



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2017년02월10일
(11) 등록번호 10-1705804
(24) 등록일자 2017년02월06일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 17/27 (2006.01)
(52) CPC특허분류
G06F 17/30731 (2013.01)
G06F 17/2745 (2013.01)
(21) 출원번호 10-2015-0161629
(22) 출원일자 2015년11월18일
심사청구일자 2015년11월18일
(56) 선행기술조사문헌
KR1020140073717 A*
JP2015032166 A*
JP2011150450 A
US20120095952 A1
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(주)삼부시스템
서울특별시 강남구 선릉로82길 13, 삼부빌딩 (대치동)
(72) 발명자
이상근
서울특별시 강남구 선릉로 221, 205동 1104호 (도곡동, 도곡렉슬아파트)
류우중
서울특별시 성북구 한천로 509, 106동 502호 (석관동, 코오롱아파트)
(74) 대리인
특허법인엠에이피에스

전체 청구항 수 : 총 9 항

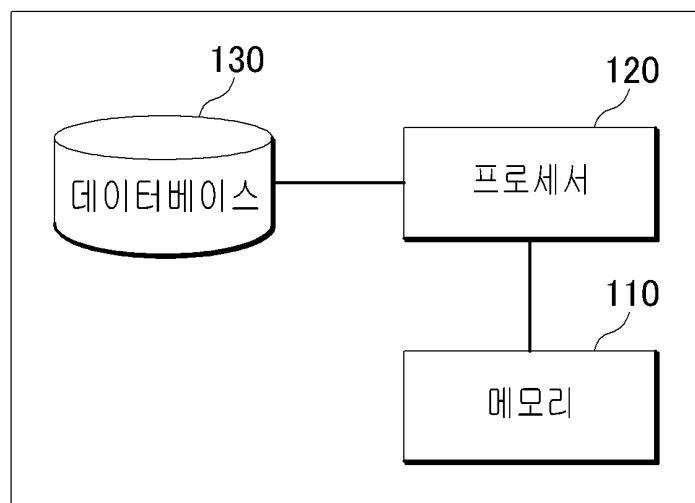
심사관 : 박승철

(54) 발명의 명칭 시맨틱 콘텐츠 분류 장치 및 이를 이용한 콘텐츠 분류 방법

(57) 요약

본 발명은 시맨틱 콘텐츠 분류 프로그램이 저장된 메모리 및 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하되, 프로세서는 프로그램의 실행에 따라, 콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하며, 산출된 문장 각각의 지역 토픽-감성에 대한 확 (뒷면에 계속)

대표도 - 도1



100

를 분포 및 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여 콘텐츠에 포함된 복수의 토픽에 대한 확률 및 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산하고, 계산된 복수의 토픽에 대한 확률 및 계산된 복수의 감성에 대한 확률 중 적어도 하나 이상에 기초하여, 콘텐츠를 분류한다. 이때, 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출되며, 토픽-감성의 어휘 분포는 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출되는 것이며, 토픽의 감성 분포는 토픽-감성의 어휘 분포에 대한 사전확률과 상이한, 토픽의 감성 분포에 대한 사전확률에 기초하여 산출된다.

(52) CPC특허분류

G06F 17/2785 (2013.01)

G06F 17/30705 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1425095209

부처명 중소기업청

연구관리전문기관 중소기업기술정보진흥원

연구사업명 2015년도 산학연협력 기술개발사업(자율편성형) 첫걸음R&D-참여기관(컨소시엄과제)

연구과제명 사용자 관심사를 반영한 맞춤형 뉴스 추천 서비스

기 여 율 1/1

주관기관 한양대학교산학협력단

연구기간 2015.06.01 ~ 2016.05.31

명세서

청구범위

청구항 1

시맨틱 콘텐츠 분류 장치에 있어서,

시맨틱 콘텐츠 분류 프로그램이 저장된 메모리 및

상기 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하되,

상기 프로세서는 상기 프로그램의 실행에 따라, 콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하며,

상기 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 상기 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여 상기 콘텐츠에 포함된 복수의 토픽에 대한 확률 및 상기 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산하고,

상기 계산된 콘텐츠에 포함된 복수의 토픽에 대한 확률에 기초하여 토픽 별로 상기 콘텐츠를 분류한 이후, 상기 계산된 콘텐츠에 포함된 복수의 감성에 대한 확률에 기초하여, 상기 콘텐츠를 긍정 콘텐츠 및 부정 콘텐츠 중 어느 하나로 분류하되,

상기 전역 토픽-감성에 대한 확률 분포 및 상기 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출되며,

상기 토픽-감성의 어휘 분포는 상기 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출되는 것이며,

상기 토픽의 감성 분포는 상기 토픽-감성의 어휘 분포에 대한 사전확률과 상이한, 상기 토픽의 감성 분포에 대한 사전확률에 기초하여 산출되는 것인, 시맨틱 콘텐츠 분류 장치.

청구항 2

삭제

청구항 3

제 1 항에 있어서,

상기 프로세서는 상기 콘텐츠로부터 추출된 단어 및 상기 단어에 대응하여 상기 콘텐츠로부터 추출된 감성에 기초하여, 상기 토픽-감성의 어휘 분포에 대한 사전확률 및 상기 토픽의 감성 분포에 대한 사전확률을 산출하는 것인, 시맨틱 콘텐츠 분류 장치.

청구항 4

제 3 항에 있어서,

상기 프로세서는 상기 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 상기 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 상기 토픽-감성의 어휘 분포에 대한 사전확률을 결정하는, 시맨틱 콘텐츠 분류 장치.

청구항 5

제 3 항에 있어서,

상기 프로세서는 상기 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 상기 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 상기 토픽의 감성 분포에 대한 사전확률을 결정하는, 시맨틱 콘텐츠 분류 장치.

