수(CGI: Cluster Growth Index)를 계산한다.

[0078] 참고로, 상기 뉴스 기사의 방송 시기 또는 기간을 연속하는 2개의 기간으로 분할한 후 상대적으로 이전 기간을 1기라고 하고 이후 기간을 2기라고 지칭할 때, 상기 군집 성장 지수는 하기 수학식 4를 통해 계산할 수 있다.

[0079] [수학식 4]

[0080] $$CGI = \frac{\text{2기 뉴스 기사 수} - \text{1기 뉴스 기사 수}}{\text{2기 뉴스 기사 수} + \text{1기 뉴스 기사 수}}$$

[0081] 이때, 군집 성장 지수는 1기에 비해 2기에 해당 뉴스 기사 군집의 뉴스 기사 수가 상대적으로 성장한 지표를 나타낸다.

[0082] 또한, 군집 성장도 분석 모듈(142)은 각 뉴스 기사 군집의 절대적 성장 규모를 산출하기 위해 하기 수학식 5를 통해 가중 군집 성장 지수(WCGI: Weighted Cluster Growth Index)를 산출한다.

[0083] [수학식 5]

[0084] $$WCGI = |\text{2기 뉴스 기사 수} - \text{1기 뉴스 기사 수}| \times CGI$$

[0085] 이와 같이, 수학식 5를 통해 산출된 가중 군집 성장 지수는 각 뉴스 기사 군집의 절대적인 뉴스 기사 수의 증가를 반영하는 지표로써, 증가한 뉴스 기사 수에 비례하여 값이 커진다.

[0086] 예를 들어, 도 5에서는 군집 성장도 분석 모듈(142)이 도 4에서 분석 처리 기준으로 설정된 방송 기간을 8월과 10월로 구분하여 2개의 기간으로 분할하고, 도 2에서 나타낸 '남북 정상 회담' 이라는 주제에 대한 19개의 뉴스 기사 군집들의 성장도를 분석하여 군집 성장도 그래프를 생성하는 것을 나타내었다.

[0087] 도 5에서와 같이, 군집 성장도 그래프를 통해 8월에 비해 10월에 방송된 뉴스 기사의 수가 증가한 뉴스 기사 군집들의 플러스 성장도와 반면 10월에 뉴스 기사 수가 감소한 뉴스 기사 군집들의 마이너스 성장도를 확인할 수 있다.

[0088] 패스파인더 네트워크 분석 모듈(143)은 상기 뉴스 기사 방송 시기 또는 기간 기준에 상응하는 일정 기간 동안 상기 뉴스 기사의 주제 또는 카테고리 기준에 상응하는 뉴스 기사 군집들의 연관도를 분석한다.

[0089] 구체적으로, 패스파인더 네트워크 분석 모듈(143)은 뉴스 기사 군집들에 대해 가중치가 있는 모든 링크를 생성하고, 각 링크의 경로 중 삼각 부등식(triangle inequality)을 위반하는 경로를 제거하는 패스파인더 네트워크 분석을 수행하여, 뉴스 기사 군집 간 연관도를 명확하게 식별할 수 있는 네트워크 모델을 생성한다.

[0090] 이때, 패스파인더 네트워크 분석 모듈(143)은 하기 수학식 6을 통해 각 뉴스 기사 군집들 간의 거리, 즉 연관도를 산출한다.

[0091] [수학식 6]

[0092] $$W(n_i, n_J) \le \left(\sum_{l=1}^{k-1} W^r(n_l, n_{l+1})\right)^{1/r} \qquad \forall k = 2,3,\ldots,q$$

[0093] 수학식 6에서 파라미터 q는 n개의 노드에 대해 각 경로 거리를 산출하는데 고려되는 최대 링크의 수를 의미하며, 2 ~ (n-1)의 값을 갖는다. 또한, 파라미터 r은 민코프스키(Minkowski) 거리 공식의 제곱수로서, 두 노드 n_i와 n_j 사이의 특정 경로를 구성하는 복수의 링크들이 가지고 있는 가중치를 경로 거리 $W(n_i, n_j)$에 반영

한 것이다. 이때, 파라미터 r의 값이 1이면 각 링크 가중치의 합이 경로 거리로써 산출되며, 파라미터 r의 값이 무한대이면 경로를 구성하는 링크의 가중치 중 최대 값이 경로 거리로 산출된다. 따라서, 파라미터 r의 값이 커질수록 경로의 거리가 짧아지게 되어 남는 링크의 수가 줄어들게 된다.

