



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년08월10일
(11) 등록번호 10-1887469
(24) 등록일자 2018년08월06일

- (51) 국제특허분류(Int. Cl.)
G06Q 30/06 (2012.01) G06F 17/27 (2006.01)
G06Q 30/02 (2012.01)
- (52) CPC특허분류
G06Q 30/0631 (2013.01)
G06F 17/2785 (2013.01)
- (21) 출원번호 10-2017-0015442
- (22) 출원일자 2017년02월03일
심사청구일자 2017년02월03일
- (56) 선행기술조사문헌
JP2013210787 A*
JP2016006593 A*
KR1020100066862 A*
KR1020160098084 A*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
(주)브레인콜라
경기도 용인시 기흥구 죽전로 10 ,6층601호(보정동)
- (72) 발명자
김강산
경기도 용인시 기흥구 기흥역로16, 지웰푸르지오 101동 3302호
박지훈
서울특별시 강서구 강서로54나길 37, 308동 506호
- (74) 대리인
최승욱

전체 청구항 수 : 총 3 항

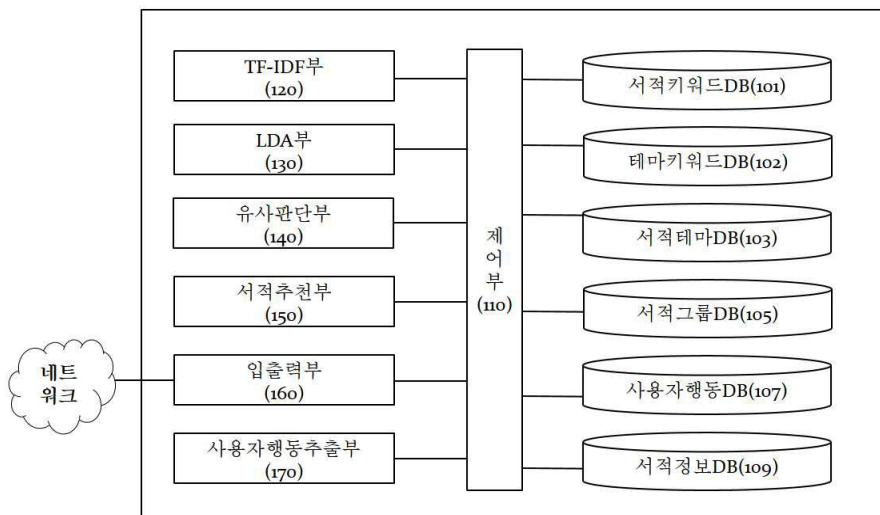
심사관 : 안창민

(54) 발명의 명칭 서적 간 유사도를 활용한 서적추천 방법

(57) 요약

서적 간 유사도를 활용한 서적추천방법이 개시된다. 본 발명의 일 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터로부터 하나 이상의 키워드를 추출하는 단계; (3) 상기 서적추천서버가 제1 서적의 키워드와 제2 서적의 키워드 간의 유사도를 연산하는 단계; 및 (4) 상기 서적추천서버가 상기 유사도에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도를 활용한 서적추천 방법을 포함하는 서적 간 유사도를 기초한 서적추천 방법이 제공된다.

대표도 - 도1



(52) CPC특허분류
G06Q 30/0255 (2013.01)

명세서

청구범위

청구항 1

삭제

청구항 2

삭제

청구항 3

삭제

청구항 4

(1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계;

(2) 상기 서적추천서버가, 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하거나, 상기 키워드를 기초로 하나 이상의 서적테마를 배정하거나, LDA(Latent Dirichlet Allocation)알고리즘을 기반으로 상기 서적데이터 내에서 하나 이상의 그룹을 생성하는 단계;

(4) 상기 서적추천서버가, 제1 서적의 키워드과 제2 서적의 키워드 간의 제1 유사도를 연산하거나, 제1 서적의 서적테마와 제2 서적의 서적테마 간의 제2 유사도를 연산하거나, 제1 서적의 그룹과 제2 서적의 그룹 간의 제3 유사도를 연산하는 단계;

(5) 상기 서적추천서버가 상기 제1 유사도 내지 제3 유사도 중 어느 하나 이상에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하고,

상기 (2)단계는,

상기 서적추천서버가, 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하거나, 상기 키워드를 기초로 하나 이상의 서적테마를 배정하거나, LDA(Latent Dirichlet Allocation)알고리즘을 기반으로 상기 서적데이터 내에서 하나 이상의 그룹을 생성하거나, 복수의 서적 각각의 스토리플로우 감정그래프를 생성하는 단계이고,

상기 (4)단계는,

상기 서적추천서버가, 제1 서적의 키워드과 제2 서적의 키워드 간의 제1 유사도를 연산하거나, 제1 서적의 서적테마와 제2 서적의 서적테마 간의 제2 유사도를 연산하거나, 제1 서적의 그룹과 제2 서적의 그룹 간의 제3 유사도를 연산하거나 제1 서적의 스토리플로우 감정그래프와 제2 서적의 스토리플로우 감정그래프 간의 제4 유사도를 연산하는 단계이고,

상기 (5)단계는,

상기 서적추천서버가 상기 제1 유사도 내지 제4 유사도 중 어느 하나 이상에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계이며,

상기 (2) 단계는,

각 서적에 포함된 전체 단어들의 개수에 대한 소정의 비율의 개수 만큼의 단어들을 분석하는 윈도우(window)를 생성하는 단계; 및

상기 윈도우를 각 서적에 포함된 제1 번 위치에 있는 단어에서 제2 위치에 있는 단어 방향인 서적의 스토리방향으로 순차적으로 이동시켜, 상기 이동된 서적 내 위치의 윈도우 내에 포함된 단어들의 감정을 분석하여 스토리

플로우 감정그래프를 생성하는 단계를 더 포함하는 서적 간 유사도를 활용한 서적추천 방법.

청구항 5

삭제

청구항 6

삭제

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

제4항에 있어서,

상기 윈도우를 각 서적에 포함된 제1 번 위치에 있는 단어에서 제2 위치에 있는 단어 방향인 서적의 스토리방향으로 순차적으로 이동시켜, 상기 이동된 서적 내 위치의 윈도우 내에 포함된 단어들의 감정을 분석하여 스토리플로우 감정그래프를 생성하는 단계는,

기정의된 감정단어테이블 내에 포함된 단어들을 참조하여 상기 윈도우 내에 포함된 각 단어에 대한 감정값을 부여하는 단계;

상기 윈도우 내 각 단어에 대한 감정값들을 합산한 단위감정합산값을 도출하는 단계; 및

상기 단위감정합산값에 기초하여 스토리플로우 감정그래프에 대한 데이터를 생성하는 단계를 더 포함하는 서적 간 유사도를 활용한 서적추천 방법.

청구항 12

제4항에 있어서,

상기 (4) 단계는,

상기 스토리방향에 따른 상기 제1 서적의 단위감정합산값과 상기 스토리방향에 따른 상기 제2 서적의 단위감정합산값 간의 차이를 나타내는 제1 유사판단기초값; 및

상기 제1 유사판단기초값의 변화량을 나타내는 제2 유사판단기초값 중 어느 하나 이상에 기초하여 제4 유사도를 연산하는 것임을 특징으로 하는 서적 간 유사도를 활용한 서적추천 방법.

발명의 설명

기술 분야

[0001] 본 발명은 서적 간 유사도를 활용한 서적추천 방법에 관한 것이다.

배경 기술

[0002] 과거로부터 읽은 책을 기록하고 관리하며, 책에 대한 평가를 공유하는 서비스는 존재하여 왔다. 또한, 이와 연관된 도서추천 서비스도 시행 중으로 국내에서는 그 서비스가 형성되어가는 단계이지만 해외, 특히 미국에서는

도서추천 서비스가 확고히 자리를 잡아 가고 있다. 이런 도서추천 서비스는 최근 인터넷 웹사이트뿐만 아니라 스마트폰 애플리케이션과 연동되어 서비스 중에 있다.

- [0003] 국내의 책 추천 서비스를 수행과 관련 특허를 검토하면 '도서 SNS 시스템 및 그 제공방법' (제10-2014-0038017호, 공개일자 2014.03.28)이 존재한다. 본 공개특허에는 도서추천을 위해 SNS(Social Networking Service)를 기반으로 다른 사용자의 콘텐츠를 공유하여 도서추천에 활용한다는 개념이 개시되어 있다. 다른 국내 책 추천 서비스에 관련 특허인 '독서 정보 공유를 통한 소셜 네트워크 서비스 제공 방법 및 시스템' (제10-2014-0133647호, 공개일자 2104.11.20)에도 책을 읽은 독자가 책에 대한 정보를 SNS를 통하여 다른 독자와 공유하겠다는 개념 정도만 개시되어 있다.
- [0004] 선행 도서 추천 방법 모두 SNS나 개인화된 정보를 이용하여 독자에게 도서를 추천하는 것에 한정되어 있을 뿐이며 실제 책의 내용을 분석하여 분류하고 추천하는 방법에 대해서는 명확히 개시한 바 없다.
- [0005] 이에 본 발명의 발명가는 책의 내용을 분석하고 이를 바탕으로 독자가 흥미를 느낄 수 있는 책을 추천하는 방법을 제시하였다.

선행기술문헌

특허문헌

- [0007] (특허문헌 0001) 대한민국 공개특허공보 제10-2014-0038017호(2014.03.28.)
- (특허문헌 0002) 대한민국 등록특허공보 제10-2014-0133647호 (2104.11.20.)

발명의 내용

해결하려는 과제

- [0008] 본 발명 서적 간 유사도를 활용한 서적추천 방법을 제공하는 것이다.