청구항 6

시맨틱 콘텐츠 분류 장치에서의 시맨틱 콘텐츠 분류 방법에 있어서,

콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포를 산출하는 단계;

상기 문장 각각에 대하여, 전역 토픽-감성에 대한 확률 분포를 산출하는 단계;

상기 지역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 상기 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 상기 전역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 상기 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 상기 콘텐츠에 포함된 복수의 토픽에 대한 확률을 계산하는 단계;

상기 지역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 상기 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 상기 전역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 상기 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 상기 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산하는 단계;

상기 토픽에 대한 확률을 산출하는 단계로부터 계산된 상기 콘텐츠에 포함된 복수의 토픽에 대한 확률에 기초하여, 토픽 별로 상기 콘텐츠를 분류하는 단계; 및

상기 토픽 별로 상기 콘텐츠를 분류하는 단계 이후에, 상기 감성에 대한 확률을 계산하는 단계로부터 상기 계산된 콘텐츠에 포함된 복수의 감성에 대한 확률에 기초하여, 상기 콘텐츠를 긍정 콘텐츠 및 부정 콘텐츠 중 어느 하나로 분류하는 단계를 포함하되,

상기 전역 토픽-감성에 대한 확률 분포 및 상기 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출되며,

상기 토픽-감성의 어휘 분포는 상기 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출되는 것이며,

상기 토픽의 감성 분포는 상기 토픽-감성의 어휘 분포에 대한 사전확률과 상이한, 상기 토픽의 감성 분포에 대한 사전확률에 기초하여 산출되는 것인, 시맨틱 콘텐츠 분류 방법.

청구항 7

삭제

청구항 8

제 6 항에 있어서,

상기 지역 토픽-감성에 대한 확률 분포를 산출하는 단계 및 상기 전역 토픽-감성에 대한 확률 분포를 산출하는 단계 이전에, 상기 콘텐츠에 대하여, 상기 토픽-감성의 어휘 분포에 대한 사전확률을 결정하는 단계; 및

상기 콘텐츠에 대하여, 상기 토픽의 감성 분포에 대한 사전확률을 결정하는 단계를 더 포함하는, 시맨틱 콘텐츠 분류 방법.

청구항 9

제 8 항에 있어서,

상기 토픽-감성의 어휘 분포에 대한 사전확률을 결정하는 단계는,

상기 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 상기 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 상기 토픽-감성의 어휘 분포에 대한 사전확률을 결정하는, 시맨틱 콘텐츠 분류 방법.

청구항 10

제 8 항에 있어서,

상기 토픽의 감성 분포에 대한 사전확률을 결정하는 단계는,

상기 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 상기 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 상기 토픽의 감성 분포에 대한 사전확률을 결정하는, 시맨틱 콘텐츠 분류 방법.

청구항 11

제 6 항 및 제 8 항 내지 제 10 항 중 어느 한 항에 기재된 방법을 컴퓨터 상에서 수행하기 위한 프로그램을 기록한 컴퓨터 판독 가능한 기록 매체.

발명의 설명

기술 분야

[0001] 본 발명은 시맨틱 콘텐츠 분류 장치 및 이를 이용한 분류 방법에 관한 것이다.

배경 기술

[0002] 인터넷의 발전과 더불어 뉴스, 웹 페이지 및 도서와 같은 텍스트 형태의 온라인 콘텐츠(on-line content)의 양이 급속도로 증가하고 있다. 이에 따라, 콘텐츠의 사용자는 수많은 온라인 콘텐츠 속에서 자신이 원하는 콘텐츠를 찾기 위하여 많은 시간과 노력을 소비해야 한다. 그러나 콘텐츠의 사용자는 최소한의 시간과 노력으로 원하는 콘텐츠를 제공 받기를 원하고 있다. 따라서 서비스 제공 업체에서는 이러한 콘텐츠 사용자의 요구를 충족시키기 위하여 콘텐츠 사용자의 의도를 파악하고, 콘텐츠 사용자의 의도에 맞는 콘텐츠를 사용자에게 제공하는 개인화(personalization) 기술을 이용하고 있다.

[0003] 콘텐츠 사용자의 의도에 맞는 콘텐츠를 제공하기 위한 개인화 기술 중 하나로는 콘텐츠를 주제에 따라 분류하고, 사용자의 의도를 추론해 추론한 사용자 의도와 연관된 주제에 해당하는 콘텐츠를 제공하는 방법이 있다. 이러한 방법은 콘텐츠를 주제에 따라 분류하는 과정에서 콘텐츠의 잠재된 감성(sentiment)을 고려하지 않기 때문에, 콘텐츠 사용자가 원하는 콘텐츠를 제대로 제공하지 못할 수 있다.

[0004] 이와 관련되어, 한국 공개특허공보 제10-2012-0077873호(발명의 명칭: "사용자 정보 및 상황 정보를 이용한 콘텐츠 추천 시스템 및 방법")는 사용자 정보 및 상황 정보를 이용하여 사용자에게 매칭되는 콘텐츠를 추천하는 시스템 및 방법을 개시하고 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 콘텐츠로부터 시맨틱 토픽 및 시맨틱 감성을 추출하여 콘텐츠를 분류하는 시맨틱 콘텐츠 분류 장치 및 이를 이용한 콘텐츠 분류 방법을 제공한다.

