(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년06월02일
(11) 등록번호 10-2260287
(24) 등록일자 2021년05월28일

(51) 국제특허분류(Int. Cl.)
G06Q 50/10 (2012.01) *G06F 16/904* (2019.01)
G06F 40/20 (2020.01)
(52) CPC특허분류
G06Q 50/10 (2013.01)
G06F 16/904 (2019.01)
(21) 출원번호 10-2019-0156828
(22) 출원일자 2019년11월29일
심사청구일자 2019년11월29일
(56) 선행기술조사문헌
KR1020110043369 A*
KR1020180071021 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
조선대학교산학협력단
광주광역시 동구 필문대로 309 (서석동)
(72) 발명자
김판구
광주광역시 북구 설죽로 600, 101동 904호 (일곡동, 삼호아파트)
홍택은
전라남도 장흥군 용산면 용산대로 2212
(뒷면에 계속)
(74) 대리인
특허법인아이엠

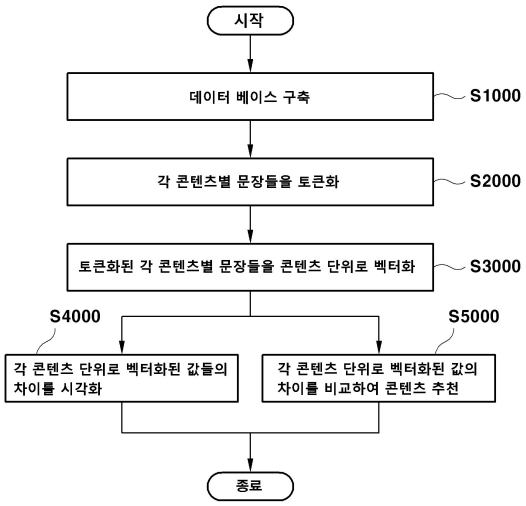
전체 청구항 수 : 총 6 항 심사관 : 고재용

(54) 발명의 명칭 **스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법.**

(57) 요 약

본 발명은 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법에 관한 것으로, 상세하게는 창작자가 스토리텔링에 있어서 다시쓰기를 하고자하는 콘텐츠와 유사 콘텐츠를 시각화된 그래프를 통해 확인할 수 있고, 더 나아가 유사 콘텐츠들의 리스트를 추천할 수 있는 방법에 관한 것이다.

대 표 도 - 도1



(52) CPC특허분류
G06F 40/205 (2020.01)
(72) 발명자
김형주
광주광역시 동구 중앙로 160번길 16-33
유경호
광주광역시 풍암2로 62, 201-1802
노주현
광주광역시 북구 동문대로86번길 55, 202동 108호 (두암동, 두암동현대2차아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711081128
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW전문인력역량강화(R&D)
연구과제명	SW중심대학(조선대)
기 여 율	1/1
과제수행기관명	조선대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

창작물, 전설, 설화 또는 민담과 관련된 콘텐츠를 수집하여 데이터베이스를 구축하는 단계;

상기 데이터베이스에 저장된 각 콘텐츠별 문장들을 토큰화하는 단계;

토큰화된 각 콘텐츠별 문장들을 콘텐츠 단위로 임베딩하여 벡터화하는 단계; 및

상기 콘텐츠 단위로 벡터화된 값(이하, "콘텐츠 단위 벡터"이라함)들을 시각화하는 단계;를 포함하고,

상기 시각화하는 단계:는

각 콘텐츠 단위 벡터들을 2차원으로 차원 축소하는 단계;

차원 축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 또는 코사인 유사도를 계산하는 단계; 및

계산된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 또는 코사인 유사도를 Confusion Matrix를 이용하여 시각화하는 단계;를 포함하는 것을 특징으로 하는 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법.

청구항 2

제 1 항에 있어서,

상기 벡터화하는 단계;는 PV-DM 알고리즘을 이용하여 각 콘텐츠별 문장들을 벡터화하는 것을 특징으로 하는 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법.

청구항 3

제 1 항에 있어서,

상기 벡터화하는 단계; 이후에,

각 콘텐츠 단위 벡터들을 비교하여 콘텐츠를 추천하는 단계;를 더 포함하는 것을 특징으로 하는 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법.

청구항 4

제 3 항에 있어서,

상기 콘텐츠를 추천하는 단계:는

각 콘텐츠 단위 벡터들을 2차원으로 차원 축소하는 단계;

차원 축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리와 코사인 유사도를 계산하는 단계; 및

각 콘텐츠 단위 벡터 간에 계산된 유클리디안 거리가 임계 거리 이하이고, 계산된 코사인 유사도가 임계 유사도 이상에 해당되는 콘텐츠를 추천하는 단계;를 포함하는 것을 특징으로 하는 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법.