과제의 해결 수단

- [0009] 본 발명의 일 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터로부터 하나 이상의 키워드를 추출하는 단계; (3) 상기 서적추천서버가 제1 서적의 키워드와 제2 서적의 키워드 간의 유사도를 연산하는 단계; 및 (4) 상기 서적추천서버가 상기 유사도에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도에 기초한 서적추천 방법이 제공된다.
- [0010] 본 발명의 다른 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하는 단계; (3) 상기 서적추천서버가 상기 키워드를 기초로 하나 이상의 서적테마를 배정하는 단계; (4) 상기 서적추천서버가 제1 서적의 서적테마와 제2 서적의 서적테마 간의 유사도를 연산하는 단계; (5) 상기 서적추천서버가 상기 유사도에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도에 기초한 서적추천방법이 제공된다.
- [0011] 본 발명의 다른 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가 LDA(Latent Dirichlet Allocation)알고리즘을 활용하여 상기 서적데이터 내에서 하나 이상의 그룹을 생성하는 단계; (3) 상기 서적추천서버가 제1 서적의 그룹과 제2 서적의 그룹 간의 유사도를 연산하는 단계; 및 (4) 상기 서적추천서버가 상기 유사도에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도에 기초한 서적추천방법이 제공된다.
- [0012] 본 발명의 다른 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가, 단어 빈도와

역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하거나, 상기 키워드를 기초로 하나 이상의 서적테마를 배정하거나, LDA(Latent Dirichlet Allocation)알고리즘을 기반으로 상기 서적데이터 내에서 하나 이상의 그룹을 생성하는 단계; (4) 상기 서적추천서버가, 제1 서적의 키워드과 제2 서적의 키워드 간의 제1 유사도를 연산하거나, 제1 서적의 서적테마와 제2 서적의 서적테마 간의 제2 유사도를 연산하거나, 제1 서적의 그룹과 제2 서적의 그룹 간의 제3 유사도를 연산하는 단계; (5) 상기 서적추천서버가 상기 제1 유사도 내지 제3 유사도 중 어느 하나 이상에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도에 기초한 서적추천방법이 제공된다.

[0013] 본 발명의 다른 측면에 따르면, (1) 서적추천서버가 서적의 제목, 서적의 본문, 서적의 소개정보 중 어느 하나 이상을 포함하는 복수 개의 서적에 대한 서적데이터를 입력받는 단계; (2) 상기 서적추천서버가 제1 또는 제3 유사도 연산 방식 중 하나를 선택하는 단계; (3) 제1 유사도 연산 방식이 선택된 경우, 상기 서적추천서버가 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하고, 제3 유사도 연산 방식이 선택된 경우, 상기 서적추천서버가 LDA(Latent Dirichlet Allocation)알고리즘을 기반으로 상기 서적데이터 내에서 하나 이상의 그룹을 생성하는 단계; (4) 제1 유사도 연산 방식이 선택된 경우, 상기 서적추천서버가 제1 서적의 키워드과 제2 서적의 키워드 간의 제1 유사도를 연산하고, 제3 유사도 연산 방식이 선택된 경우, 제1 서적의 그룹과 제2 서적의 그룹 간의 제3 유사도를 연산하는 단계; (5) 상기 서적추천서버가 상기 제1 유사도 또는 제3 유사도에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계를 포함하는 서적 간 유사도에 기초한 서적추천방법이 제공된다.

[0014] 또한, 상기 (2) 단계는, 서적의 정보 혹은 사용자의 정보에 따라 상기 제1 또는 상기 제3 유사도 연산 방식 중 하나를 선택하는 단계일 수 있다.

[0015] 또한, 상기 제1 서적은, 사용자에게 의해 소정의 기준값 보다 높은 평가점수를 부여받은 서적, 온라인 쇼핑물의 장바구니에 포함된 서적 또는 사용자에게 의해 온라인 공간상에서 관련 서적 정보가 열람된 적이 있는 서적일 수 있다.

[0016] 또한, 상기 유사서적을 출력하는 단계는, 상기 하나 이상의 유사서적 중 사용자가 기독서한 서적은 제외하는 단계; 상기 하나 이상의 유사서적 중 서적 평점에 순서대로 상기 유사서적을 출력하는 단계; 상기 하나 이상의 유사서적 중 사용자 정보를 고려하여 상기 유사서적을 출력하는 단계 중 어느 하나 이상의 단계를 더 포함할 수 있다.

[0017] 또한, 상기 (2)단계는, 상기 서적추천서버가, 단어 빈도와 역문서 빈도(Term Frequency - Inverse Document Frequency)를 기반으로 상기 서적데이터에서 하나 이상의 키워드를 추출하거나, 상기 키워드를 기초로 하나 이상의 서적테마를 배정하거나, LDA(Latent Dirichlet Allocation)알고리즘을 기반으로 상기 서적데이터 내에서 하나 이상의 그룹을 생성하거나, 복수의 서적 각각의 스토리플로우 감정그래프를 생성하는 단계이고 상기 (4)단계는, 상기 서적추천서버가, 제1 서적의 키워드과 제2 서적의 키워드 간의 제1 유사도를 연산하거나, 제1 서적의 서적테마와 제2 서적의 서적테마 간의 제2 유사도를 연산하거나, 제1 서적의 그룹과 제2 서적의 그룹 간의 제3 유사도를 연산하거나 제1 서적의 스토리플로우 감정그래프와 제2 서적의 스토리플로우 감정그래프 간의 제4 유사도를 연산하는 단계; 상기 (5)단계는, 상기 추천서버가 상기 제1 유사도 내지 제4 유사도 중 어느 하나 이상에 따라 제1 서적과 유사한 하나 이상의 유사서적을 출력하는 단계일 수 있다.

[0018] 또한, 상기 (2) 단계는, 각 서적에 포함된 전체 단어들의 개수에 대한 소정의 비율의 개수 만큼의 단어들을 분석하는 윈도우(window)를 생성하는 단계; 및 상기 윈도우를 각 서적에 포함된 제1 번 위치에 있는 단어에서 제2 위치에 있는 단어 방향으로 순차적으로 이동시켜, 상기 이동된 서적 내 위치의 윈도우 내에 포함된 단어들의 감정을 분석하여 스토리플로우 감정그래프를 생성하는 단계를 더 포함할 수 있다.

[0019] 상기 윈도우를 각 서적에 포함된 제1 번 위치에 있는 단어에서 제2 위치에 있는 단어 방향으로 순차적으로 이동시켜, 상기 이동된 서적 내 위치의 윈도우 내에 포함된 단어들의 감정을 분석하여 스토리플로우 감정그래프를 생성하는 단계는, 기정의된 감정단어테이블 내에 포함된 단어들을 참조하여 상기 윈도우 내에 포함된 각 단어에 대한 감정값을 부여하는 단계; 상기 윈도우 내 각 단어에 대한 감정값들을 합산한 단위 감정합산값을 도출하는 단계; 및 상기 단위감정합산값에 기초하여 스토리플로우 감정그래프에 대한 데이터를 생성하는 단계를 더 포함할 수 있다.

[0020] 또한, 상기 윈도우를 각 서적에 포함된 제1 번 위치에 있는 단어에서 제2 위치에 있는 단어 방향으로 순차적으로 이동시켜, 상기 이동된 서적 내 위치의 윈도우 내에 포함된 단어들의 감정을 분석하여

스토리플로우 감정그래프를 생성하는 단계는, 상기 윈도우를 분석대상 서적의 각 페이지에 포함된 단어들의 개수의 평균개수의 단어만큼 이동시키는 서적 간 유사도를 활용할 수 있다.

- [0021] 상기 (4) 단계는, 상기 스토리방향에 따른 상기 제1 서적의 단위감정합산값과 상기 스토리방향에 따른 상기 제2 서적의 단위감정합산값 간의 차이를 나타내는 제1 유사판단기초값; 및 상기 제1 유사판단기초값의 변화량을 나타내는 제2 유사판단기초값 중 어느 하나 이상에 기초하여 제4 유사도를 연산하는 것일 수 있다.
- [0022] 본 발명의 다른 측면에 따르면, 상기 방법을 실행시키는 기록매체에 저장된 컴퓨터 프로그램일 수 있다.

발명의 효과

- [0024] 본 발명에 따르면 서적추천방법을 제공할 수 있다.

도면의 간단한 설명

- [0026] 도 1은 본 발명의 실시예에 따른 서적추천시스템의 구성도.
- 도 2 내지 5는 본 발명의 실시예에 따른 순서도.
- 도 6은 본 발명의 다른 실시예에 따른 서적추천시스템의 구성도.
- 도 7은 윈도우와 스토리플로우 감정그래프 생성을 설명하기 위한 개념도.
- 도 8은 감정단어 데이터베이스에 포함된 행복이라는 감정을 나타내는 단어들의 감정값 테이블.
- 도 9는 감정값이 부여된 단어가 존재하는 서적의 페이지를 나타낸 도면.
- 도 10은 스토리플로우 방향에 따른 단위감정합산값의 변화를 나타낸 그래프.
- 도 11은 실시예에 따른 서적들의 단위감정합산값을 나타낸 테이블
- 도 12는 스토리플로우에 따른 감정그래프.
- 도 13은 유사판단기초값을 나타내는 테이블.
- 도 14 내지 17은 본 발명의 실시예에 따른 순서도.
- 도 18은 본 발명의 다른 실시예에 따른 서적추천시스템의 구성도.

발명을 실시하기 위한 구체적인 내용

- [0027] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세한 설명에 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0028] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.
- [0029] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0030] 이하, 본 발명에 따른 실시예를 첨부도면을 참조하여 상세히 설명하기로 하며, 첨부 도면을 참조하여 설명함에 있어, 동일하거나 대응하는 구성 요소는 동일한 도면번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.
- [0031] 또한, 이하 사용되는 제1, 제2 등과 같은 용어는 동일 또는 상응하는 구성 요소들을 구별하기 위한 식별 기호에 불과하며, 동일 또는 상응하는 구성 요소들이 제1, 제2 등의 용어에 의하여 한정되는 것은 아니다.