[0006] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0007] 상술한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본 발명의 제 1 측면에 따른 시맨틱 콘텐츠 분류 장치는 시맨틱 콘텐츠 분류 프로그램이 저장된 메모리 및 메모리에 저장된 프로그램을 실행하는 프로세서를 포함한다. 이때, 프로세서는 프로그램의 실행에 따라, 콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하며, 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여 콘텐츠에 포함된 복수의 토픽에 대한 확률 및 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산하고, 계산된 복수의 토픽에 대한 확률 및 계산된 복수의 감성에 대한 확률 중 적어도 하나 이상에 기초하여, 콘텐츠를 분류한다. 그리고 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출되며, 토픽-감성의 어휘 분포는 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출되는 것이며, 토픽의 감성 분포는 토픽-감성의 어휘 분포에 대한 사전확률과 상이한, 토픽의 감성 분포에 대한 사전확률에 기초하여 산출된다.

[0008] 또한, 본 발명의 제 2 측면에 따른 시맨틱 콘텐츠 분류 장치에서의 시맨틱 콘텐츠 분류 방법은 콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포를 산출하는 단계; 문장 각각에 대하여, 전역 토픽-감성에 대한 확률 분포를 산출하는 단계; 지역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산

출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 콘텐츠에 포함된 복수의 토픽에 대한 확률을 계산하는 단계; 지역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하는 단계로부터 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산하는 단계; 및 토픽에 대한 확률을 계산하는 단계로부터 계산된 콘텐츠에 포함된 복수의 토픽에 대한 확률 및 감성에 대한 확률을 계산하는 단계로부터 계산된 콘텐츠에 포함된 복수의 감성에 대한 확률 중 적어도 하나 이상에 기초하여, 콘텐츠를 분류하는 단계를 포함한다. 이때, 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출되며, 토픽-감성의 어휘 분포는 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출되는 것이며, 토픽의 감성 분포는 토픽-감성의 어휘 분포에 대한 사전확률과 상이한, 토픽의 감성 분포에 대한 사전확률에 기초하여 산출된다.

발명의 효과

[0009] 전술한 본 발명의 과제 해결 수단 중 어느 하나에 의하면, 본 발명은 기 설정된 비대칭적인 사전 확률에 따라, 시맨틱 토픽 및 시맨틱 감성을 추출하므로, 기존의 대칭적인 사전 확률에 기초한 콘텐츠 분류 방법보다 정확하게 콘텐츠에 포함된 시맨틱 토픽 및 시맨틱 감성을 추출할 수 있다. 또한, 본 발명은 추출된 시맨틱 토픽 및 시맨틱 감성에 따라 콘텐츠를 자동으로 분류할 수 있다.

도면의 간단한 설명

[0010] 도 1은 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치의 블록도이다.
 도 2는 본 개시에 사용된 표기를 도시한 것이다.
 도 3은 본 발명의 일 실시예에 따른 분류된 콘텐츠의 예시도이다.
 도 4는 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 프로그램의 블록도이다.
 도 5는 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치의 시맨틱 분류 방법의 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0011] 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.

[0012] 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 의미하며, 하나 또는 그 이상의 다른 특징이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0013] 이하의 실시예는 본 발명의 이해를 돕기 위한 상세한 설명이며, 본 발명의 권리 범위를 제한하는 것이 아니다. 따라서 본 발명과 동일한 기능을 수행하는 동일 범위의 발명 역시 본 발명의 권리 범위에 속할 것이다.

[0014] 다음은 도 1을 참조하여 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100)를 설명한다.

[0015] 도 1은 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100)의 블록도이다.

[0016] 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠에 포함된 하나 이상의 문장에 대하여 시맨틱 토픽 및 시맨틱 감성을 추출한다. 그리고 시맨틱 콘텐츠 분류 장치(100)는 추출된 시맨틱 토픽 및 시맨틱 감성에 기초하여, 콘텐츠를 분류한다. 이때, 시맨틱 콘텐츠 분류 장치(100)는 메모리(110) 및 프로세서(120)를 포함한다. 또한, 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠 및 분류된 콘텐츠를 저장하는 데이터베이스(130)를 포함할 수 있다.

[0017] 메모리(110)는 시맨틱 콘텐츠 분류 프로그램(400)이 저장된다. 이때, 메모리(110)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 비휘발성 저장장치 및 저장된 정보를 유지하기 위하여 전력이 필요한 휘발성 저장장치를 통칭하는 것이다.