[0094] 예를 들어, 도 6에서는 패스파인더 네트워크 분석 모듈(143)이 상기 설정된 뉴스 비디오 방송 시기 또는 기간 기준에 상응하는 일정 기간 동안 도 2에서 나타낸 '남북 정상 회담' 이라는 주제에 대한 19개의 뉴스 기사 군집(C1 ~ C19)들에 대해 패스파인더 네트워크 분석을 처리한 결과를 예로써 나타내었다.

[0095] 참고로, 패스파인더 네트워크 모델에서는 군집 개체의 수가 많이 포함된 군집일수록 원의 지름이 크게 표현되며, 각 원의 링크들의 굵기에 따라 가중치가 표현된다.

[0096] 이때, 패스파인더 네트워크 분석 모듈(143)의 분석 처리 결과 도 6에서와 같이, 상기 설정된 뉴스 비디오 방송 시기 또는 기간 기준 내에서는 '회담일정및배경' 의 주제에 따른 뉴스 기사 군집(C12)에 포함되는 뉴스 기사 클립의 수가 가장 많은 것을 알 수 있으며, 각 뉴스 기사 군집 간에 생성된 링크들을 통해 뉴스 기사 군집 간의 연관도를 편리하게 확인할 수 있다.

[0097] 이하, 도 7 및 도 8을 참조하여 본 발명의 실시예에 따른 뉴스 데이터 분석 방법에 대해서 상세히 설명하도록 한다.

[0098] 도 7은 본 발명의 실시예에 따른 뉴스 데이터 분석 방법을 설명하기 위한 순서도이다.

[0099] 그리고, 도 8은 본 발명의 실시예에 따른 군집 분석 처리 과정을 설명하기 위한 순서도이다.

[0100] 도 7에서와 같이, 뉴스 데이터 분석 방법에서는 먼저 적어도 하나의 뉴스 비디오에 대해서 뉴스 데이터 전처리를 수행한다(S710). 이때, 뉴스 데이터 전처리 과정에서 뉴스 비디오를 복수의 뉴스 기사 클립으로 분할하고, 각 뉴스 기사 클립들의 단어 및 대표 영상을 추출하여 뉴스 기사 클립들에 대한 워드, 프레임 및 카테고리 클러스터를 생성한다.

[0101] 그리고, 뉴스 데이터 전처리 결과에 대한 계층적 군집화 처리를 수행한다(S720).

[0102] 이때, 계층적 군집화 처리 단계에서는 동일 카테고리 클러스터에 포함되는 복수의 뉴스 기사 클립들에 대해서 유사도 행렬을 생성하고, 유사도 행렬에 평균 연결법을 적용하여 복수의 뉴스 기사 클립들 간의 유사도를 산출한다. 그리고, 산출된 유사도에 따라 적어도 하나의 뉴스 기사 클립을 포함하는 복수의 뉴스 기사 군집을 생성한다. 또한, 계층적 군집화 처리 단계에서는 생성된 뉴스 기사 군집들 간의 유사도를 산출하여 상위 유사 군집을 생성하며, 하나의 최상위 군집이 생성될 때까지 상기 유사 군집 생성을 반복한다.

[0103] 다음으로, 계층적 군집화 처리를 통해 생성된 뉴스 기사 군집들에 대한 지적 구조 분석 처리를 수행한다(S730).

[0104] 이때, 지적 구조 분석 처리 단계에서는 설정된 군집 분석 처리 기준에 상응하여 뉴스 기사 군집들에 대한 기간별 군집 성향 분석, 군집 성장도 분석 및 패스파인더 네트워크 분석 등의 지적 구조 분석 처리를 수행한다.