- [0032] 또한, 결합이라 함은, 각 구성 요소 간의 접촉 관계에 있어, 각 구성 요소 간에 물리적으로 직접 접촉되는 경우만을 뜻하는 것이 아니라, 다른 구성이 각 구성 요소 사이에 개재되어, 그 다른 구성에 구성 요소가 각각 접촉되어 있는 경우까지 포괄하는 개념으로 사용하도록 한다.
- [0034] 도 1은 본 발명에 따른 서적추천시스템의 구성도이다. 도 1을 참조하면, 서적추천시스템은 제어부(110), TF-IDF부(120), LDA부(130), 유사판단부(140), 서적추천부(150), 입출력부(160), 사용자행동추출부(170), 서적키워드데이터베이스(101), 테마키워드데이터베이스(103), 서적그룹데이터베이스(105), 사용자행동데이터베이스(107) 및 서적정보데이터베이스(109)이 도시되어 있다. 본 서적추천시스템은 서적추천서버가 될 수 있다. 한편, 상기 데이터베이스는 하나의 데이터베이스로 구성하여도 무방하다. 또한, 본 발명의 상세한 설명 및 청구항에서는 서적에 대해서 언급하고 있으나 본 발명의 적용은 서적에 한정된 것이 아니며, 웹소셜, 만화 등 활자를 포함하는 구별된 텍스트에 대해서는 모두 적용이 가능하다.
- [0036] 먼저, 제어부(110)는 TF-IDF부, LDA부, 유사판단부, 서적추천부, 입출력부 및 사용자행동추출부 등 구성들을 제어하는 역할을 한다. 예를 들어, 사용자로부터 제어신호 등을 입력받아 제어신호에 따라 서적의 키워드를 추출하도록 TF-IDF부를 제어하는 역할을 할 수 있다. 제어부가 제어부에 연결된 각각의 데이터베이스들을 제어할 수 있음은 물론이다. 그 밖에 서적의 키워드에 따라 서적의 테마를 부여하는 역할을 할 수 있다. 또한, 서적의 장르나 사용자의 취향 등을 고려하여 서적의 유사도 측정 방법을 선택하는 역할을 할 수 있다. 이에 대해서는 하기에서 더욱 상세히 설명한다. 한편, 도 1에서는 각종 데이터베이스가 서적추천시스템인 서버 내에 포함되어 있는 것을 도시되어 있으나 상기 데이터베이스들이 서버 외부에 존재하여도 무방하다.
- [0038] 다음으로, TF-IDF부(120)는 서적의 텍스트 내 단어 빈도와 역문서 빈도를 참조하여 서적을 대표할 수 있는 키워드들을 추출하는 역할을 한다. TF-IDF(Term Frequency - Inverse Document Frequency)는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군, 문서집합이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치이다. 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정하거나, 문서들 사이의 비슷한 정도를 구하는 등의 용도로 사용할 수 있다. TF-IDF로 키워드를 추출하는 것은 널리 알려진 것으로 종래의 방법을 따를 수 있다.
- [0039] 더욱 상세하게, TF-IDF부는 서적의 본문, 서적의 제목, 서적을 소개하는 소갯글 등을 서적정보데이터베이스로부터 읽어 드린다. 이후 상기 서적의 본문, 서적의 제목, 서적을 소개하는 소갯글 중 어느 하나 이상의 텍스트를 TF-IDF 기반으로 분석하여 키워드를 추출하는 것이다. 서적을 대표하는 키워드는 하나 이상 추출될 수 있으며 예를 들어 50 여개가 추출될 수 있다. 이렇게 추출된 서적에 대한 키워드들은 서적키워드데이터베이스에 저장된다. 즉, 서적을 대표하는 복수 개의 키워드들이 서적키워드데이터베이스에 저장되는 것이다. 상기 키워드 각각은 가중치를 가지고 있게 되고, 서적을 대표하는 키워드들은 가중치를 바탕으로 벡터로 표현될 수 있다.
- [0041] 한편, 제어부는 TF-IDF부가 추출한 키워드들을 바탕으로 각 서적의 테마 혹은 장르를 부여할 수 있다. 더욱 구체적으로 제어부는 TF-IDF부가 추출한 각 서적에 대한 키워드들을 테마키워드데이터베이스 내에 저장된 키워드와 비교하여 테마를 부여할 수 있다. 테마키워드데이터베이스에는 각 키워드에 대한 테마가 배정되어 있다. 예를 들어, "가계부"라는 키워드에 대해서는 "경제"라는 테마가 배정된다. 다른 예를 들어, "골프"라는 키워드에 대해서는 "스포츠"라는 테마가 배정될 수 있다. 테마키워드데이터베이스에는 수천 가지의 다양한 테마가 존재할 수 있다. 이런 테마분류는 외부 온라인 서점의 데이터베이스의 분류를 활용하여 생성될 수 있다.
- [0042] 한편, 각 키워드와 테마는 일대일로 배정되는 것에 한정되지 않으며 일대다의 형태로 배정될 수 있다. 예를 들어, "가계부"는 "경제"라는 테마에 배정될 수 있지만 "경제"외에도 "다이어리" 및 "문구"라는 테마에도 배정될 수 있다. 각 키워드에 대한 배정된 테마는 벡터형태로 표현될 수 있다. TF-IDF부에 의해 분석된 서적의 키워드들에 대하여, 제1 제어부가 테마를 배정하여 최종적으로 하나의 서적은 하나 이상의 테마값이 배정될 수 있고 이런 테마값은 테마벡터를 형성할 수 있다. 이런 서적테마들은 서적테마데이터베이스에 저장될 수 있다.
- [0044] 다음으로 LDA부(130)는 LDA(Latent Dirichlet Allocation)알고리즘을 활용하여 상기 서적데이터 내에서 하나 이상의 그룹을 생성하는 역할을 한다. 더욱 구체적으로 LDA부는 서적의 본문, 제목, 소갯글 등 다양한 서적데이터를 분석하여 상기 서적데이터 내의 그룹을 생성한다. 예를 들어, 제1 서적의 서적데이터를 분석하였더니 제1 그룹, 제3 그룹이 생성되고, 제2 서적의 서적데이터를 분석하였더니 제2 그룹, 제4 그룹이 생성되었다고 가정할 수 있다. 이와 같이 이후 유사판단부를 통하여 판단한 결과 제1 그룹과 제2 그룹의 성격이 유사하여 동일한 그룹으로 판단할 수 있는 것이다. 이에 대한 설명은 하기에서 상술한다. LDA 알고리즘은 종래의 발명을 따를 수 있다.

- [0046] 유사판단부(140)는 서적들 간의 유사여부를 판단하는 역할을 한다. 더욱 구체적으로, 유사판단부는 상기 TF-IDF 부가 추출한 서적에 대한 키워드집합들을 연산하여 각 서적의 유사성을 연산한다. 유사판단부는 앞서 설명한 각 서적을 대표하는 키워드집합들 간에 연산, 예를 들어 코사인 유사도(cosine similarity) 연산하여 키워드들 간의 유사도를 도출할 수 있다. 이를 통하여 특정 서적과 유사한 서적이 어떤 서적인지 리스트를 생성할 수 있다. 이런 유사정보는 서적키워드데이터베이스 또는 서적정보데이터베이스 혹은 별도의 데이터베이스(미도시)에 저장될 수 있다.
- [0047] 또한, 유사판단부는 제어부가 도출한 각 서적에 대한 서적테마들 간의 유사도를 연산하는 역할을 할 수도 있다. 앞서 설명한 바와 같이, 유사판단부는 서적테마데이터베이스에 저장된 각 서적에 대한 서적테마값을 연산하여 각 서적 간의 유사도를 도출할 수 있다.
- [0048] 또한, 유사판단부는 LDA부가 생성하여 서적그룹데이터베이스에 저장한 서적그룹정보를 이용하여 서적 간의 유사도를 판단할 수 있다. 앞서 설명한 바와 같이 예를 들어, 제1 서적의 데이터가 제1 그룹, 제2 그룹, 제3 그룹으로 그룹핑이 되고, 제2 서적이 제4 그룹, 제5 그룹, 제3 서적이 제6 그룹, 제7 그룹으로 그룹핑된다고 가정한다. 이때 유사판단부가 제1 그룹, 제2 그룹, 제3 그룹과 제4 그룹, 제5 그룹 간 비교연산하고, 제1 그룹, 제2 그룹, 제3 그룹과 제6 그룹, 제7 그룹 간 비교연산한다. 그 결과 유사판단부는 제1 서적과 제2 서적 간의 유사도가 제1 서적과 제3 서적간의 유사도 보다 판단할 수 있다. 그룹들 간에도 코사인유사도 연산을 통하여 유사도가 도출될 수 있다.
- [0050] 다음으로 서적추천부(150)는 상기 유사판단부를 통해 도출된 서적들 간의 유사도를 바탕으로 서적을 추천하는 역할을 할 수 있다. 앞서 설명한 바와 같이, 키워드 기반으로 측정된 제1 유사도, 테마기반으로 측정된 제2 유사도, 그룹 기반으로 측정된 제3 유사도 등을 바탕으로 유사도 순으로 추천서적을 출력할 수 있다.
- [0051] 제1 유사도는 키워드를 기반으로 하며 기설정된 테마키워드데이터베이스를 참조하지 않는 방식이므로 서적추천부는 서적의 장르를 불문하고 유사한 서적을 찾을 수 있다. 이에 반해 제2 유사도는 서적의 테마, 즉 장르 기반으로 하며, 더욱 상세하게는 기디렉토리화된 테마가 저장된 테마키워드데이터베이스를 활용하여 연산되기 때문에 서적추천부는 서적의 장르에 기초하여 유사한 서적을 찾을 수 있다. 또한, 제3 유사도는 서적 내용에 대한 LDA 알고리즘 기반의 그룹에 기초하여 유사도를 판단하는 것이므로 서적추천부는 키워드나 장르에 한정되지 않고 서적 간의 유사도를 도출할 수 있다.
- [0052] 그러므로 서적추천부는 사용자의 취향이나 니즈, 그 동안의 사용자의 독서 사용행태, 서적의 특징 등을 고려하여 제1 유사도, 제2 유사도, 제3 유사도를 혼합 적용하여 최종적으로 서적추천할 수 있다. 예를 들어, 제1 서적이 전문의학적적으로 대부분 전문용어를 포함하는 서적인 경우, 의학용어에 특징이 있으므로 키워드 기반의 제1 유사도의 가중치를 높여서 유사서적들을 추천할 수 있다. 이에 반해 제2 서적이 소설로 장르성이 강한 서적인 경우, 테마, 즉, 장르 기반의 제2 유사도의 가중치를 높여서 유사서적을 추천할 수 있다. 이와 같이 서적의 특징이나 장르에 기초하여 서적추천부는 제1 유사도 내지 제3 유사도의 가중치를 조정하여 최종적으로 서적을 추천할 수 있다.
- [0053] 한편, 제어부는 앞서 설명한 바와 같이 서적의 장르 등 서적에 종속되는 정보와 사용자 행태 등 사용자에게 종속되는 정보를 바탕으로 제1 유사도 또는 제3 유사도를 선택할 수 있다. 제어부가 선택한 방식에 따라 유사판단부가 서적의 유사도를 연산하고 서적추천부가 추천함으로써 최종적으로 가장 적절한 유사 서적을 도출할 수 있는 것이다. 서적의 정보는 서적의 장르, 서적의 가격정보, 서적의 페이지 수, 서적의 출판연도, 서적의 저자 국적, 저자의 수, 서적 출판사, 서적 제목, 인쇄횟수, 재판횟수, 판매서점, 가격할인률 등 다양한 정보가 될 수 있다. 또한, 사용자의 정보는 사용자 나이, 사용자 성별, 사용자 직업, 사용자 종교, 사용자의 제1 언어, 사용자 주거 지역, 사용자 온라인 웹사이트 방문 기록, 사용자 대학 전공, 사용자 가족관계, 사용자 수입 등이 될 수 있다.
- [0055] 다음으로, 서적추천부는 제1 유사도 내지 제3 유사도 뿐만 아니라 사용자 행동을 반영하여 서적을 추천할 수 있다. 구체적으로, 서적추천부는 사용자에게 의해 소정의 기준값보다 높은 평가점수를 부여받은 서적, 온라인 쇼핑 물의 장바구니에 포함된 서적 또는 사용자에게 의해 온라인 공간상에서 서적 정보가 열람된 적이 있는 서적과 유사한 서적들을 선택하여 사용자에게 추천하는 역할을 할 수 있다. 상기 역할은 서적추천부 내의 유사서적추출모듈(미도시)에서 할 수 있다.
- [0056] 더욱 구체적으로 서적추천시스템은 궁극적으로 사용자가 흥미를 느낄만한 서적을 추천하는 역할을 한다. 이때 서적추천시스템의 서적추천부는 사용자의 행동패턴 등을 분석하여 이와 관련된 서적에 대한 유사한 서적들을 추천할 수 있다. 더더욱 구체적으로 사용자의 개인화된 정보, 사용자 맞춤형 추천 서적을 제공하기 위해 사용자의