- [0018] 데이터베이스(130)는 콘텐츠를 저장할 수 있다. 이때, 콘텐츠는 뉴스, 도서 및 웹 문서와 같은 텍스트 형태의 온라인 콘텐츠일 수 있다. 데이터베이스(130)에 저장되는 콘텐츠는 시맨틱 콘텐츠 분류 장치(100)의 관리자가 수집한 후 데이터베이스(130)에 저장된 것이거나, 크롤러(crawler)와 같은 자동 콘텐츠 수집 모듈을 이용하여 자동으로 수집된 후, 데이터베이스(130)에 저장된 것일 수 있다.
- [0019] 또한, 데이터베이스(130)는 감성어 집합을 저장할 수 있다. 그리고 데이터베이스(130)는 콘텐츠로부터 단어 및 감성을 추출하기 위하여 사용될 수 있는 단어 및 감성 사전 또는 문법 사전을 저장할 수 있다.
- [0020] 이때, 감성어 집합은 단어의 극성에 따라, 감성어를 구분하여 저장할 수 있다. 즉, 감성어 집합은 긍정의 의미를 가지는 단어 및 부정의 감성을 가지는 단어를 포함할 수 있다. 예를 들어, 감성어 사전에 포함되는 긍정의 의미를 가지는 단어는 "좋다", "훌륭하다", "예쁘다" 및 "최고" 등이 될 수 있다. 또한, 감성어 사전에 포함되는 부정의 의미를 가지는 단어는 "나쁘다", "싫다", "짜증" 및 "최악" 등이 될 수 있다.
- [0021] 프로세서(120)는 콘텐츠 집합으로부터 시맨틱 토픽 및 시맨틱 감성을 추출한다. 이때, 콘텐츠 집합은 데이터베이스(130)에 기 저장된 콘텐츠들의 집합이거나, 온라인을 통하여 수집된 콘텐츠 집합일 수 있다. 또한, 콘텐츠 집합은 데이터베이스(130)에 기 저장된 콘텐츠 집합과 온라인을 통해 수집된 콘텐츠 집합이 병합된 형태일 수 있다.
- [0022] 프로세서(120)는 추출된 시맨틱 토픽 및 시맨틱 감성을 데이터베이스(130)에 저장한다. 그리고 프로세서(120)는 추출된 시맨틱 토픽 및 시맨틱 감성에 기초하여, 콘텐츠를 분류한다.
- [0023] 이때, 시맨틱 토픽 및 시맨틱 감성은 토픽-감성에 대한 어휘 분포, 각 콘텐츠에 대한 토픽 분포, 토픽에 대한 감성 분포, 토픽의 문맥을 결정하기 위한 토픽 문맥 분포 및 각 문장에 대해 슬라이딩 윈도우 크기를 설정하기 위한 카테고리 분포를 포함할 수 있다.
- [0024] 구체적으로 프로세서(120)는 콘텐츠에 포함된 복수의 문장으로부터 단어를 추출할 수 있다. 프로세서(120)는 문장으로부터 추출된 단어에 대응하는 감성을 추출할 수 있다.
- [0025] 이때, 프로세서(120)는 문장에서 단어 및 감성을 추출하기 위하여 데이터베이스(130)에 기 저장된 단어 및 감성 사전을 사용할 수 있다. 즉, 프로세서(120)는 콘텐츠에 포함된 문장에 단어 및 감성 사전에 포함된 단어 및 감성이 포함되는 경우, 해당 문장을 대표하는 단어 또는 감성으로 추출할 수 있다. 또는 프로세서(120)는 형태소 분석기를 통하여 해당 문장의 구문 분석(parsing) 및 형태소 분석(morphological analysis)을 수행하여 단어 또는 감성을 추출할 수 있다. 그러나 단어 및 감성을 추출하는 방법은 이에 한정되지 않으며, 다양한 방법을 적용할 수 있다.
- [0026] 프로세서(120)는 단어 및 감성을 추출하면, 추출된 단어 및 감성에 기초하여 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출할 수 있다.
- [0027] 이때, 프로세서(120)는 토픽-감성에 대한 어휘 분포 및 토픽에 대한 감성 분포에 비대칭적 사전 확률(asymmetric prior)을 부여할 수 있다. 예를 들어, 토픽-감성에 대한 어휘 분포의 사전 확률(β^{prior})은 [수학식 1]과 같으며, 토픽에 대한 감성 분포의 사전 확률(δ^{prior})은 [수학식 2]와 같다.

수학식 1

$$\beta^{prior} = \begin{cases} C_{bp} & \text{if } w \in Q_{pos}, s = pos \\ C_{bn} & \text{if } w \in Q_{neg}, s = neg \\ C_{bo} & \text{otherwise} \end{cases}$$

[0029]

- [0031] [수학식 1]에서 w 는 단어이며, C_{bp} 는 단어가 긍정의 감성어 집합에 포함되고, 단어의 감성도 긍정인 경우에 대한 사전 확률이다. C_{bn} 는 단어가 부정의 감성어 집합에 포함되고, 단어의 감성도 부정인 경우에 대한 사전 확률이며, C_{bo} 는 그 외의 경우에 대한 사전 확률이다.

[0032] 토픽-감성에 대한 어휘 분포에서 각각의 사전 확률은 비대칭적으로 부여될 수 있다. 이때, 각각의 사전 확률은 시맨틱 콘텐츠 분류 장치(100)의 관리자가 미리 정한 것일 수 있다. 또한, 각각의 사전 확률은 테스트 콘텐츠를 이용하여 실험적으로 산출된 것일 수 있으나 이에 한정된 것은 아니다.

수학식 2

$$\delta^{prior} = \begin{cases} C_{dp} & \text{if } w \in Q_{pos}, s=pos \\ C_{dn} & \text{if } w \in Q_{neg}, s=neg \\ C_{do} & \text{otherwise} \end{cases}$$

[0034]

[0036] [수학식 2]에서 w 는 단어이며, C_{dp} 는 단어가 긍정의 감성어 집합에 포함되고, 단어의 감성도 긍정인 경우에 대한 사전 확률이다. C_{dn} 는 단어가 부정의 감성어 집합에 포함되고, 단어의 감성도 부정인 경우에 대한 사전 확률이며, C_{do} 는 그 외의 경우에 대한 사전 확률이다. 토픽에 대한 감성 분포에서 각각의 사전 확률은 시맨틱 콘텐츠 분류 장치(100)의 관리자가 미리 정한 것일 수 있다.

[0037] 토픽에 대한 감성 분포에서 각각의 사전 확률은 비대칭적으로 부여될 수 있다. 이때, 각각의 사전 확률은 토픽-감성에 대한 어휘 분포와 유사하게 시맨틱 콘텐츠 분류 장치(100)의 관리자가 미리 정한 것일 수 있다. 또한, 각각의 사전 확률은 테스트 콘텐츠를 이용하여 실험적으로 산출된 것일 수 있으나 이에 한정된 것은 아니다.

[0038] 이와 같이, 프로세서(120)는 각각 문장에서 추출된 단어가 긍정 또는 부정 감성어 집합에 포함되는지 여부와 단어의 감성이 긍정 또는 부정에 포함되는지 여부에 따라, 토픽-감성에 대한 어휘 분포에 대한 사전 확률 및 토픽에 대한 감성 분포에 대한 사전 확률을 부여할 수 있다.