[0105] 구체적으로, 단계 S730에서는 도 8에서와 같은 군집 분석 처리를 수행한다.

[0106] 도 8에서와 같이, 본 발명의 실시예에 따른 군집 분석 처리 방법에서는, 먼저 군집 분석 처리 기준을 확인한다(S810).

[0107] 이때, 군집 분석 처리 기준은 뉴스 비디오 방송 시기 또는 기간, 뉴스 기사의 주제 또는 카테고리, 및 처리할 군집 분석 종류 등의 기준을 포함한다. 이러한 군집 분석 처리 기준은 기설정되어 있던 뉴스 기사 주제 및 카테고리에 해당하는 뉴스 기사에 대한 계층적 군집화 처리 시 자동으로 적용되거나, 외부(예를 들어, 뉴스 기사에 대한 군집 분석 처리를 요청하는 사용자 등)로부터 기준 값들을 입력 받아 설정할 수 있다.

[0108] 다음으로, 군집 분석 처리 기준에 따라 지적 구조 분석을 수행한다(S820).

[0109] 본 발명의 실시예에 따른 군집 분석 처리 방법에서는 군집 분석 처리 기준에 군집 분석 종류가 포함된 경우 해당 군집 분석을 수행하고, 군집 분석 종류가 포함되지 않는 경우 임의의 군집 분석을 수행하거나 모든 군집 분석을 임의의 순서에 따라 수행하는 것이 가능하다.

[0110] 이때, 단계 S820은 뉴스 기사 군집들에 대한 기간별 군집 성향 분석을 수행하는 단계(S821), 군집 성장도 분석을 수행하는 단계(S822) 및 패스파인더 네트워크 분석을 수행하는 단계(S823)를 포함한다. 이와 같은, 각 군집 분석 단계들은 별도의 모듈을 통해 동시에 수행되거나, 임의의 순서에 따라 수행될 수 있다.

[0111] 단계 S820을 수행한 후, 지적 구조 분석하여 생성된 새로운 지적 정보들을 인덱싱하여 저장한다(S830).

[0112] 참고로, 상기 새로운 지적 정보들은 뉴스 기사 그룹들에 대한 문자 또는 그래프 정보들을 포함한다.

[0113] 다시 도 7을 참조하면, 뉴스 기사 군집들에 대한 지적 구조 분석 처리를 수행한 후, 군집 분석 결과로써 생성된 새로운 지적 정보들을 저장 및 제공한다(S740).

[0114] 구체적으로, 새로운 지적 정보들은 설정된 기간 내에 방송된 뉴스 비디오들에 포함된 뉴스 기사들을 설정된 주제 또는 카테고리에 따라, 적어도 하나의 계층의 뉴스 기사 군집으로 계층적 군집화한 결과, 상기 뉴스 기사 군집들이 상기 설정된 기간 내에 실제 방송된 성향 분석 결과, 상기 뉴스 기사 군집들의 상기 설정된 기간 동안의 성장도 분석 결과 및 상기 뉴스 기사 군집들 간의 연관도 분석 결과 등을 포함한다. 그리고, 새로운 지적 정보들은 웹 페이지 형식으로 생성되어 사용자 단말기(미도시) 또는 뉴스 데이터 분석 처리 장치에 자체적으로 구비되는 모니터(미도시) 등의 화면에 출력될 수 있으며, 문자 또는 그래프 정보로 생성되어 프린터(미도시) 등의 출력 장치를 통해 인쇄될 수 있다.

[0115] 본 발명의 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 판독가능 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.

[0116] 본 발명의 장치 및 방법은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 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.