청구항 5

제 4 항에 있어서,

상기 차원 축소하는 단계;는 T-SNE 알고리즘 이용하여 각 콘텐츠 단위 벡터들을 2차원으로 차원 축소하는 것을 특징으로 하는 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법.

청구항 6

컴퓨터와 결합하여 제 1 항 내지 제 5 항 중 어느 한 항의 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방

법을 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램.

청구항 7

삭제

청구항 8

삭제

발명의 설명

기 술 분 야

[0001] 본 발명은 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법에 관한 것으로, 상세하게는 창작자가 스토리텔링에 있어서 다시쓰기를 하고자하는 콘텐츠와 유사 콘텐츠를 시각화된 그래프를 통해 확인할 수 있고, 더 나아가 유사 콘텐츠들의 리스트를 추천할 수 있는 방법에 관한 것이다.

배 경 기 술

[0002] 스토리텔링이란 스토리(Story)와 텔링(Telling) 의 합성어로서 말 그대로 "이야기하다 라는 의미를 지니며, 상대방에게 알리고자 하는 바를 재미있고 생생한 이야기로 설들력 있게 전달하는 행위이다.

[0003] 이러한, 스토리텔링은 문학, 영화, 교육학 등에 사용되고 있으며, 알리고자 하는 바를 단어, 이미지, 소리를 통해 사건, 이야기를 전달하는 것이다.

[0004] 한편, 스토리텔링과 같은 저작 작업 중 "다시쓰기"는 기존에 존재하는 이야기를 패러디, 모방, 반복 등을 이용하여 각색하는 것으로, 스토리텔링에서 "다시쓰기'로 다양한 콘텐츠가 생산되고 있다.

[0005] 그러나, "다시쓰기" 작업을 수행하기 위해서는 다시 쓰고자 하는 콘텐츠와 연관성이 있는 유사 콘텐츠로부터 정보를 수집해야 하는데, 종래에는 많은 시간과 노력을 소모하여, 인터넷과 서적 등을 통해 유사 콘텐츠를 수작업으로 일일이 찾아야하는 문제가 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) KR10-1706167 B1 "콘텐츠 추천을 위한 장치 및 방법"

(특허문헌 0002) KR10-1559459 B1 "콘텐츠 추천 방법"

발명의 내용

해결하려는 과제

[0007] 본 발명은 상술한 문제점을 해결하기 위해 안출된 것으로, 본 발명의 목적은 저작자가 스토리텔링의 저작 작업 중 다시쓰기를 수행할 때 다시 쓰고자 하는 콘텐츠와 유사한 콘텐츠를 저작자에게 추천하여 이를 통해 창의적이고 자유로운 저작 활동에 도움을 줄 수 있는 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법을 제공하는 데 있다.

과제의 해결 수단

[0008] 상술한 목적을 해결하기 위해 본 발명은 창작물, 전설, 설화 또는 민담과 관련된 콘텐츠를 수집하여 데이터베이스를 구축하는 단계; 상기 데이터베이스에 저장된 각 콘텐츠별 문장들을 토큰화하는 단계; 토큰화된 각 콘텐츠별 문장들을 콘텐츠 단위로 임베딩하여 벡터화하는 단계; 및 상기 콘텐츠 단위로 벡터화된 값(이하, "콘텐츠 단위 벡터"이라함)들을 시각화하는 단계;를 포함하는 것을 특징으로 하는 스토리텔링 작업 중 다시쓰기를 위한 콘

텐츠 추천 방법을 제공한다.

[0009] 바람직한 실시예에 있어서, 상기 벡터화하는 단계;는 PV-DM 알고리즘을 이용하여 각 콘텐츠별 문장들을 벡터화할 수 있다.

[0010] 바람직한 실시예에 있어서, 상기 시각화하는 단계:는 각 콘텐츠 단위 벡터들을 2차원으로 차원 축소하는 단계; 차원 축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 또는 코사인 유사도를 계산하는 단계; 및 계산된 유클리디안 거리 또는 코사인 유사도를 색의 밝기 차로 표현하여 시각화하는 단계;를 포함한다.

[0011] 바람직한 실시예에 있어서, 상기 색의 밝기 차로 표현하여 시각화하는 단계;는 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 또는 코사인 유사도를 Confusion Matrix를 이용하여 시각화할 수 있다.

[0012] 바람직한 실시예에 있어서, 상기 벡터화하는 단계; 이후에, 각 콘텐츠 단위 벡터들을 비교하여 콘텐츠를 추천하는 단계;를 더 포함할 수 있다.

[0013] 바람직한 실시예에 있어서, 상기 콘텐츠를 추천하는 단계:는 각 콘텐츠 단위 벡터들을 2차원으로 차원 축소하는 단계; 차원 축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리와 코사인 유사도를 계산하는 단계; 및 각 콘텐츠 단위 벡터 간에 계산된 유클리디안 거리가 임계 거리 이하이고, 계산된 코사인 유사도가 임계 유사도 이상에 해당되는 콘텐츠를 추천하는 단계;를 포함한다.