[0039] 그리고 프로세서(120)는 토픽-감성에 대한 어휘 분포 및 토픽에 대한 감성 분포에 대한 사전 확률에 기초하여, 콘텐츠에 포함된 각 문장에 대하여 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포를 산출할 수 있다.

[0040] 프로세서(120)는 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포가 산출되면, 콘텐츠에 포함된 복수의 토픽 및 복수의 감성에 대한 확률을 각각 계산할 수 있다. 그리고 프로세서(120)는 계산된 복수의 토픽에 대한 확률 및 복수의 감성에 대한 확률 중 적어도 하나에 기초하여, 콘텐츠를 분류할 수 있다.

[0041] 예를 들어, 프로세서(120)는 콘텐츠에 포함된 각 토픽에 대한 확률에 따라, 토픽별로 콘텐츠를 분류할 수 있다. 또한, 프로세서(120)는 콘텐츠에 포함된 각 감성에 대한 확률에 따라, 감성별로 콘텐츠를 분류하거나, 토픽 및 감성을 모두 고려하여 콘텐츠를 분류할 수 있다.

[0042] 콘텐츠의 분류 과정을 토픽을 기준으로 상세하게 살펴보면, 프로세서(120)는 각 콘텐츠가 각 토픽에 포함될 확률에 따라 각 콘텐츠를 분류할 수 있다. 그리고 프로세서(120)는 토픽으로 분류된 콘텐츠를 다시 각 감성에 포함될 확률에 기초하여 2단계로 분류할 수 있다.

[0043] 이를 위하여, 먼저 프로세서(120)는 [수학식 3] 및 [수학식 4]에 기초하여, 각 문장의 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포를 산출할 수 있다.

수학식 3

$$\theta_{d,s}^{gl}(z,l) = \sum_{v \in s} \sum_{w \in v} \frac{n_{w_{d,i}}^{gl,z,l} + \beta^{gl}}{n^{gl,z,l} + V\beta^{gl}} \times \frac{n_v^{d,s} + \gamma}{n^{d,s} + T\gamma} \\ \times \frac{n_{gl}^{d,s} + \alpha_{gl}^{mix}}{n^{d,v} + \sum_{\gamma' \in \{gl, loc\}} \alpha_{\gamma'}^{mix}} \times \frac{n_l^{d,gl} + \delta^{gl}}{n^{d,gl} + Y\delta^{gl}} \times \frac{n_z^{d,l,gl} + \alpha^{gl}}{n_l^{d,gl} + K^{gl}\alpha^{gl}}$$

[0044]

수학식 4

$$P(z,l|d) = \sum_s \theta_{d,s}^{gl}(z,l)$$

[0046]

[0048] 도 2는 본 개시에 사용된 표기를 도시한 것이다. 즉, 다음의 [수학식 3] 내지 [수학식 7]에서 각 상수 및 변수의 상세한 설명은 도 2를 참조한다.

[0049] 프로세서(120)는 [수학식 3]을 참조하여, 각 문장의 전역 토픽-감성에 대한 확률 분포를 산출할 수 있다. 또한, 프로세서(120)는 [수학식 3]을 통하여 산출된 전역 토픽-감성에 대한 확률 분포에 기초하여 [수학식 4]를 통하여, 각 콘텐츠가 각 전역 토픽 및 각 감성에 해당할 확률을 산출할 수 있다.

[0050] 또한, 프로세서(120)는 [수학식 5] 및 [수학식 6]에 기초하여, 각 콘텐츠에서 각 문장의 지역 토픽-감성에 대한 확률 분포를 산출할 수 있다.

수학식 5

$$\theta_{d,s}^{loc}(z,l) = \sum_{v \in s} \sum_{w \in v} \frac{n_{w_{d,i}}^{loc,z,l} + \beta^{loc}}{n^{loc,z,l} + V\beta^{loc}} \times \frac{n_v^{d,s} + \gamma}{n^{d,s} + T\gamma} \\ \times \frac{n_{loc}^{d,s} + \alpha_{loc}^{mix}}{n^{d,v} + \sum_{\gamma' \in \{gl, loc\}} \alpha_{\gamma'}^{mix}} \times \frac{n_l^{d,v,loc} + \delta^{loc}}{n_{loc}^{d,v} + Y\delta^{loc}} \times \frac{n_z^{d,v,l,loc} + \alpha^{loc}}{n_l^{d,v,loc} + K^{loc}\alpha^{loc}}$$

[0051]

수학식 6

$$P(z,l|d) = \sum_s \theta_{d,s}^{loc}(z,l)$$

[0053]

[0055] [수학식 5]는 각 콘텐츠에서 각 문장의 지역 토픽-감성에 대한 확률 분포를 산출하는 수학식이다. 또한, [수학식 6]은 [수학식 5]를 통하여 산출된 지역 토픽-감성에 대한 확률 분포를 이용해 각 콘텐츠가 각 지역 토픽과

각 감성에 해당할 확률을 산출하는 수식이다.

[0056] 즉, 프로세서(120)는 [수학식 5]를 사용하여 주어진 콘텐츠의 각 문장이 지역 토픽의 각 토픽에 해당할 확률을 산출할 수 있다. 그리고 프로세서(120)는 [수학식 5]를 통하여 산출된 각 문장이 지역 토픽의 각 토픽에 해당할 확률을 [수학식 6]에 적용해 주어진 콘텐츠가 지역 토픽의 각 토픽에 해당할 확률을 산출할 수 있다.

[0057] 프로세서(120)는 산출된 각 콘텐츠가 각 토픽에 해당할 확률에 기초하여, 각 콘텐츠의 시맨틱 토픽을 산출할 수 있다. 그리고 프로세서(120)는 산출된 시맨틱 토픽에 기반하여 콘텐츠를 분류할 수 있다. 이때, 임의의 콘텐츠로부터 복수의 시맨틱 토픽이 추출되는 경우, 프로세서(120)는 해당 콘텐츠를 복수의 토픽으로 분류할 수 있다.