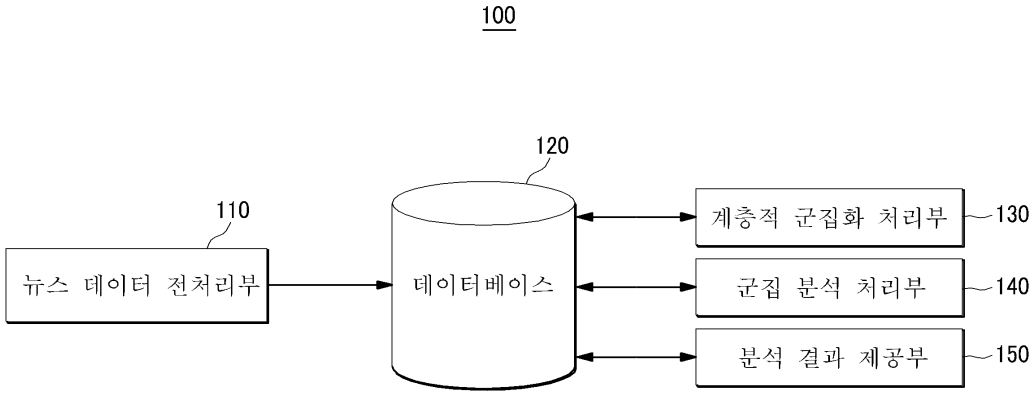
부호의 설명

[0119]

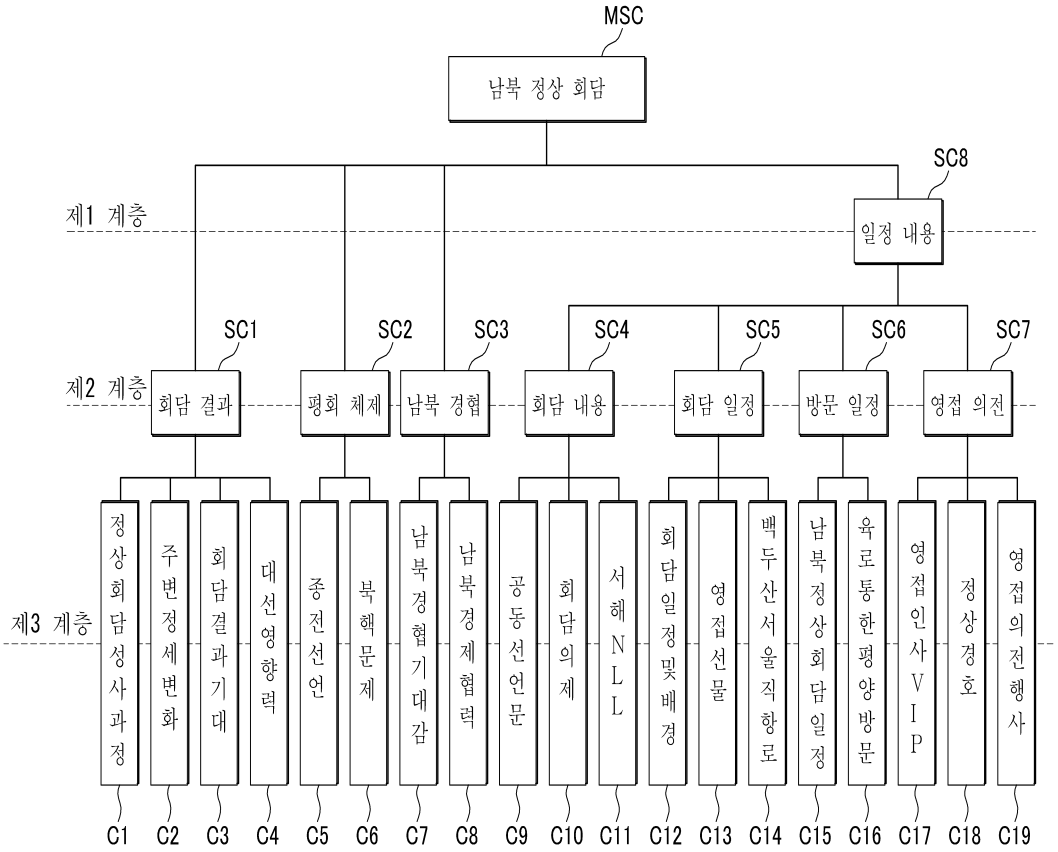
110: 뉴스 데이터 전처리부	120: 데이터베이스
130: 계층적 군집화 처리부	140: 군집 분석 처리부
150: 분석 결과 제공부	141: 기간별 군집 성향 분석 모듈
142: 군집 성장도 분석 모듈	143: 패스파인더 네트워크 분석 모듈