[0014] 바람직한 실시예에 있어서 상기 차원 축소하는 단계;는 T-SNE 알고리즘 이용하여 각 콘텐츠 단위 벡터들을 2차원으로 차원 축소한다.

[0015] 또한, 본 발명은 컴퓨터와 결합하여 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법을 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램을 제공한다.

발명의 효과

[0016] 본 발명은 다음과 같은 우수한 효과가 있다.

[0017] 본 발명의 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법에 따르면, 콘텐츠 간의 유사도를 밝기 차이로 시각화된 그래프 또는 콘텐츠 간의 유사도가 높은 리스트들을 저작자에게 제공될 수 있어, 저작자가 참고하고자 하는 유사 콘텐츠들을 쉽고 빠르게 파악할 수 있다.

[0018] 또한, 본 발명의 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법에 따르면, 저작자가 유사 콘텐츠를 빠르게 파악할 수 있어서, 자료 수집에 투입되는 시간이 절약될 수 있다는 장점이 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 일 실시예에 따른 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법의 순서도,

도 2는 본 발명의 일 실시예에 따른 시각화 과정을 설명하기 위한 순서도,

도 3은 본 발명의 일 실시예에 따른 2차원 콘텐츠 단위 벡터의 시각화 그래프,

도 4는 본 발명의 일 실시예에 따른 2차원 콘텐츠 단위 벡터 간의 유클리디안 거리를 시각화한 그래프,

도 5는 본 발명의 일 실시예에 따른 2차원 콘텐츠 단위 벡터 간의 코사인 유사도를 시각화한 그래프,

도 6은 본 발명의 일 실시예에 따른 유사 콘텐츠 추천 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0020] 본 발명에서 사용되는 용어는 가능한 현재 널리 사용되는 일반적인 용어를 선택하였으나, 특정한 경우는 출원인이 임의로 선정한 용어도 있는데 이 경우에는 단순한 용어의 명칭이 아닌 발명의 상세한 설명 부분에 기재되거나 사용된 의미를 고려하여 그 의미가 파악되어야 할 것이다.

[0021] 이하, 첨부한 도면에 도시된 바람직한 실시예들을 참조하여 본 발명의 기술적 구성을 상세하게 설명한다.

[0022] 그러나 본 발명은 여기서 설명되는 실시예에 한정되지 않고 다른 형태로 구체화될 수도 있다. 명세서 전체에 걸쳐 동일한 참조번호는 동일한 구성요소를 나타낸다.

[0023] 도 1은 본 발명의 일 실시예에 따른 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법의 순서도로, 도 1을 참조하

면, 본 발명의 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법은 저작자에게 다시쓰기 작업에 필요한 유사한 콘텐츠를 시각화하거나 리스트로 만들어 제공할 수 있는 방법이다.

[0024] 또한, 상기 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법은 실질적으로 컴퓨터에 의해 수행되며 상기 컴퓨터에는 상기 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법이 수행하기 위한 컴퓨터 프로그램이 저장된다.

[0025] 또한, 상기 컴퓨터는 일반적인 퍼스널 컴퓨터뿐만 아니라, 스마트폰이나 태블릿 PC와 같은 스마트 기기를 포함하는 광의의 컴퓨팅 장치를 의미한다.

[0026] 또한, 상기 컴퓨터 프로그램은 별도의 기록 매체에 저장되어 제공될 수 있으며, 상기 기록 매체는 본 발명을 위하여 특별히 설계되어 구성된 것들이거나 컴퓨터 소프트웨어 분야에서 통상의 지식을 가진 자에게 공지되어 사용 가능한 것일 수 있다.

[0027] 예를 들면, 상기 기록 매체는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD, DVD와 같은 광 기록 매체, 자기 및 광 기록을 겸할 수 있는 자기-광 기록 매체, 롬, 램, 플래시 메모리 등 단독 또는 조합에 의해 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치일 수 있다.

[0028] 또한, 상기 컴퓨터 프로그램은 프로그램 명령, 로컬 데이터 파일, 로컬 데이터 구조 등이 단독 또는 조합으로 구성된 프로그램일 수 있고, 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라, 인터프리터 등을 사용하여 컴퓨터에 의해 실행될 수 있는 고급 언어 코드로 짜여진 프로그램일 수 있다.

[0029] 이하에서는 본 발명의 일 실시예에 따른 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법에 대해 상세히 설명한다.

[0030] 먼저, 본 발명의 스토리텔링 다시쓰기를 위한 콘텐츠 추천 방법은 콘텐츠들을 수집하여 데이터베이스를 구축한다(S1000).