[0058] 한편, 프로세서(120)는 시맨틱 감성을 이용하여 각 토픽으로 분류된 콘텐츠를 다시 토픽 내에서 감성에 따라 분류할 수 있다. 예를 들어, 분류하고자 하는 감성이 긍정과 부정인 경우, 프로세서(120)는 [수학식 7]에 기초하여, 주어진 콘텐츠에 대해 긍정 및 부정 여부에 대한 확률을 산출할 수 있다.

수학식 7

$$P(l|d) = \frac{n_l^{d,gl} + \delta^{gl}}{n^{d,gl} + Y\delta^{gl}} + \sum_{w \in v} \frac{n_l^{d,v,loc} + \delta^{loc}}{n_{loc}^{d,v} + Y\delta^{loc}}$$

[0060]

[0062] 즉, 프로세서(120)는 [수학식 7]을 사용하여, 각 콘텐츠가 분류된 토픽 내에서 긍정의 의미가 있을 확률 및 부정의 의미가 있을 확률을 산출할 수 있다. 그리고 프로세서(120)는 긍정의 의미가 있을 확률이 부정의 의미가 있을 확률에 비하여 클 경우, 해당 콘텐츠를 긍정으로 분류할 수 있다. 이와 반대로, 프로세서(120)는 부정의 의미가 있을 확률이 긍정의 의미가 있을 확률에 비하여 작은 경우, 해당 콘텐츠를 부정으로 분류할 수 있다.

[0063] 도 3은 본 발명의 일 실시예에 따른 분류된 콘텐츠의 예시도이다.

[0064] 도 3을 참조하면 프로세서(120)는 먼저, 토픽을 기준으로 콘텐츠를 분류할 수 있다. 그리고 프로세서(120)는 각 토픽에 포함된 콘텐츠를 다시 긍정 및 부정으로 감성에 따라 분류할 수 있다.

[0065] 예를 들어, '토픽 1'에 포함된 '콘텐츠 1'에 대하여, 수학식 7을 이용하여 도출된 긍정에 대한 확률이 부정에 대한 확률보다 크므로, 프로세서(120)는 도 3과 같이, '콘텐츠 1'을 긍정에 해당하는 콘텐츠로 분류할 수 있다. 또한, '토픽 1'에 포함된 '콘텐츠 10'의 경우, 부정에 대한 확률이 크므로 프로세서(120)는 도 3과 같이 '콘텐츠 10'을 부정에 해당하는 콘텐츠로 분류할 수 있다. 이렇게 프로세서(120)는 콘텐츠를 분류하고, 분류된 콘텐츠를 데이터베이스(130)에 저장할 수 있다.

[0066] 이와 같이, 프로세서(120)는 시맨틱 토픽 및 시맨틱 감성을 고려하므로, 의미 및 감성적으로 연관된 콘텐츠를 같은 그룹으로 분류할 수 있다.

[0067] 프로세서(120)는 하나의 콘텐츠를 여러 그룹에 포함되도록 분류할 수 있다. 이때, 프로세서(120)는 동일한 콘텐츠의 토픽 및 감성을 각각 상이하게 분류할 수 있다. 즉, 프로세서(120)는 동일한 콘텐츠를 상이한 토픽 및 상이한 감성의 그룹에 포함되도록 분류할 수 있다. 예를 들어, '콘텐츠 10'은 '토픽 1' 및 '부정'으로 분류될 수 있으며, '토픽 3' 및 '긍정' 그룹으로도 분류될 수 있다.

[0068] 위에서는 콘텐츠로부터 시맨틱 토픽 및 시맨틱 감성을 추출하고, 추출된 시맨틱 토픽으로 콘텐츠를 분류한 다음 다시 추출된 시맨틱 감성에 기초하여 콘텐츠를 분류하는 과정을 설명하였다. 그러나 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100)는 이외에도 추출된 시맨틱 감성에 기초하여 콘텐츠를 분류한 다음, 다시 추출된 시맨틱 감성에 기초하여 콘텐츠를 분류하거나, 추출된 시맨틱 토픽 및 시맨틱 감성을 동시에 고려하여 콘텐츠를 분류할 수 있다.

[0069] 다음은 도 4를 참조하여, 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 프로그램(400)을 설명한다.

[0070] 도 4는 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 프로그램(400)의 블록도이다.

[0071] 시맨틱 콘텐츠 분류 프로그램(400)은 시맨틱 토픽 추출 모듈(410), 시맨틱 토픽 분류 모듈(420), 시맨틱 감성

분류 모듈(430), 콘텐츠 저장 모듈(440), 토픽 모델 저장 모듈(450) 및 시맨틱 콘텐츠 저장 모듈(460)을 포함할 수 있다.