도면

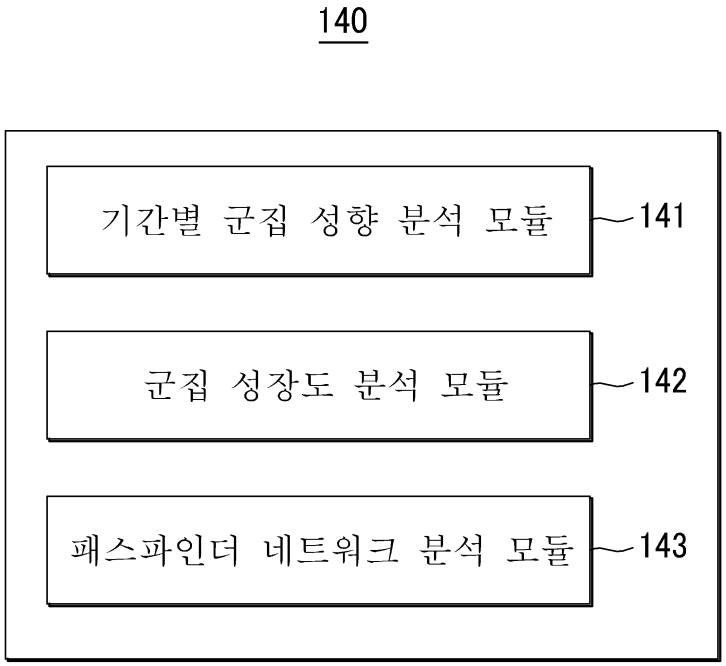
도면1


100
120
110
뉴스 데이터 전처리부
데이터베이스
계층적 군집화 처리부
130
군집 분석 처리부
140
분석 결과 제공부
150

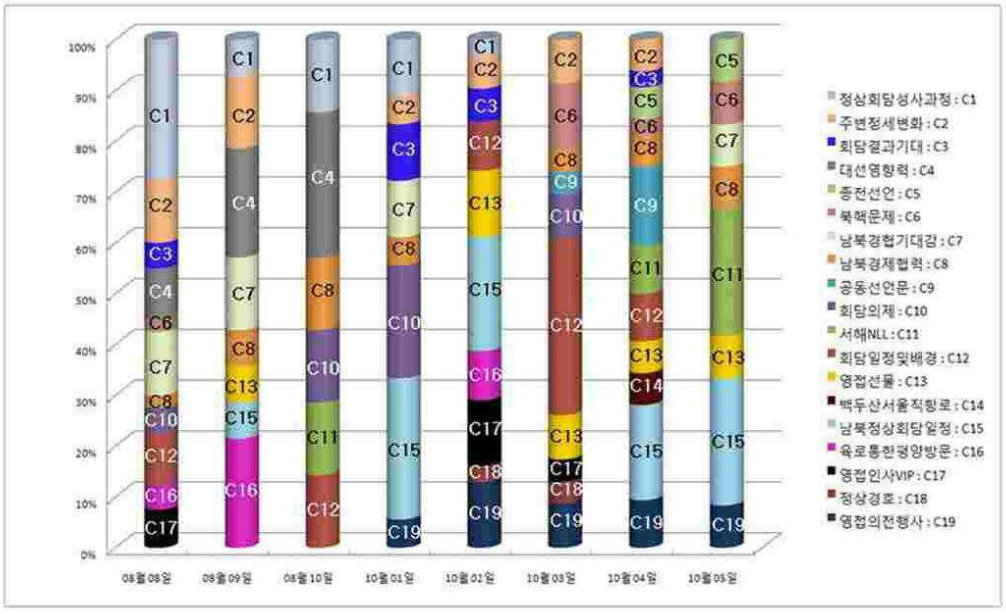
도면2


MSC
남북 정상 회담
SC8
제1 계층
일정 내용
SC1
SC2
SC3
SC4
SC5
SC6
SC7
제2 계층
회담 결과
평화 체제
남북 경협
회담 내용
회담 일정
방문 일정
영접 의전
제3 계층
정상회담성사과정
주변정세변화
회담결과기대
대선영향력
종전선언
북핵문제
남북경협기대감
남북경제협력
공동선언문
회담의제
서해NLL
회담일정및배경
영접선물
백두산서울직항로
남북정상회담일정
육로통한평양방문
영접인사VIP
정상경호
영접의전행사
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
C11
C12
C13
C14
C15
C16
C17
C18
C19