[0031] 상기 콘텐츠는 소설, 영화, 드라마, 시나리오 등의 창작물, 전설, 설화 또는 민담을 포함한다.

[0032] 또한, 상기 데이터베이스는 아래의 표 1,2와 같이 제1 데이터 베이스와 제2 데이터 베이스로 구축될 수 있다.

표 1

[0033]

제1 데이터베이스	
인덱스	텐츠 제목
0	설씨녀와 가실
1	감생설화
2	강감찬설화
3	의견(義犬)설화
.	.
.	.
.	.

표 2

[0034]

제2 데이터베이스	
인덱스	텐츠 문장
0	설씨녀(薛氏女)와 가실(嘉實) 설씨녀(薛氏女)는 신라 경주의 민가 여자로 비록 평범한 가문과 가난한 집안 사람이었으나, 안색이 단정하고 행실이 바르므로 보는 사람마다 곱다고 하지 않는 사람이 없었다
.	.
.	.
.	.

[0035] 또한, 상기 제1 데이터 베이스 인덱스와 콘텐츠의 제목이 저장되고, 상기 제2 데이터 베이스는 인덱스와 문장이 저장되며, 상기 제1 데이터 베이스와 상기 제2 데이터 베이스는 하나로 통합될 수도 있다.

[0036] 한편, 상기 인덱스는 콘텐츠의 종류를 나타내는 것으로 예를 들면, 인덱스가 "0"인 경우 설씨녀와 가실에 관한 콘텐츠를 의미하고, 인덱스가 "1"인 경우 감생설화에 관한 콘텐츠를 의미한다.

[0037] 즉, 상기 인덱스는 해당 콘텐츠를 나타내는 고유 ID이다.

[0038] 다음, 상기 데이터베이스에 저장된 각 콘텐츠 문장들을 토큰화한다(S2000).

[0039] 상기 토큰화는 문장들을 단어 단위로 나뉘는 작업을 의미하며, 이를 위해 형태소 분석을 수행하여 품사별로 단어를 토큰화한다.

[0040] 아래의 표 3은 토큰화된 문장의 예를 보여준다.

표 3

[0041]

토큰화된 문장
['설씨녀', '(', '薛', '氏', '女', ')', '와', '가실', '(', '嘉實', ')', '설씨녀', '(', '薛', '氏', '女', ')', '는', '신라', '경주', '의', '민가', '여자', '로', '비록', '평범', '한', '가문', '과', '가난', '한', '집안', '사람', '이', '었', '으나', ',', '안색', '이', '단정', '하', '고', '행실', '이', '바르', '므로', '보', '는', '사람', '마다', '곱', '다고', '하', '지', '않', '는', '사람', '이', '없', '었', '다']

[0042] 다음, 토큰화된 각 콘텐츠의 문장들을 콘텐츠 단위로 벡터화한다(S3000).

[0043] 상기 벡터화는 토큰화된 문장의 단어들 간의 관계를 임베딩하여 벡터로 표현해주는 것으로 본 발명에서는 상기 벡터화를 위해 PV-DM 알고리즘에 상기 인덱스와 문장들을 입력하여 벡터화한다.

[0044] 한편, 상기 PV-DM 알고리즘으로 임베딩되어 출력되는 벡터는 각 인덱스에 대한 벡터 행렬(이하,"콘텐츠 단위 벡터"이라함)로 나오게되며, 사용자의 파라메타 설정에 따라 고차원으로 출력될 수 있다.

[0045] 다음, 각 콘텐츠 단위 벡터들 시각화하여 표현한다(S4000).

[0046] 도 2는 본 발명의 일실시예에 따른 시각화 과정을 설명하기 위한 순서도이다.

[0047] 도 2를 참조하면, 시각화를 위해 먼저, 각 콘텐츠 단위 벡터를 2차원으로 차원 축소한다(S4100).

[0048] 이는 상기 PV-DM 알고리즘을 통해 벡터화된 각 콘텐츠 단위 벡터가 고차원 행렬로 이루어져 있어, 시각화하기 위해 필요한 작업으로 2차원으로 축소하는 것이 바람직하다.

[0049] 또한, 상기 각 콘텐츠 단위 벡터를 2차원으로 축소하기 위해 PCA(Principal Component Analysis), SVD(singular value decomposition), NMF(Non-negative Martrix Factorization), T-SNE(T-Distributed Stochastic Neighbor Embedding 등 공지된 다양한 차원축소 알고리즘을 사용할 수 있으나 T-SNE 알고리즘을 사용하는 것이 바람직하다.

[0050] 도 3은 본 발명의 일 실시예에 따른 2차원으로 축소된 콘텐츠 단위 벡터들을 시각화한 그래프이다.