- [0072] 콘텐츠 저장 모듈(440)은 데이터베이스(130)에 콘텐츠를 저장할 수 있다. 이때, 데이터베이스(130)에 저장되는 콘텐츠는 앞에서 설명한 바와 같이, 관리자가 직접 수집한 뒤 콘텐츠 저장 모듈(440)을 통하여 데이터베이스(130)에 저장된 것일 수 있다. 또한, 데이터베이스(130)에 저장된 콘텐츠는 콘텐츠 수집 모듈(440)을 통하여 자동으로 수집된 것일 수 있다.
- [0073] 먼저, 시맨틱 토픽 추출 모듈(410)은 콘텐츠 저장 모듈(440)을 통하여 데이터베이스(130)에 저장된 콘텐츠 집합으로부터 시맨틱 토픽 및 시맨틱 감성을 추출할 수 있다. 그리고 시맨틱 토픽 추출 모듈(410)은 시맨틱 토픽 및 시맨틱 감성 추출 시 생성한 토픽 모델을 토픽 모델 저장 모듈(450)로 전달하여 데이터베이스(130)에 저장할 수 있다.
- [0074] 시맨틱 토픽 분류 모듈(420) 및 시맨틱 감성 분류 모듈(430)은 시맨틱 토픽 추출 모듈(410)로부터 추출된 시맨틱 토픽 및 시맨틱 감성에 따라 콘텐츠를 분류할 수 있다. 먼저, 시맨틱 토픽 분류 모듈(420)을 통하여, 추출된 시맨틱 토픽에 따라 콘텐츠를 분류한 후, 시맨틱 감성 분류 모듈(430)을 통하여, 분류된 콘텐츠를 시맨틱 감성에 기초하여 다시 분류할 수 있다. 그리고 시맨틱 감성 분류 모듈(430)은 시맨틱 콘텐츠 저장 모듈(460)을 통하여, 데이터베이스(130)에 분류된 콘텐츠를 저장할 수 있다.
- [0075] 다음은 도 5를 참조하여, 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100)에서의 시맨틱 분류 방법을 설명한다.
- [0076] 도 5는 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100)의 시맨틱 분류 방법의 순서도이다.
- [0077] 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠에 포함된 복수의 문장 각각에 대하여, 지역 토픽-감성에 대한 확률 분포를 산출한다(S500). 또한, 시맨틱 콘텐츠 분류 장치(100)는 문장 각각에 대하여, 전역 토픽-감성에 대한 확률 분포를 산출한다(S510). 이때, 전역 토픽-감성에 대한 확률 분포 및 지역 토픽-감성에 대한 확률 분포는 토픽-감성의 어휘 분포 및 토픽의 감성 분포에 기초하여 산출된다. 토픽-감성의 어휘 분포는 토픽-감성의 어휘 분포에 대한 사전확률에 기초하여 산출된다. 그리고 토픽의 감성 분포는 토픽의 감성 분포에 대한 사전확률에 기초하여 산출된다.
- [0078] 이를 위하여, 시맨틱 콘텐츠 분류 장치(100)는 지역 토픽-감성에 대한 확률 분포 및 전역 토픽-감성에 대한 확률 분포를 산출하기 이전에, 콘텐츠에 대하여, 토픽-감성 쌍의 어휘 분포에 대한 사전확률 및 토픽의 감성 분포에 대한 사전확률을 결정할 수 있다.
- [0079] 구체적으로 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 토픽-감성의 어휘 분포에 대한 사전확률을 결정할 수 있다. 이때, 토픽-감성의 어휘 분포에 대한 사전확률은 각각의 감성에 따라 비대칭적으로 설정될 수 있다.
- [0080] 그리고 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠로부터 추출된 단어가 긍정 감성어 사전 및 부정 감성어 사전 중 어느 것에 포함되는지 여부와 콘텐츠로부터 추출된 단어에 대응하는 감성의 극성에 따라, 토픽의 감성 분포에 대한 사전확률을 결정할 수 있다. 이때, 토픽의 감성 분포에 대한 사전확률은 각각의 감성에 따라 비대칭적으로 설정될 수 있다. 또한, 토픽의 감성 분포에 대한 사전확률은 토픽-감성의 어휘 분포에 대한 사전확률과 상이하게 설정될 수 있다.
- [0081] 한편, 시맨틱 콘텐츠 분류 장치(100)는 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 콘텐츠에 포함된 복수의 토픽에 대한 확률을 계산한다(S520).
- [0082] 시맨틱 콘텐츠 분류 장치(100)는 산출된 문장 각각의 지역 토픽-감성에 대한 확률 분포 및 산출된 문장 각각의 전역 토픽-감성에 대한 확률 분포에 기초하여, 콘텐츠에 포함된 복수의 감성에 대한 확률을 계산한다(S530).
- [0083] 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠에 포함된 복수의 토픽에 대한 확률 및 콘텐츠에 포함된 복수의 감성에 대한 확률 중 적어도 하나 이상에 기초하여, 콘텐츠를 분류한다(S540).
- [0084] 예를 들어, 시맨틱 콘텐츠 분류 장치(100)는 2단계로 콘텐츠를 분류할 수 있다. 시맨틱 콘텐츠 분류 장치(100)는 콘텐츠에 포함된 복수의 토픽에 대한 확률에 기초하여, 토픽별로 콘텐츠를 분류할 수 있다. 그리고 시맨

틱 콘텐츠 분류 장치(100)는 토픽별로 콘텐츠를 분류한 이후에, 콘텐츠에 포함된 복수의 감성에 대한 확률에 기초하여, 콘텐츠를 긍정 콘텐츠 및 부정 콘텐츠 중 어느 하나로 분류할 수 있다.

[0085] 본 발명의 일 실시예에 따른 시맨틱 콘텐츠 분류 장치(100) 및 이를 이용한 콘텐츠 분류 방법은 기 설정된 비대칭적인 사전 확률에 따라, 시맨틱 토픽 및 시맨틱 감성을 추출하므로, 기존의 대칭적인 사전 확률에 기초한 콘텐츠 분류 방법보다 정확하게 콘텐츠에 포함된 시맨틱 토픽 및 시맨틱 감성을 추출할 수 있다.

[0086] 또한, 시맨틱 콘텐츠 분류 장치(100) 및 이를 이용한 콘텐츠 분류 방법은 추출된 시맨틱 토픽 및 시맨틱 감성에 따라 콘텐츠를 자동으로 분류할 수 있다.

[0087] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능한 기록 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능한 기록 매체는 컴퓨터 저장 매체를 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.