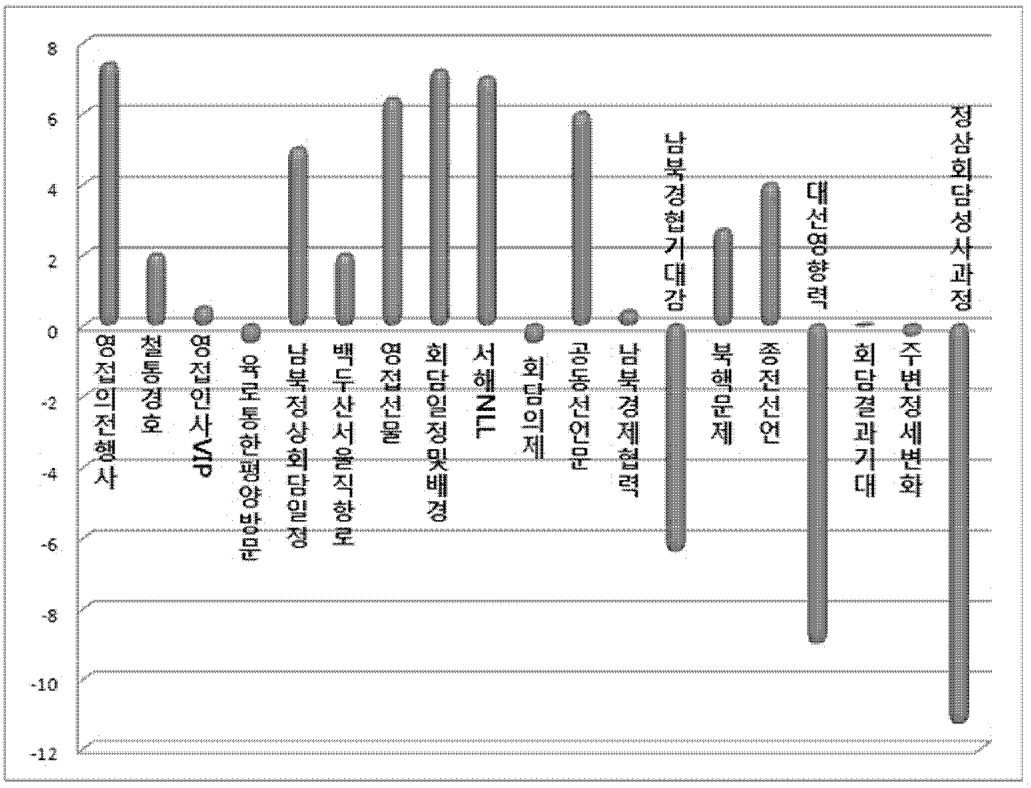
도면3


140
기간별 군집 성향 분석 모듈
141
군집 성장도 분석 모듈
142
패스파인더 네트워크 분석 모듈
143

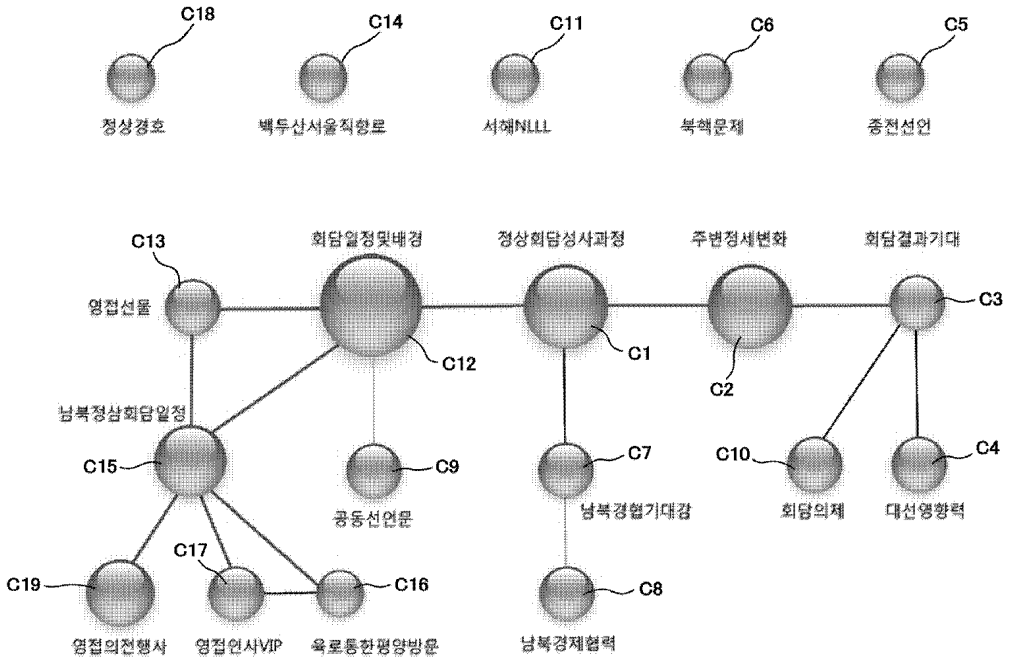
도면4


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
08월08일
08월09일
08월10일
10월01일
10월02일
10월03일
10월04일
10월05일
정상회담성사과정 : C1
주변정세변화 : C2
회담결과기대 : C3
대선영향력 : C4
종전선언 : C5
북핵문제 : C6
남북경협기대감 : C7
남북경제협력 : C8