온라인상의 행동을 파악하여 추천서적 선택에 활용하는 것이다. 예를 들어, 사용자가 온라인 웹사이트 등에서 특정 서적에 대해서 높은 평가점수를 부여한 경우, 상기 서적과 유사한 서적들을 유사판단부를 통해 선별하여 추천한다. 또한, 사용자가 아마존(등록상표) 등의 온라인 서점에서 특정 서적을 구매하였거나 구매를 하기 위해 장바구니에 담은 적이 있는 경우, 상기 특정 서적과 유사한 서적들을 선택하여 사용자에게 추천정보를 제공할 수 있다. 또한, 사용자가 온라인 커뮤니티나 평론 사이트에서 특정 서적에 대한 서평을 열람한 적이 있는 경우, 상기 서적과 유사한 서적들을 선택할 수 있다. 상기 설명하는 유사한 서적들의 선택은 유사판단부를 통해 상기 언급한 제1 유사도 내지 제3 유사도를 바탕으로 소트 또는 분류된 유사서적들을 대상으로 진행될 수 있음은 물론이다.

[0057] 상기 외부 온라인 공간상의 웹사이트, 예를 들어, SNS 사이트, 온라인 서점 사이트, 온라인 서평 사이트 등의 다양한 서적 정보 및 서적에 대한 사용자의 행동 정보는 상기 사이트로부터 사용자행동추출부(170)를 통해 추출될 수 있다. 사용자행동추출부를 통해 추출된 사용자 행동정보는 사용자행동데이터베이스에 저장될 수 있다.

[0058] 또한, 서적추천부는 선택된 서적 중을 임의대로 추천할 수 있다. 앞서 서적추출부, 특히 유사서적추출모듈이 사용자의 행동 등을 고려하여 유사한 서적들을 선별하였다면, 서적추출부의 선별추천모듈(미도시)은 선별된 유사서적들을 기설정된 기준에 따라 최종적으로 출력하는 역할을 할 수 있다. 즉, 상기 출력 역할은 서적추천부 내의 선별추천모듈이 할 수 있다. 선별추천모듈은 유사서적추출모듈이 추출한 유사서적들 중 사용자가 이미 읽은 적이 있는 서적을 제외하고 남은 서적들을 추천할 수 있다. 또한, 사용자, 타사용자, 평론가 등의 평점이 낮은 서적들은 제외하고 남은 서적들을 추천할 수 있다. 혹은 사용자 정보를 고려하여 사용자가 읽기 부적절한 등급의 서적들은 제외하고 추천할 수 있다. 선별추천모듈은 유사서적들 중 기설정된 기준에 기초하여 추천하지 않을 서적들을 제외하는 역할을 한다. 한편, 선별추천모듈이 서적 선별을 위해 서적정보데이터베이스를 참조할 수 있다. 서적정보데이터베이스 내에 각 서적에 대한 다양한 정보가 포함될 수 있다. 또한, 선별추천모듈은 사용자행동데이터베이스와 서적정보데이터베이스를 연계하여 서적을 최종적으로 추천할 수 있다. 예를 들어, 특정 사용자가 이미 독서한 서적에 대한 정보는 사용자행동데이터베이스에 저장될 수 있기 때문이다.

[0059] 또한, 선별추천모듈은 유사서적추출모듈에서 추출한 서적들을 특정 기준에 따라 분류하여 출력할 수 있다. 하나 이상의 서적들을 서적 평점 순서에 따라 출력할 수 있다. 이 경우 서적정보데이터베이스를 이용하여 각 서적의 평점을 불러올 수 있고, 이를 활용한다. 또한, 유사서적 각각의 누적 판매량, 기간별 판매량 등의 정보에 기초하여 서적들을 소트하여 출력할 수 있다.

[0060] 또한, 사용자 개개인에 맞추어 개인화된 정보를 활용하여 유사서적들을 소트하여 출력할 수도 있다. 예를 들어, 사용자 개인의 정보 중 나이에 대한 정보를 이용하여 유사한 연령대의 사용자가 선호하는 서적들을 호감도 순서대로 출력할 수 있다. 또한, 그 밖에 개인의 정보 중 성별이나 성향 등을 고려하여 유사서적들 중 특정 유사서적을 먼저 출력하게 할 수 있다. 이러한 동작들은 사용자행동데이터베이스 내에 저장된 사용자 개인 정보를 이용하여 수행될 수 있다.

[0062] 데이터베이스를 살펴보면, 앞서 설명한 바와 같이 서적키워드데이터베이스는 각 서적을 대표하는 키워드집합을 포함하고 있고, 테마키워드데이터베이스는 테마별 키워드, 키워드별 테마 등 키워드와 테마의 관계에 대한 정보를 포함하고 있다. 키워드에 따른 테마 가중치를 포함할 수 있다. 그 역도 마찬가지이다. 서적테마데이터베이스는 각 서적을 대표하는 테마집합을 포함하고 있다. 서적그룹데이터베이스는 각 서적을 대표하는 그룹을 포함하고 있다. 사용자행동데이터베이스는 앞서 설명한 바와 같이 SNS 사이트, 온라인 서점 사이트, 온라인 서평 사이트 등의 다양한 서적 정보 및 서적에 대한 사용자의 행동 정보를 포함할 수 있다. 마지막으로 서적정보데이터베이스는 서적에 관련된 정보 즉, 서적 제목, 서적 저자, 서적 출판연도 등 서적에 관한 정보를 포함할 수 있다.

[0064] 도 2 내지 도 5는 본 발명에 따른 동작 순서도들이다. 도 2 내지 도 5의 본 발명에 따른 서적추천시스템 설명과 중복된 설명은 제외하고 설명한다. 도 2를 참조하면, 서적추천서버가 서적 데이터를 입력받는다(S810). 다음으로, 서적추천서버는 TF-IDF 방식으로 서적 데이터로부터 키워드를 추출한다(S830). 다음으로, 서적추천서버는 추출된 키워드집합들 간의 유사도를 연산한다(S850). 이후 서적추천서버는 유사도 연산 등에 기초하여 추천 서적들을 출력한다(S870).

[0065] 도 3을 참조하면, S810 및 S830 단계는 동일하며, 이후 단계가 상이하다. 서적추천서버가 추출된 키워드들에 따른 서적테마를 배정한다(S835). 이후 서적추천서버가 배정된 서적테마들 간의 유사도를 연산한다(S837). 이후 연산된 유사도 등을 기초로 하여 추천 서적을 출력한다(S870).

[0066] 도 4를 참조하면, S810 단계는 동일하며, 이후 단계가 상이하다. 먼저 LDA 방식으로 서적의 내용을 그루핑한다

(S840). 이후 그룹핑된 서적 내 그룹 간의 유사도를 연산한다(S845). 이후 연산된 유사도 등을 기초로 하여 추천 서적을 출력한다(S870).

[0067] 도 5를 참조하면, S810 단계는 동일하며, 이후 단계가 상이하다. 먼저 서적의 특징이나 사용자의 니즈에 따라 어떠한 방식으로 서적 간의 유사도를 연산할지 판단한다(S820). TF-IDF 방식을 선택하는 경우, TF-IDF 방식에 따라 키워드 집합 간의 유사도를 측정하여 양서적 간의 유사도를 측정한다. 이에 반해 서적의 특징이나 사용자의 니즈에 따라 LDA 방식을 선택하는 경우, LDA 방식에 따라 서적의 유사도를 측정하여 출력한다.

[0069] 도 6은 본 발명의 다른 실시예에 따른 서적추천시스템의 구성도이다. 도 6을 참조하면, 서적추천시스템은 제어부(110), 감정그래프생성부(220), 감정그래프유사판단부(230), 서적추천부(240), 입출력부(250), 감정단어데이터베이스(201), 감정그래프데이터베이스(203), 사용자행동데이터베이스(205), 서적정보데이터베이스(207)이 도시되어 있다. 본 서적추천시스템은 서적추천서버가 될 수 있다.