[0051] 도 3을 참조하면, 2차원으로 축소된 콘텐츠 단위 벡터(이하,"2차원 콘텐츠 단위 벡터"이라함)들은 좌표평면 상에 표시될 수 있으며, 이에 따라 상기 2차원 콘텐츠 단위 벡터는 x축과 y축 두 개의 좌표로 이루어진 것을 알 수 있다.

[0052] 한편, 상기 2차원 콘텐츠 단위 벡터만으로는 각 콘텐츠 간의 유사 정도를 파악하기 어려우며, 이에 따라 저작자가 한눈에 알아볼 수 있도록 아래의 방법들을 수행한다.

[0053] 다음, 상기 2차원 콘텐츠 단위 벡터를 간의 유클리디안 거리 또는 코사인 유사도를 계산한다(S4200).

[0054] 또한, 상기 유클리디안 거리는 아래의 수학식 1에 의해 계산될 수 있으며, 상기 코사인 유사도는 아래의 수학식 2에 의해 계산될 수 있다.

수학식 1

[0055] $$Distance(A.B)=\sqrt{||A||^2+||B||^2-2A\cdot B}$$

수학식 2

[0056] $$Similarity=\frac{\sum_{i=1} A_i \times B_i}{\sum_{i=1}(A_i)^2 \times \sum_{i=1}(B_i)^2}$$

[0057] A는 기준이 되는 콘텐츠의 2차원 콘텐츠 단위 벡터이며, B는 이외에 해당되는 2차원 콘텐츠 단위 벡터이다.

[0058] 이를 통해, 상기 수학식 1에서는 각 2차원 콘텐츠 단위 벡터 간의 거리를 유추할 수 있으며, 거리가 가까울수록 콘텐츠 간의 유사도가 높으며, 거리가 멀수록 콘텐츠 간의 유사도가 낮은 것을 의미한다.

[0059] 상기 수학식 2에서는 각 2차원 콘텐츠 단위 벡터 간의 코사인 유사도를 유추할 수 있으며, -1 내지 1의 값으로 출력된다.

[0060] 또한, 상기 코사인 유사도는 1에 가까울수록 콘텐츠 간의 유사도가 높은 것을 의미한다.

[0061] 다음, 2차원 콘텐츠 단위 벡터를 간에 계산된 유클리디안 거리 또는 코사인 유사도를 시각화한다(S4300).

[0062] 도 4는 본 발명의 일 실시예에 따른 2차원 콘텐츠 단위 벡터 간의 유클리디안 거리를 시각화한 그래프, 도 5는 본 발명의 일 실시예에 따른 2차원 콘텐츠 단위 벡터 간의 코사인 유사도를 시각화한 그래프이다.

[0063] 상기 시각화는 다양한 형태로 표현될 수 있으나, 색 차이를 이용하여 표현될 수 있으며, 바람직하게는 Confusion Matrix로 시각화될 수 있다.

[0064] 또한, 상기 Confusion Matrix는 상기 2차원 콘텐츠 단위 벡터들 간에 계산되 유클리디안 거리를 시각화할 경우 계산된 유클리디안 거리가 0에 가까울수록 밝게 표현하고 거리가 멀수록 어둡게 표현되는 것이 바람직하다.

[0065] 또한, 상기 Confusion Matrix로 상기 2차원 콘텐츠 단위 벡터들 간에 계산된 코사인 유사도를 시각화할 경우, 계산된 코사인 유사도가 1에 가까울수록 어둡게 표현하고, 1에서 멀어질수록 밝게 표현하는 것이 바람직하다.

[0066] 즉, 상기 시각화된 그래프는 각 콘텐츠 간의 유사한 정도를 색의 밝기로 나타낸 그래프이다.

[0067] 한편, 상기 2차원 콘텐츠 단위 벡터들 간에 계산된 유클리디안 거리 또는 상기 2차원 콘텐츠 단위 벡터들 간에 계산된 코사인 유사도 중 하나를 시각화하여 저작자에게 제공되거나 두 개 모두 시각화하여 저작자에게 제공될 수 있다.

[0068] 따라서, 본 발명의 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법에 의하면, 시각화된 각 콘텐츠 간의 유사도를 색 차이를 이용하여 보여줌으로써 저작자가 육안으로 쉽고 편리하게 유사 콘텐츠를 확인할 수 있다는 장점이 있다.

[0069] 또한, 본 발명의 추천 방법은 콘텐츠 간에 유사도를 시각화하여 저작자에게 출력하는 것 이외에 각 콘텐츠 간의 유사도가 높은 콘텐츠들을 리스트로 추천할 수 있다(S5000).

[0070] 도 6은 본 발명의 일 실시예에 따른 콘텐츠 추천 방법의 순서도로, 도 6을 참조하면, 콘텐츠 추천 방법은 시각화를 수행하는 과정(S4000)과 유사하게 상기 콘텐츠 단위 벡터를 2차원으로 차원 축소하고(S5100), 2차원 콘텐츠 단위 벡터를 상기 수학식 1과 상기 수학식 2에 각각 대입하여, 2차원 콘텐츠 단위 벡터들 간의 유클리디안 거리 및 코사인 유사도를 추출한다(S5200).