공동선언문 : C9
회담의제 : C10
서해NLL : C11
회담일정및배경 : C12
영접선물 : C13
백두산서울직항로 : C14
남북정상회담일정 : C15
육로통한평양방문 : C16
영접인사VIP : C17
정상경호 : C18
영접의전행사 : C19

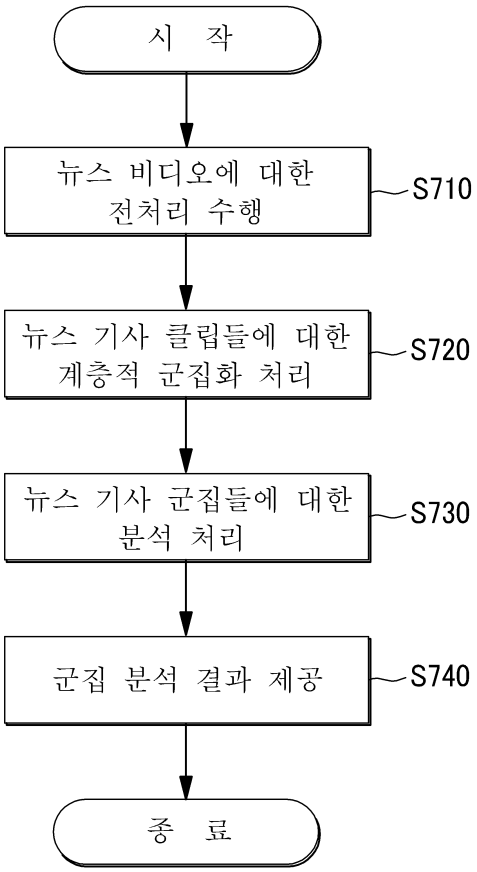
도면5


8
6
4
2
0
-2
-4
-6
-8
-10
-12
영접의전행사
철통경호
영접인사VIP
육로통한평양방문
남북정상회담일정
백두산서울직항로
영접선물
회담일정및배경
서해NLL
회담의제
공동선언문
남북경제협력
남북경협기대감
북핵문제
종전선언
대선영향력
회담결과기대
주변정세변화
정상회담성사과정