[0088] 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.

[0089] 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.

[0090] 본 발명의 범위는 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

[0091] 100: 시맨틱 콘텐츠 분류 장치

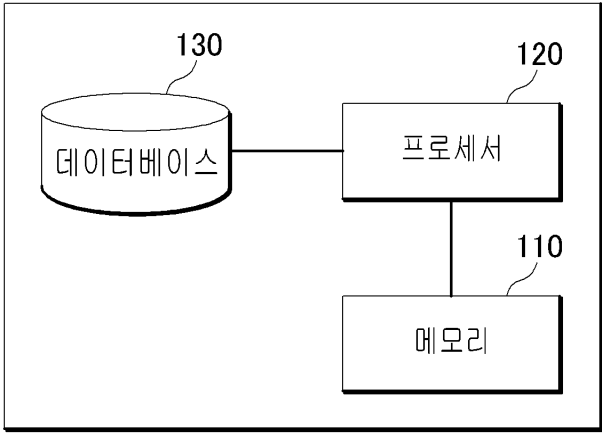
110: 메모리

120: 프로세서

130: 데이터베이스

도면

도면1



100

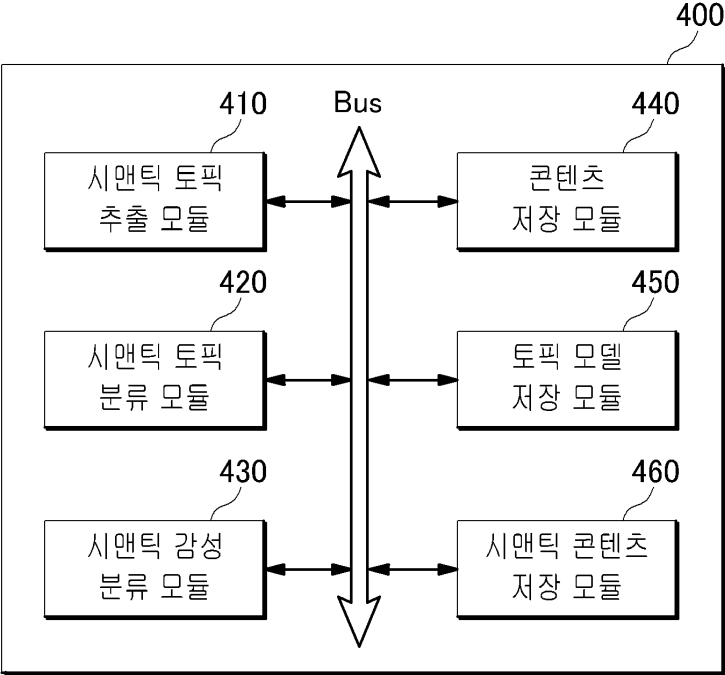
도면2

M, N Y, S T K^{gl}, K^{loc} v, w z, l, d r $\theta_{d,l}^{gl}$ $\theta_{d,l}^{loc}$ $\alpha^{gl}, \alpha^{loc}$ α_r^{mix} β, δ, γ	각각 콘텐츠의 개수 및 콘텐츠 내 단어의 개수 각각 감성의 개수 및 문장의 개수 슬라이딩 윈도우 내 문장의 개수 각각 전역 토픽의 개수 및 지역 토픽의 개수 각각 슬라이딩 윈도우 및 단어 각각 토픽, 감성, 및 문서 전역 토픽인지 지역 토픽인지 여부를 나타내는 토픽 문맥 변수 전역 토픽의 다항 분포 지역 토픽의 다항 분포 각각 전역 토픽 분포 및 지역 토픽 분포에 대한 사전확률 문맥 분포에 대한 사전확률 각각 어휘 분포, 감성 분포, 및 슬라이딩 윈도우 분포에 대한 사전확률
$n^{gl,z,l}$ $n_w^{gl,z,l}$ $n^{loc,z,l}$ $n_w^{loc,z,l}$ $n_v^{d,s}$ $n^{d,s}$ $n^{d,v}$ $n_{gl}^{d,v}$ $n_{loc}^{d,v}$ $n^{d,gl}$ $n_l^{d,gl}$ $n_z^{d,gl,l}$ $n_l^{d,v,loc}$ $n_z^{d,v,l,loc}$	전역 토픽 z 및 감성 l에 할당된 단어 w의 개수 전역 토픽 z 및 감성 l에 할당된 단어 w의 횟수 지역 토픽 z 및 감성 l에 할당된 단어 w의 개수 지역 토픽 z 및 감성 l에 할당된 단어 w의 횟수 문서 d의 슬라이딩 윈도우 v에 할당된 문장 s에서 추출한 단어의 개수 문서 d의 문장 s에 있는 단어의 개수 문서 d의 전역 토픽에 할당된 윈도우 v에서 추출한 단어의 개수 문서 d의 윈도우 v에 할당된 단어의 개수 문서 d의 전역 토픽에 할당된 단어의 개수 문서 d의 전역 토픽 및 감성 l에 할당된 단어의 개수 문서 d의 전역 토픽 z 및 감성 l에 할당된 단어의 개수 문서 d의 지역 토픽에 할당된 단어의 개수 문서 d의 지역 토픽 및 감성 l에 할당된 단어의 개수 문서 d의 지역 토픽 z 및 감성 l에 할당된 단어의 개수

도면3

	토픽 1	토픽 2	토픽 3	...
긍정	콘텐츠 1 콘텐츠 14 ...	콘텐츠 20 콘텐츠 3 ...	콘텐츠 10 콘텐츠 2 ...	...
부정	콘텐츠 10 콘텐츠 21 ...	콘텐츠 14 콘텐츠 5 ...	콘텐츠 7 콘텐츠 11 ...	...

도면4



도면5