[0071] 다음, 계산된 2차원 콘텐츠 단위 벡터를 간의 유클리디안 거리와 계산된 2차원 콘텐츠 단위 벡터들 간의 코사인

유사도에 각각 임계값을 주어 조건 범위에 해당되는 콘텐츠를 추출하여 추천한다(S5300).

[0072] 바람직하게는, 상기 2차원 콘텐츠 단위 벡터들 간의 유클리디안 거리가 임계 거리 이하에 해당되고, 상기 2차원 콘텐츠 단위 벡터들 간의 코사인 유사도가 임계 유사도 이상에 해당되는 콘텐츠를 추천한다.

[0073] 또한, 상기 임계 거리는 4 내지 6, 상기 임계 유사도는 0.6 내지 0.8일 수 있으며, 바람직하게는 상기 임계 거리는 5, 상기 임계 유사도는 0.7일 수 있다.

[0074] 또한, 상기 조건 범위에 만족하는 2차원 콘텐츠 단위 벡터는 상기 데이터베이스로부터 해당 콘텐츠의 제목과 문장들을 불러와 저작자에게 제공될 수 있다.

[0075] 아래의 표 4는 상기 조건 범위에 만족하는 콘텐츠가 추천되는 리스트의 예를 보여준다.

표 4

[0076]

	유사 콘텐츠1	유사 콘텐츠2	유사 콘텐츠3
1	함몰(陷沒) 설화	인사이드 아웃	로건
2	트윈스터즈	에이 아이	-
.	.	.	.
.	.	.	.
.	.	.	.

[0077] 따라서, 본 발명의 스토리텔링 작업 중 다시쓰기를 위한 콘텐츠 추천 방법은 저작자가 시각적으로 각 콘텐츠 간의 유사도를 확인할 수 있으며, 더 나아가 각 콘텐츠 간의 유사도가 높은 콘텐츠들을 리스트로 제공하기 때문에, 자료 수집에 투입되는 시간이 절약될 수 있다는 장점이 있다.

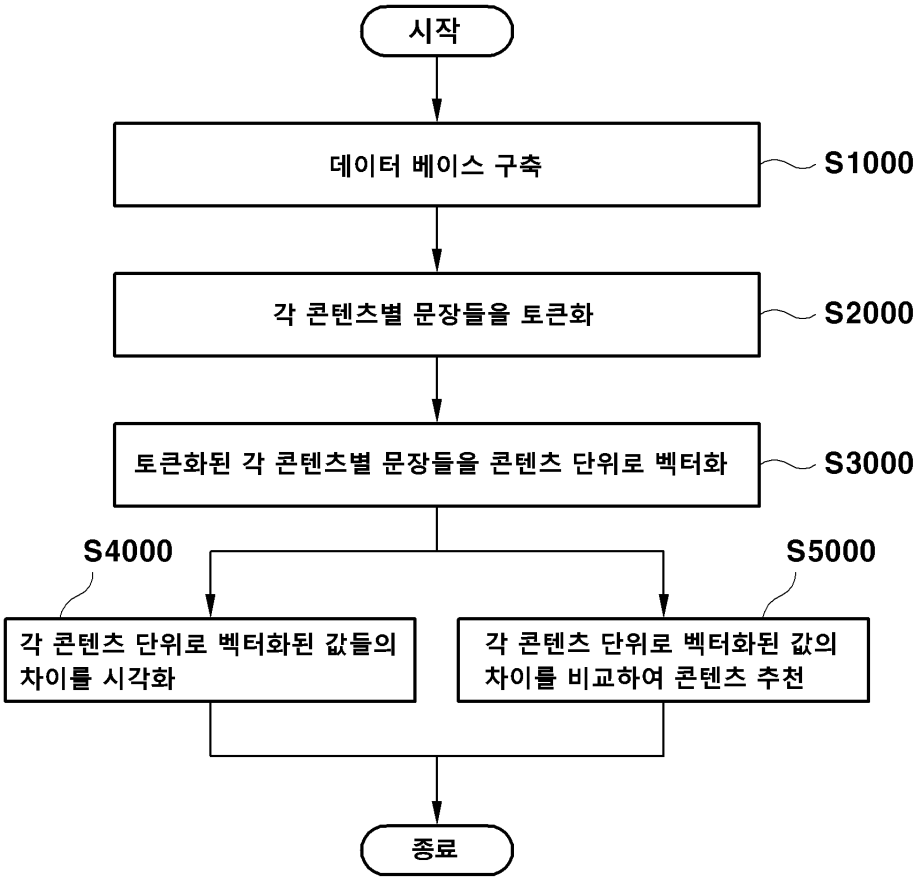
[0078] 이상에서 살펴본 바와 같이 본 발명은 바람직한 실시예를 들어 도시하고 설명하였으나, 상기한 실시예에 한정되지 아니하며 본 발명의 정신을 벗어나지 않는 범위 내에서 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변경과 수정이 가능할 것이다.