[0070] 먼저, 제어부(110)는 감정그래프생성부, 감정그래프유사판단부, 서적추천부, 입력부, 출력부 등 구성들을 제어하는 역할을 한다. 예를 들어, 사용자로부터 제어신호 등을 입력받아 제어신호에 따라 감정그래프를 생성하도록 감정그래프생성부를 제어하는 역할을 할 수 있다. 제어부가 제어부에 연결된 각각의 데이터베이스들을 제어할 수 있음은 물론이다. 도 6에서는 각종 데이터베이스가 서적추천시스템인 서버 내에 포함되어 있는 것을 도시되어 있으나 상기 데이터베이스들이 서버 외부에 존재하여도 무방하다. 또한, 본 발명의 상세한 설명 및 청구항에 서는 서적에 대해서 언급하고 있으나 본 발명의 적용은 서적에 한정된 것이 아니며, 웹소설, 만화 등 활자를 포함하는 구별된 텍스트에 대해서는 모두 적용이 가능하다.

[0071] 다음으로, 감정그래프생성부(220)는 서적들의 스토리플로우에 대한 감정그래프를 생성하는 역할을 한다. 스토리플로우는 서적의 첫 장에서 마지막 장까지의 시간순서에 따른 이야기 전개를 의미할 수 있다. 일반적으로 설명문이나 논설문을 제외한 소설책이나 수필, 기행문 등은 문학적 요소를 포함하고 전체적으로 본문 내에 감정의 흐름이 존재한다. 예를 들어, 윌리엄 셰익스피어의 소설 '로미오와 줄리엣'의 경우, 양 가문의 다툼, 로미오와 줄리엣의 사랑, 줄리엣과 패리스의 결혼 강요, 로미오와 패리스의 결투, 로미오의 자살, 줄리엣의 자결, 양 가문의 화해 등의 여러 가지 감정이 포함된 스토리플로우를 갖는다. 구체적으로, 감정 중 일종이 행복이라는 감정의 관점에서 검토하면, 소설의 초기의 양 가문의 다툼에서는 감정 정도가 낮고, 중반부에는 로미오와 줄리엣의 사랑에서는 감정이 강하며, 종반에서는 결투로 인해 감정 정도가 낮으나 최종부에는 화해로 인하여 다시 행복이라는 감정이 올라가게 된다. 이와 같이 소설이나 수필과 같은 문학적 요소가 포함된 서적의 경우 스토리플로우에 따른 감정변화가 존재한다. 감정그래프생성부는 이런 감정변화를 감정그래프로 생성하는 역할을 한다.

[0072] 또한, 감정그래프생성부는 서적정보데이터베이스에 저장되어 있을 수 있는 서적의 장르정보를 이용하여 스토리플로우에 따른 감정그래프가 생성되어야 할 서적을 선별할 수 있다. 더욱 구체적으로 감정그래프생성부는 기획된 서적의 장르정보가 소설인 경우에는 감정그래프생성을 위해 감정그래프생성부에 서적의 스토리플로우를 입력한다. 이에 반해, 서적의 장르정보가 논설문인 경우, 추출할 감정이 없다고 판단하여 감정그래프생성을 시도하지 않는다.

[0073] 또한, 감정그래프생성부는 윈도우(window)를 통해 서적 내 감정을 분석할 수 있다. 도 7은 본 발명에 따른 스토리플로우 감정그래프 생성을 설명하기 위한 개념도이다. 도 7을 참조하면, 서적의 스토리플로우를 구성하는 단어들(단어)을 메모리(미도시)에 입력하고, 본 메모리에 저장된 단어들(단어)을 부분 부분으로 나누어 순차적으로 읽어드려 감정 분석을 수행한다. 더욱 구체적으로 예를 들어, 서적이 만 개의 단어들로 이루어진 소설이라고 가정한다. 이만 개의 단어들로 이루어진 서적에 대해 백 개의 단어 단위로 감정을 분석하며 이때 감정분석의 대상이 되는 단어들(단어)을 그룹을 윈도우라고 칭할 수 있다. 이 윈도우를 서적의 첫 단어부터 마지막 단어 순서의 방향으로 이동시켜가며, 윈도우에 포함된 단어들의 감정값을 산출할 수 있다.

[0074] 윈도우의 크기는 사용자가 임의로 정할 수 있다. 본 발명의 일 실시예에서는 서적의 전체 단어 수의 1%에 해당하는 수만큼의 단어들(단어)을 윈도우할 수 있는 크기의 윈도우를 정하였다. 예를 들어, 전체 소설에 만 개의 단어들(단어)이 포함되어 있다면 윈도우는 백 개의 단어들(단어)을 읽어들이 수 있다. 물론 윈도우의 크기가 상기 예에 한정되는 것은 아니며 다양한 크기의 윈도우를 가질 수 있다.

[0075] 윈도우의 크기는 서적의 페이지에 포함된 단어들 개수의 평균치의 단어들(단어)을 갖도록 하여도 무방하다. 이를 통해 윈도우의 크기는 서적 각 페이지 정도가 될 수 있다.

[0076] 또한, 윈도우는 서적의 스토리플로우 방향으로 이동시킬 수 있다. 예를 들어, 만 개의 단어들(단어)로 이루어진 서적에 백 개의 단어들(단어)을 읽어 드릴 수 있는 윈도우를 적용하여 이동시킨다고 가정한다. 최초 윈도우의 위치는 첫 번째

단어에서 백 번째 단어에 위치하게 되며, 최초 윈도우가 위치한 곳에 대한 서적의 내용을 분석하여 감정을 추출할 수 있다. 이후 윈도우를 스토리플로우 방향으로 한 단어씩 이동시킨다. 그 결과, 윈도우가 두 번째 단어에서 백일 번째 단어에 위치하게 되며, 이에 해당하는 단어들의 감정을 분석할 수 있다. 이러한 과정을 반복하여 윈도우가 마지막 구천구백일 번째 단어에서 만 번째 단어에 위치하게 되며, 이에 해당하는 단어들의 감정을 분석할 수 있다. 이와 같은 과정을 통해 구천구백일 번의 감정 분석이 수행되어 서적 전체에 대한 감정분석이 완료될 수 있다. 한편, 첫 번째 감정분석 결과값부터 구천구백일 번째 감정분석 결과값은 한 단어의 추가 및 삭제가 반복되는 것이므로 그 결과값이 부드럽게 변화할 수 있다. 위에 대한 내용을 정리하면 다음과 같다.

- [0078] 전체 서적의 단어 개수 = n
- [0079] 윈도우의 크기 = 윈도우에 포함될 수 있는 단어 개수 = m
- [0080] 서적 내 제 a 위치의 윈도우의 분석대상단어 = 서적 내 a 번째 단어부터 $a+m-1$ 번째 단어
- [0081] 서적의 스토리플로우 순서로 n 번째에 위치하는 단어 = 단어(n)
- [0083] 예시)
- [0084] 전체 서적의 단어 개수 = 10,000
- [0085] 윈도우의 크기 = 100
- [0086] 제1 위치의 윈도우의 분석대상단어들 = 단어 (1) 내지 단어 (100)
- [0087] 제2 위치의 윈도우의 분석대상단어들 = 단어 (2) 내지 단어 (101)
- [0088]
- [0089] 제99,901 위치의 윈도우의 분석대상단어들 = 단어 (99,901) 내지 단어 (10,000)
- [0091] 한편, 윈도우의 이동은 한 단어 단위로 이동할 수 있지만 이에 한정된 것은 아니며 한 페이지씩 이동하여 감정값을 연산하여도 무방하다. 이 경우 윈도우 내에 단어가 전면 교체되므로 감정값의 변화가 다소 클 수 있다. 예를 들어, 윈도우의 크기를 서적의 한 페이지 크기로 설정하고, 이동 단위도 한 페이지로 설정하는 경우, 각 페이지에 대한 감정 결과값이 순차적으로 도출될 수 있다. 또한, 감정값이 이산(discrete)하게 분포될 수 있다.
- [0093] 다음으로, 윈도우의 감정값 연산 방법에 대해 설명한다. 감정그래프생성부는 윈도우에 포함된 단어들을 분석하여 감정 정도를 산출한다. 구체적으로 윈도우 내 단어들 중 특정 감정에 연관된 단어가 있는지 검색하고, 검색된 단어에 부여된 점수를 합산하여 특정 감정 정도에 대한 값을 산출한다. 도 8은 감정단어 데이터베이스(101)에 포함된 행복(happiness)이라는 감정을 나타내는 단어들의 감정값 테이블이다. 행복이라는 감정값 테이블을 살펴보면, '웃음'이라는 단어는 감정값이 4이고, '행복'이라는 감정값은 '3'이 될 수 있다. 반드시 양수의 감정값만을 갖는 것은 아니며 음수의 감정값을 갖는 단어도 존재한다. 예를 들어, '미움'이라는 단어는 감정값이 '-1'이고, 복수라는 단어는 감정값이 '-3'이다. 감정값의 표현은 반드시 양수 또는 음수에 한정될 필요는 없고 감정값의 정도를 표현할 수 있으면 족하다. 예를 들어 상, 중, 하로 나타낼 수도 있다.
- [0094] 포함된 단어들이 감정값 테이블에 존재하는 경우, 각 단어에 따른 감정값을 합산하여 최종 단위감정합산값을 산출한다. 즉, 단위감정합산값이란 하나의 윈도우에 포함된 단어들의 감정값을 합산한 것을 의미할 수 있다. 도 9는 감정값이 부여된 단어들이 존재하는 페이지에 관한 것이다. 예를 들어, 도 9의 페이지에서 행복이라는 감정에 관한 단어인 '웃음', '사랑', '행복', '증오'가 검색된다. 이에 각 감정에 관한 단어의 감정값을 합산하면 본 페이지의 단위감정합산값은 웃음(4)+사랑(2)+행복(3)+증오(-2)로 7점으로 산출된다. 이런 단위감정합산값을 스토리플로우 방향으로 반복하여 연산하면 행복이라는 감정에 대한 서적의 감정그래프가 도출될 수 있다. 도 5는 이런 스토리플로우 방향에 따른 단위감정값을 그래프로 나타낸 것이다.
- [0095] 본 예시에서는 행복이라는 감정에 대해서만 감정그래프를 생성하였으나 이에 한정되지 않으며, 행복, 놀람, 후회, 고독, 두려움 등 다양한 감정그래프가 생성될 수 있다. 이를 위해 감정단어 데이터베이스에 다양한 감정테이블이 존재할 수 있으며, 상기 감정테이블에는 다양한 단어가 존재할 수 있다. 또한, 감정값을 연산할 때 단어의 동일 여부에 한정될 필요는 없으며, 뜻의 나타내는 최소 단위인 형태소 단위로 감정값을 산출하여도 무방하다.
- [0096] 감정그래프 유사판단부(230)는 각 서적의 감정그래프 간의 유사도를 판단하는 역할을 한다. 더욱 구체적으로 제 1 서적의 단위감정합산값과 제2 서적의 단위감정합산값 간의 차이를 나타내는 제1 유사판단기초값 및 상기 제1

유사판단기초값의 변화량을 나타내는 제2 유사판단기초값을 기초로 하여 제1 서적과 제2 서적 간의 유사도를 판단한다. 또한, 앞서 언급한 바와 같이 하나의 감정에 한정되는 것이 아니라 여러 개의 감정 그래프들 간에 유사도를 측정하여 최종적으로 양 서적 간의 유사도를 도출할 수 있다.