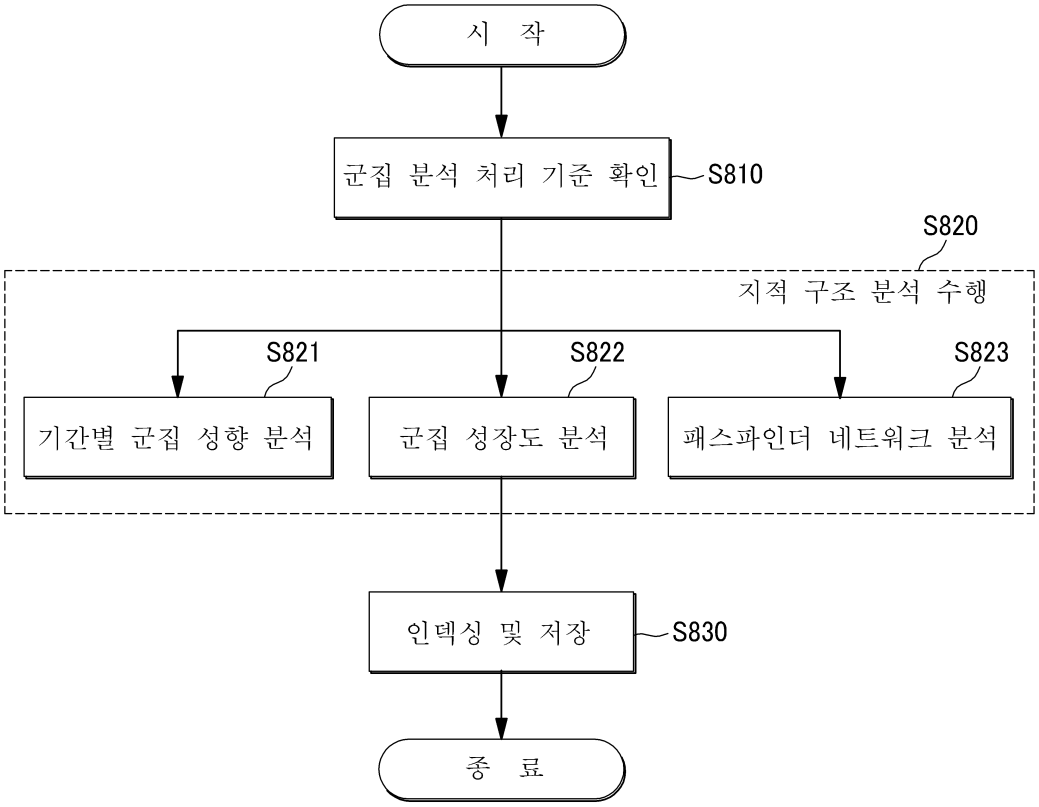
도면6


C18
정상경호
C14
백두산서울직항로
C11
서해NLLL
C6
북핵문제
C5
종전선언
C13
영접선물
회담일정및배경
정상회담성사과정
주변정세변화
회담결과기대
C3
C12
C1
C2
남북정상회담일정
C15
C9
C7
C10
C4
공동선언문
남북경협기대감
회담의제
대선영향력
C17
C19
C16
C8
영접의전행사
영접인사VIP
육로통한평양방문
남북경제협력

도면7


시 작
뉴스 비디오에 대한 전처리 수행
S710
뉴스 기사 클립들에 대한 계층적 군집화 처리
S720
뉴스 기사 군집들에 대한 분석 처리
S730
군집 분석 결과 제공
S740
종 료

도면8


시 작
군집 분석 처리 기준 확인
S810
S820
지적 구조 분석 수행
S821
S822
S823
기간별 군집 성향 분석
군집 성장도 분석
패스파인더 네트워크 분석
인덱싱 및 저장
S830
종 료