부호의 설명

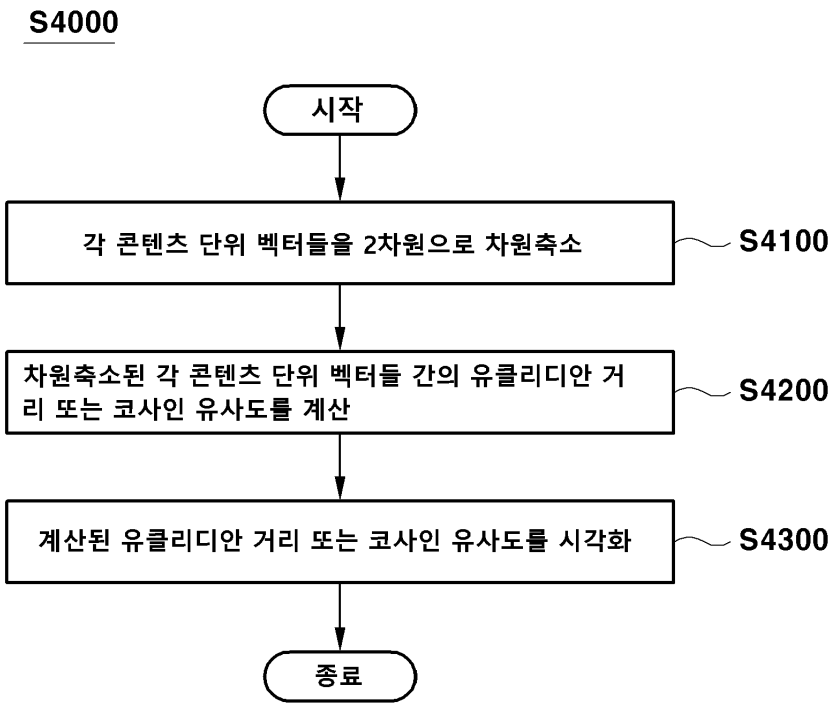
삭제

도면

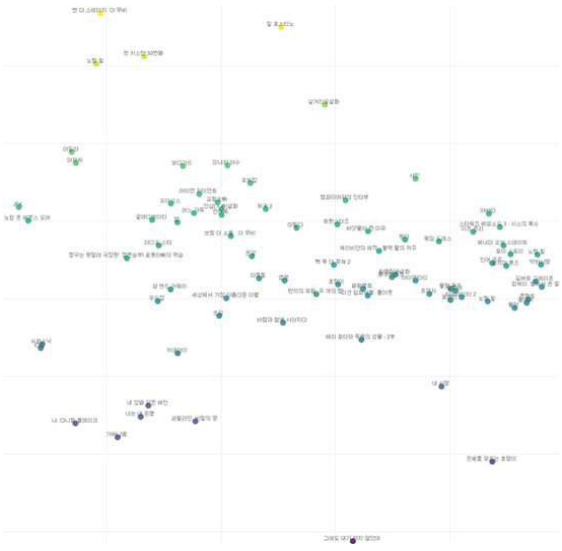
도면1


시작
데이터 베이스 구축
S1000
각 콘텐츠별 문장들을 토큰화
S2000
토큰화된 각 콘텐츠별 문장들을 콘텐츠 단위로 벡터화
S3000
S4000
S5000
각 콘텐츠 단위로 벡터화된 값들의 차이를 시각화
각 콘텐츠 단위로 벡터화된 값의 차이를 비교하여 콘텐츠 추천
종료

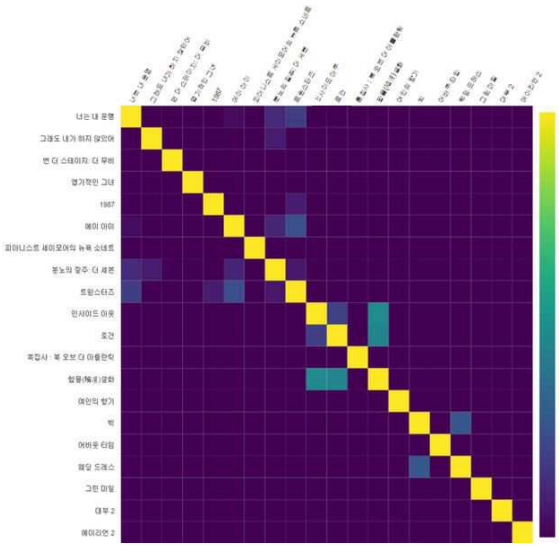
도면2


S4000
시작
각 콘텐츠 단위 벡터들을 2차원으로 차원축소
S4100
차원축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 또는 코사인 유사도를 계산
S4200
계산된 유클리디안 거리 또는 코사인 유사도를 시각화
S4300
종료