[0097] 도 11은 서적 1 내지 4 각각의 구간별 단위감정합산값을 나타낸 테이블이며, 도 12는 이를 그래프로 나타낸 것이다. 도 11 및 도 12를 참조하면, 서적1과 서적2 간의 구간별 단위감정합산값의 차이는 도 13과 같이 나타난다. 제1 유사판단기초값은 제1 서적의 단위감정합산값과 스토리방향에 따른 상기 제2 서적의 단위감정합산값 간의 차이를 의미하며, 제2 유사판단기초값은 상기 제1 유사판단기초값의 변화량을 나타낸다. 도 8을 참조하여, 서적 1과 서적 2 간의 제1 유사판단기초값은 '1, 3, 5, 7, 9, 11, 8, 6, 4, 2, 1, 1, 3, 5'와 같이 나열된다. 한편, 서적 1과 서적 3 간의 제1 유사판단기초값은 전부 '1'이다. 감정그래프유사판단부는 양 서적의 단위감정합산값의 차이가 작을수록 유사하다고 판단한다. 즉, 제1 유사판단기초값이 작을수록 양 서적이 유사하다고 판단할 수 있다. 그러므로 도 13을 참조하면, 감정그래프유사판단부는 서적 2에 비해 서적 3은 서적 1에 더 유사하다고 판단할 수 있다. 또한, 제2 유사판단기초값을 검토해보면, 서적1과 서적2 간의 제1 유사판단기초값의 변화량이 크므로 제2 유사판단기초값도 크다. 이에 반해 서적1과 서적3 간의 제1 유사판단기초값의 변화량은 없으므로 제2 유사판단기초값은 '0'이다. 그러므로 감정그래프유사판단부는 제2 유사판단기초값이 작은 서적 1과 서적 3이 서적1과 서적2에 비해 더 유사하다고 판단할 수 있다. 감정그래프유사판단부는 제1 유사판단기초값 및 제2 유사판단기초값 중 어느 하나 이상을 혼합하여 양 서적 간의 유사판단을 할 수 있다. 또한, 유사판단기초값에 서적의 성격 등에 따라 가중치를 부여하여 두 개의 서적 간의 유사도를 측정할 수 있다.

[0098] 서적 1과 서적 3 및 서적 1과 서적 4 간에는 제1 유사판단기초값의 변화가 없기 때문에 제2 유사판단기초값은 '0'이 된다. 그러나 제1 유사판단기초값이 차이가 있으므로 감정그래프유사판단부는 값이 작은 서적 1과 서적 3 간에 서적 1과 서적 4보다 더 유사하다고 판단할 수 있다.

[0099] 서적추천부(140)는 사용자에게 의해 소정의 기준값보다 높은 평가점수를 부여받은 서적, 온라인 쇼핑몰의 장바구니에 포함된 서적 또는 사용자에게 의해 온라인 공간상에서 서적 정보가 열람된 적이 있는 서적과 유사한 서적들을 선택하여 사용자에게 추천하는 역할을 한다. 상기 역할은 서적추천부 내의 유사서적추출모듈(미도시)에서 할 수 있다.

[0100] 더욱 구체적으로 서적추천시스템은 궁극적으로 사용자가 흥미를 느낄만한 서적을 추천하는 역할을 한다. 이때 서적추천시스템의 서적추천부는 사용자의 행동패턴 등을 분석하여 이와 관련된 서적에 대한 유사한 서적들을 추천할 수 있다. 더더욱 구체적으로 사용자의 개인화된 정보, 사용자 맞춤형 추천 서적을 제공하기 위해 사용자의 온라인상의 행동을 파악하여 추천서적 선택에 활용하는 것이다. 예를 들어, 사용자가 온라인 웹사이트 등에서 특정 서적에 대해서 높은 평가점수를 부여한 경우, 상기 서적과 유사한 서적들을 감정그래프유사판단부를 통해 선별하여 추천한다. 또한, 사용자가 아마존(등록상표) 등의 온라인 서점에서 특정 서적을 구매하였거나 구매를 하기 위해 장바구니에 담은 적이 있는 경우, 상기 특정 서적과 유사한 서적들을 선택하여 사용자에게 추천정보를 제공할 수 있다. 또한, 사용자가 온라인 커뮤니티나 평론 사이트에서 특정 서적에 대한 서평을 열람한 적이 있는 경우, 상기 서적과 유사한 서적들을 선택할 수 있다.

[0101] 상기 외부 온라인 공간상의 웹사이트, 예를 들어, SNS 사이트, 온라인 서점 사이트, 온라인 서평 사이트 등의 다양한 서적 정보 및 서적에 대한 사용자의 행동 정보는 상기 사이트로부터 사용자행동추출부(160)를 통해 추출될 수 있다. 사용자행동추출부를 통해 추출된 사용자 행동정보는 사용자행동데이터베이스에 저장될 수 있다.

[0102] 또한, 서적추천부는 선택된 서적 중을 임의대로 추천할 수 있다. 앞서 서적추출부, 특히 유사서적추출모듈이 사용자의 행동 등을 고려하여 유사한 서적들을 선별하였다면, 서적추출부의 선별추천모듈(미도시)은 선별된 유사서적들을 기설정된 기준에 따라 최종적으로 출력하는 역할을 한다. 즉, 상기 출력 역할은 서적추천부 내의 선별추천모듈이 할 수 있다. 선별추천모듈은 유사서적추출모듈이 추출한 유사서적들 중 사용자가 이미 읽은 적이 있는 서적을 제외하고 남은 서적들을 추천할 수 있다. 또한, 사용자, 타사용자, 평론가 등의 평점이 낮은 서적들은 제외하고 남은 서적들을 추천할 수 있다. 혹은 사용자 정보를 고려하여 사용자가 읽기 부적절한 등급의 서적들은 제외하고 추천할 수 있다. 선별추천모듈은 유사서적들 중 기설정된 기준에 기초하여 추천하지 않을 서적들을 제외하는 역할을 한다. 한편, 선별추천모듈이 서적 선별을 위해 서적정보데이터베이스를 참조할 수 있다. 서적정보데이터베이스 내에 각 서적에 대한 다양한 정보가 포함될 수 있다. 또한, 선별추천모듈은 사용자행동데이터베이스와 서적정보데이터베이스를 연계하여 서적을 최종적으로 추천할 수 있다. 예를 들어, 특정 사용자가 이미 독서한 서적에 대한 정보는 사용자행동데이터베이스에 저장될 수 있기 때문이다.

[0103] 또한, 선별추천모듈은 유사서적추출모듈에서 추출한 서적들을 특정 기준에 따라 분류하여 출력할 수 있다. 하

나 이상의 서적들을 서적 평점 순서에 따라 출력할 수 있다. 이 경우 서적정보데이터베이스를 이용하여 각 서적의 평점을 불러올 수 있고, 이를 활용한다. 또한, 유사서적 각각의 누적 판매량, 기간별 판매량 등의 정보에 기초하여 서적들을 소트하여 출력할 수 있다.

[0104] 또한, 사용자 개개인에 맞추어 개인화된 정보를 활용하여 유사서적들을 소트하여 출력할 수도 있다. 예를 들어, 사용자 개인의 정보 중 나이에 대한 정보를 이용하여 유사한 연령대의 사용자가 선호하는 서적들을 호감도 순서대로 출력할 수 있다. 또한, 그 밖에 개인의 정보 중 성별이나 성향 등을 고려하여 유사서적들 중 특정 유사서적을 먼저 출력하게 할 수 있다. 이러한 동작들은 사용자행동데이터베이스 내에 저장된 사용자 개인 정보를 이용하여 수행될 수 있다.

[0105] 도 14 내지 17은 본 발명에 따른 동작 순서도들이다. 도 1 내지 도 13의 본 발명에 따른 서적추천시스템 설명과 중복된 설명은 제외하고 설명한다. 도 14를 참조하면, 서적추천서버가 서적들의 스토리플로우에 따른 감정그래프를 생성한다(S910). 다음으로 서적추천서버가 사용자가 관심을 두는 서적(흥미서적)과 타서적들 간의 유사도를 판단한다(S930). 앞서 설명한 바와 같이 감정그래프의 유사도를 측정하여 유사도를 판단할 수 있다. 상기 유사도 연산을 통해 유사한 서적들을 추천 서적으로 출력한다(S950).

[0106] 도 15는 스토리플로우에 따른 감정그래프를 생성하는 단계를 더 분석하여 나타낸 순서도이다. 도 15를 참조하면, 서적추천서버가 서적의 장르정보를 이용하여 스토리 존재 여부를 판단한다(S911). 다음으로, 서적추천서버는 서적을 소정의 크기로 구간을 나눈다(S913). 다음으로, 서적추천서버는 각 구간에 존재하는 감정 단어의 출현을 카운팅한다(S915). 이후 서적추천서버는 감정 단어별로 가중치에 따라 연산하여 구간에 따른 단위감정합산값을 연산하고 상기 값에 따라 감정그래프를 생성한다(S917). 상기 그래프에 대한 정보를 감정그래프데이터베이스에 저장한다(S919).

[0107] 도 16은 사용자에게 추천할 서적을 추출하기 위한 전처리 단계를 나타낸 도면이다. 도 16을 참조하면, 서적추천서버는 사용자로부터 호평을 받은 서적, 사용자가 온라인 쇼핑몰의 장바구니에 담은 적이 있는 서적, 혹은 온라인 쇼핑몰에서 '찜하기' 기능으로 찜이 된 적이 있는 서적, 사용자가 서적 정보를 열람한 적이 있는 서적 등을 확인한다(S920). 서적추천서버는 이런 서적을 사용자 흥미서적이라고 판단하고 상기 흥미서적과 타서적들 간의 유사도를 분석한다(S930). 서적추천서버는 상기 유사도를 바탕으로 유사한 서적을 출력한다(S950).

[0108] 도 17은 유사도 연산 등에 기초하여 추천 서적 출력하는 단계를 분석하여 설명한 순서도이다. 도 17을 참조하면, 서적추천서버가 추천할 유사한 서적들 리스트를 추출한다(S951). 서적추천서버는 상기 추출된 서적들 중 사용자행위정보, 예를 들어, 사용자가 이미 읽은 적이 있는 서적정보를 적용하여 유사한 서적들 리스트를 필터링한다(S953). 사용자행위정보는 성별, 나이, SNS 사용 행태 등 다양한 정보가 포함될 수 있다. 다음으로, 서적추천서버는 서적정보, 예를 들어, 평점정보 또는 판매량정보 등을 서적정보를 적용하여 유사한 서적들 리스트를 필터링한다(S955). 이후 필터링된 서적을 최종적으로 출력한다(S957). 한편, 필터링으로 표현을 했으나 필터링은 서적을 소팅한다는 개념으로 이해될 수 있다. 또한, S951 단계와 S955 단계는 순차적인 것으로 한정될 필요 없고 그 순서가 바뀌어도 무방하다.

[0110] 도 18은 본 발명의 다른 실시예에 따른 서적추천시스템의 구성도이다. 앞서 설명과 중복된 설명은 제외하고 설명한다. 도 18을 참조하면 본 서적추천시스템은 TF-IDF부, LDA부, 감정그래프생성부에서 추출된 각종 유사도 정보, 예를 들어, 서적의 키워드간 유사도, 서적의 테마간 유사도, 서적의 그룹별 유사도, 서적의 스토리 플로우간의 유사도 등을 측정하여 사용자에게 추천할 수 있다.

[0112] 위와 같은 서적추천시스템 및 서적추천방법을 통하여 사용자에게 최적의 서적들을 추천할 수 있다.

[0114] 상술된 방법 및 처리는, 예를 들어, 프로세서, 컨트롤러, 또는 다른 처리 디바이스에 의한 실행을 위한 명령들로서, 인코딩되거나, 콤팩트 디스크 판독 전용 메모리 (CDROM), 자기 또는 광학 디스크, 플래시 메모리, 랜덤 액세스 메모리 (RAM) 또는 판독 전용 메모리 (ROM), 소거가능 프로그램가능 판독 전용 메모리(EPROM) 또는 다른 머신-판독가능 매체와 같은 머신 판독가능 또는 컴퓨터 판독가능 매체에 저장될 수도 있다.

[0115] 이러한 매체는, 명령 실행가능 시스템, 장치 또는 디바이스에 의해 또는 이와 연결하여 사용하기 위한 실행가능한 명령들을 포함, 저장, 통신, 전파 또는 이동시키는 임의의 디바이스로서 구현될 수도 있다. 대안으로 또는 추가적으로, 하나 이상의 집적 회로, 또는 하나 이상의 프로세서 실행 명령들과 같은 하드웨어를 이용하여 아날로그 또는 디지털 로직으로서; 또는 API (application programming interface) 또는 DLL (Dynamic Link Library), 로컬 또는 원격 절차 호출로서 정의된 또는 공유 메모리에서 이용가능한 기능들의 소프트웨어로; 또는 하드웨어 및 소프트웨어의 조합으로서 구현될 수도 있다.

[0116] 다른 구현에서, 방법은 신호 또는 전파-신호 매체로 나타내어질 수도 있다. 예를 들어, 임의의 소정의 프로그램의 로직을 구현하는 명령들은 전기, 자기, 광학, 전자기, 적외선 또는 다른 타입의 신호의 형태를 취할 수도 있다. 상술된 시스템은 광섬유 인터페이스, 안테나, 또는 다른 아날로그 또는 디지털 신호 인터페이스와 같은 통신 인터페이스에서 이러한 신호를 수신하고, 그 신호로부터 명령들을 복원하고, 이들을 머신 판독 가능 메모리에 저장하고, 그리고/또는 프로세서를 이용하여 이들을 실행시킬 수도 있다.

[0117] 이상, 본 발명의 일 실시예에 대하여 설명하였으나, 해당 기술 분야에서 통상의 지식을 가진 자라면 특허청구범위에 기재된 본 발명의 사상으로부터 벗어나지 않는 범위 내에서, 구성 요소의 부가, 변경, 삭제 또는 추가 등에 의해 본 발명을 다양하게 수정 및 변경시킬 수 있을 것이며, 이 또한 본 발명의 권리범위 내에 포함된다고 할 것이다.

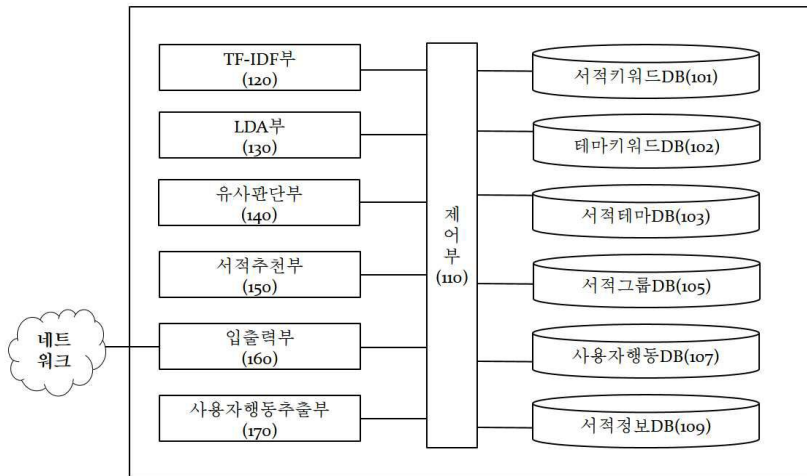
부호의 설명

[0119] 100, 200, 300 : 서적추천시스템

- 110 : 제어부
- 120 : TF-IDF부
- 130 : LDA부
- 140 : 유사판단부
- 150 : 서적추천부
- 160 : 입출력부
- 170 : 사용자행동추출부
- 101 : 서적키워드데이터베이스
- 102 : 테마키워드데이터베이스
- 103 : 서적테마데이터베이스
- 105 : 서적그룹데이터베이스
- 107 : 사용자행동데이터베이스
- 109 : 서적정보데이터베이스
- 220 : 감정그래프생성부
- 230 : 감정그래프유사판단부
- 240 : 서적추천부
- 250 : 입출력부
- 260 : 사용자행동추출부
- 201 : 감정단어데이터베이스
- 203 : 감정그래프데이터베이스
- 205 : 사용자행동데이터베이스
- 207 : 서적정보데이터베이스