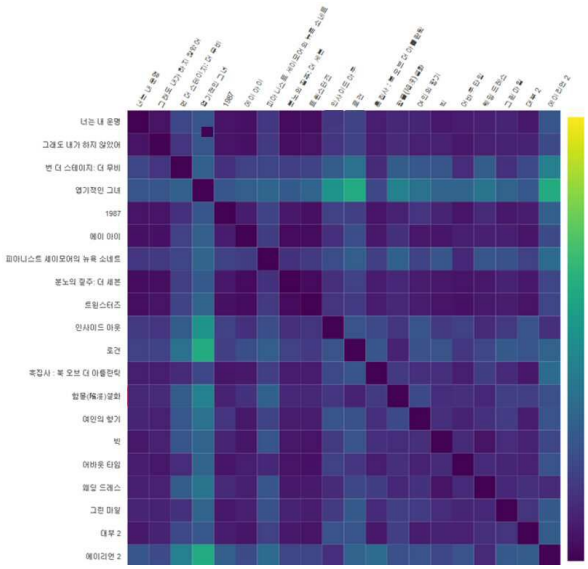
도면3



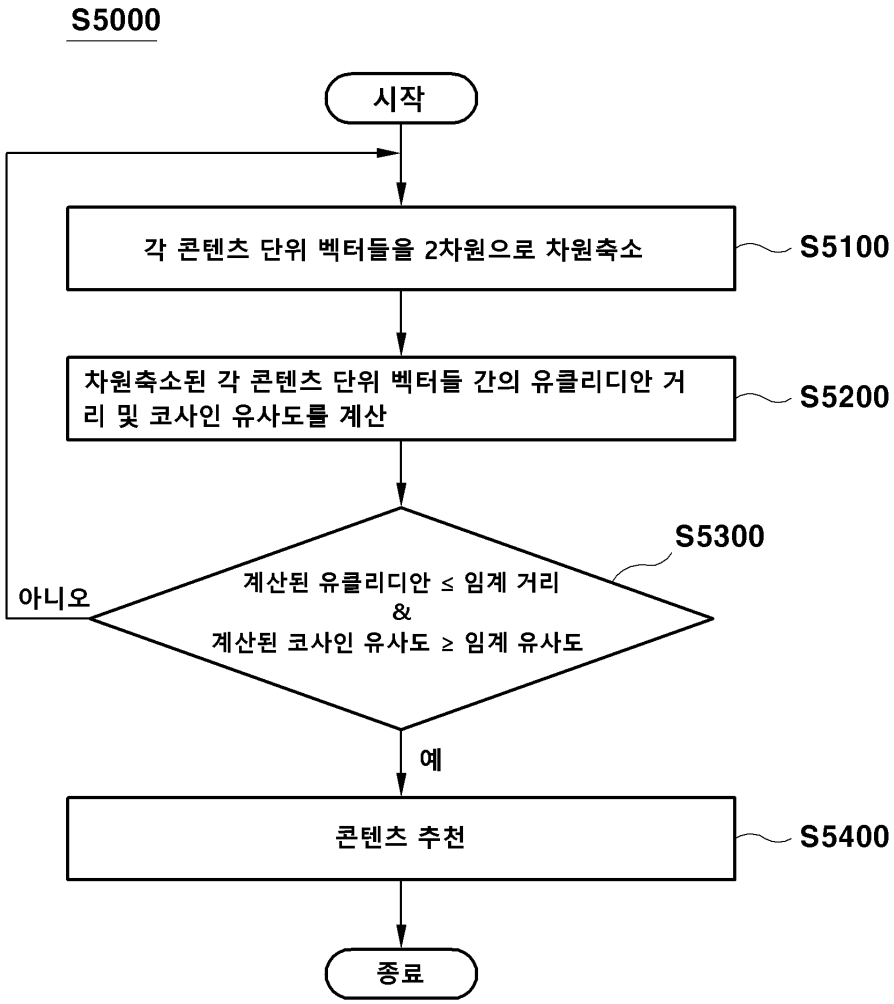
도면4



도면5



도면6


S5000
시작
각 콘텐츠 단위 벡터들을 2차원으로 차원축소
S5100
차원축소된 각 콘텐츠 단위 벡터들 간의 유클리디안 거리 및 코사인 유사도를 계산
S5200
S5300
계산된 유클리디안 ≤ 임계 거리
&
계산된 코사인 유사도 ≥ 임계 유사도
아니오
예
콘텐츠 추천
S5400
종료