도면

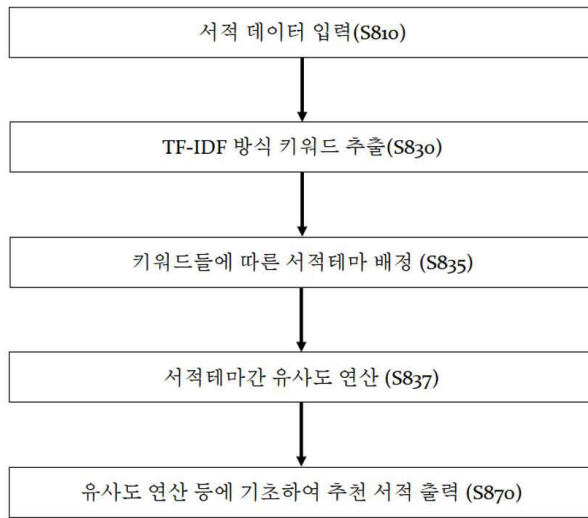
도면1



도면2



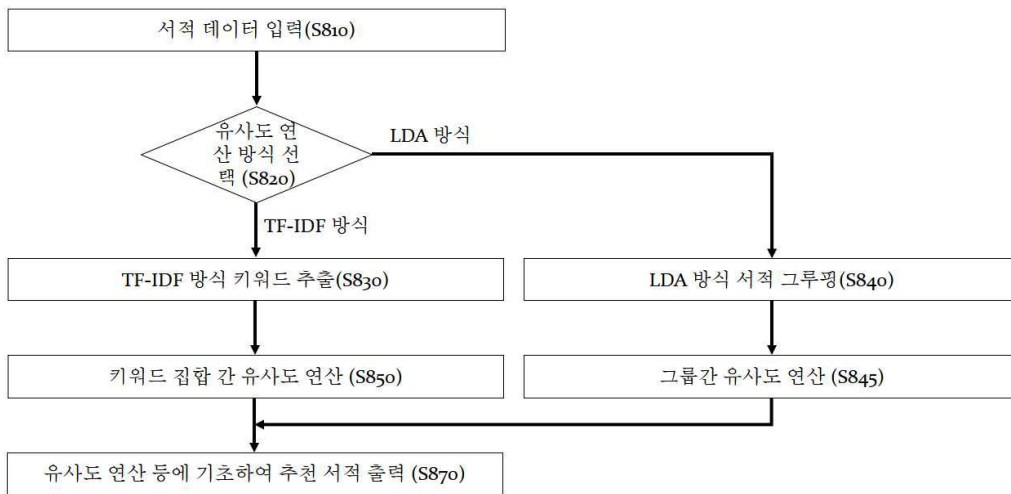
도면3



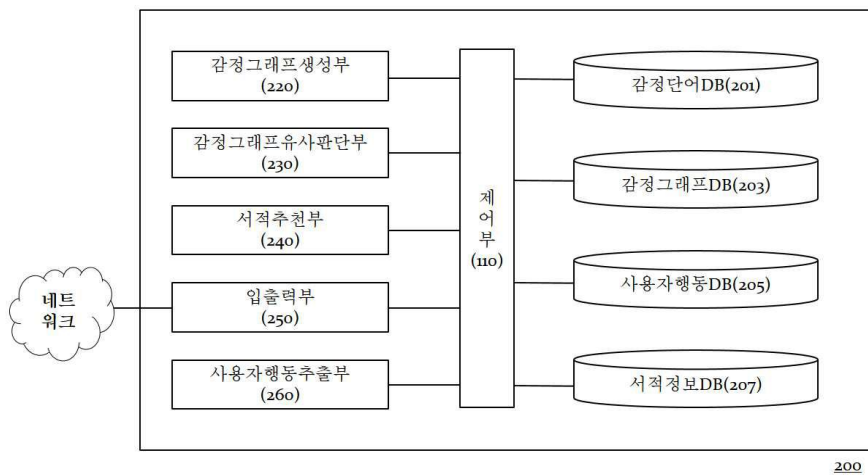
도면4



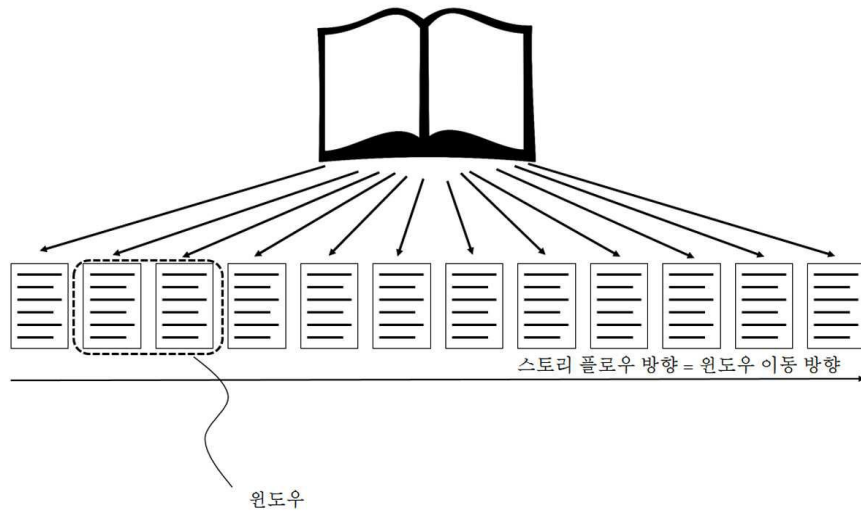
도면5



도면6



도면7



도면8

단어	감정값
웃음	4
행복	3
사랑	2
.	.
.	.
.	.
.	.
.	.
미움	-1
증오	-2
복수	-3

도면9

웃음

.....

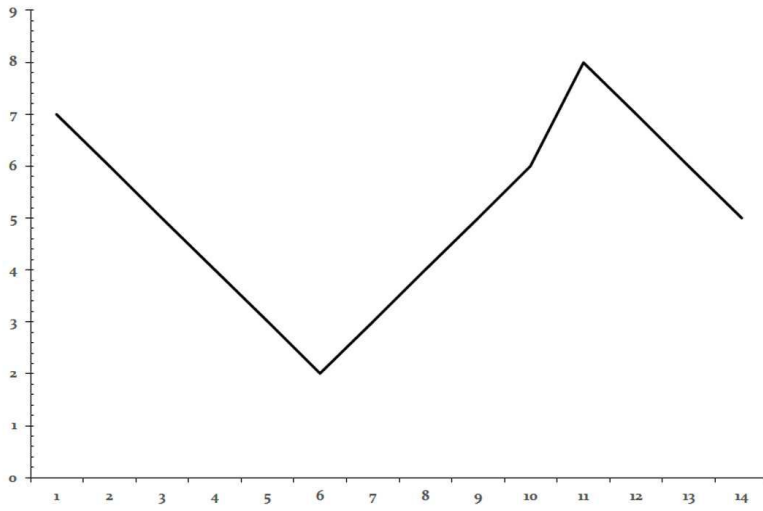
.....사랑.....

.....

.....행복.....

.....증오.....

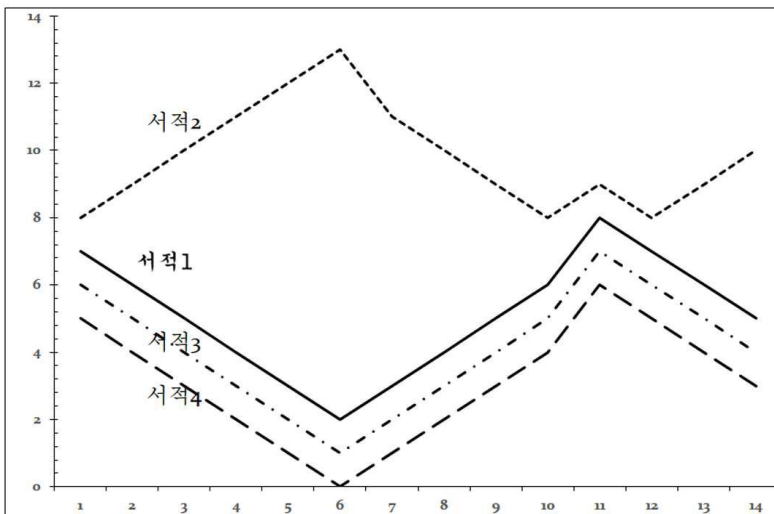
도면10



도면11

구간	서적1	서적2	서적3	서적4
1	7	8	6	5
2	6	9	5	4
3	5	10	4	3
4	4	11	3	2
5	3	12	2	1
6	2	13	1	0
7	3	11	2	1
8	4	10	3	2
9	5	9	4	3
10	6	8	5	4
11	8	9	7	6
12	7	8	6	5
13	6	9	5	4
14	5	10	4	3

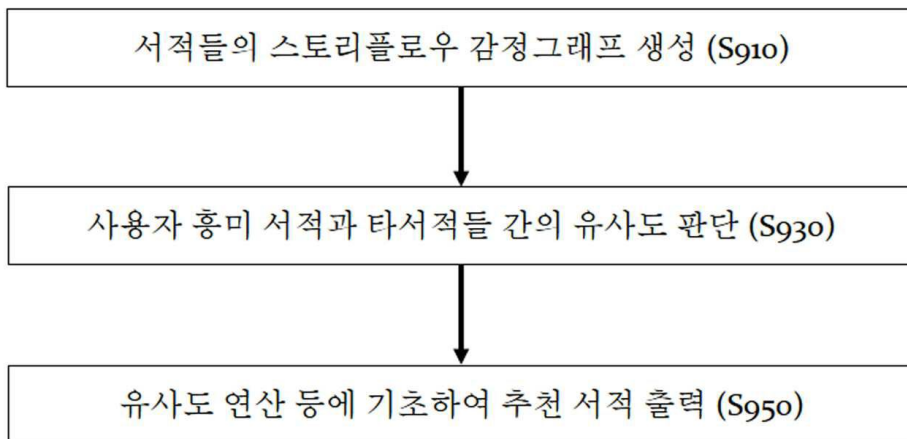
도면12



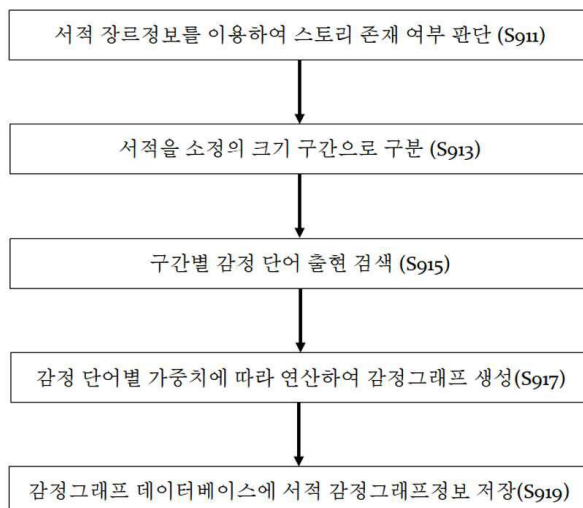
도면13

구 간	서적 1	서적 2	서적 3	서적 4	서적 ₁ 과 서적 ₂ 간 제1 유사판단 기초값	서적 ₁ 과 서적 ₃ 간 제1 유사판단 기초값	서적 ₁ 과 서적 ₄ 간 제1 유사판단 기초값
1	7	8	6	5	1	1	2
2	6	9	5	4	3	1	2
3	5	10	4	3	5	1	2
4	4	11	3	2	7	1	2
5	3	12	2	1	9	1	2
6	2	13	1	0	11	1	2
7	3	11	2	1	8	1	2
8	4	10	3	2	6	1	2
9	5	9	4	3	4	1	2
10	6	8	5	4	2	1	2
11	8	9	7	6	1	1	2
12	7	8	6	5	1	1	2
13	6	9	5	4	3	1	2
14	5	10	4	3	5	1	2

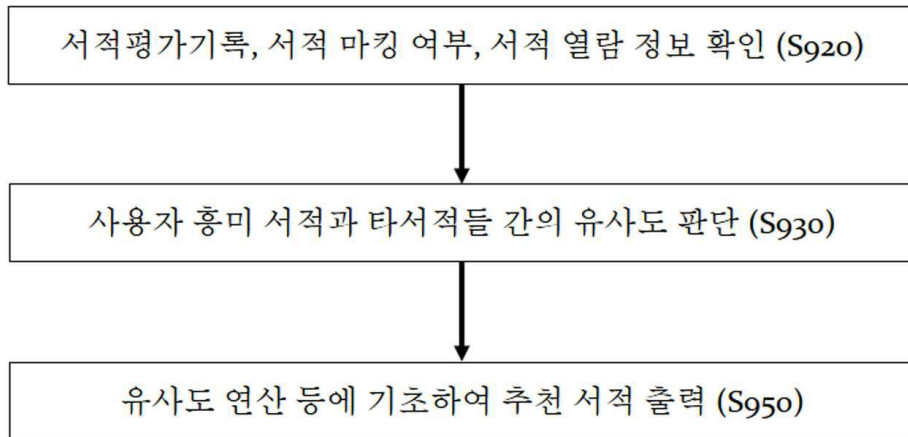
도면14



도면15



도면16



도면17



도면18

