



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2014-0053717
(43) 공개일자 2014년05월08일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01) G06F 17/20 (2006.01)
G06F 17/27 (2006.01)
(21) 출원번호 10-2012-0119977
(22) 출원일자 2012년10월26일
심사청구일자 2012년10월26일

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암
동5가)
(72) 발명자
강재우
경기 가평군 청평면 경춘로 277-245,
(74) 대리인
특허법인엠에이피에스

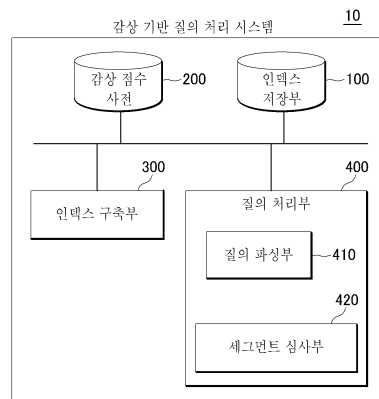
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 감상 기반 질의 처리 시스템 및 방법

(57) 요약

본 발명은 하나 이상의 문서를 하나 이상의 세그먼트(segment)로 분할하고, 상기 세그먼트에 기술된 의견의 대상(object)이 갖는 속성(aspect)을 나타내는 속성 키워드 및 상기 속성에 대한 문서 작성자의 감상(sentiment)을 나타내는 감상 키워드를 상기 세그먼트에서 추출하여 속성-감상 쌍(pair)을 생성하고, 상기 세그먼트의 내용 및 상기 속성-감상 쌍을 포함하는 인덱스(index)를 구축하는 인덱스 구축부; 상기 인덱스가 저장되는 인덱스 저장부; 및 상기 인덱스 저장부에 저장되어 있는 인덱스를 기초로 질의를 처리하여, 상기 질의와 관련된 의견을 기술하고 있는 문서 또는 상기 질의와 관련된 의견이 기술하고 있는 대상을 검색하여 반환하는 질의 처리부;를 포함하는 감상 기반 질의 처리 시스템을 제공한다.

대표도 - 도1



특허청구의 범위

청구항 1

감상 기반 질의 처리 시스템에 있어서,

하나 이상의 문서를 하나 이상의 세그먼트(segment)로 분할하고, 상기 세그먼트에 기술된 의견의 대상(object)이 갖는 속성(aspect)을 나타내는 속성 키워드 및 상기 속성에 대한 문서 작성자의 감상(sentiment)을 나타내는 감상 키워드를 상기 세그먼트에서 추출하여 속성-감상 쌍(pair)을 생성하고, 상기 세그먼트의 내용 및 상기 속성-감상 쌍을 포함하는 인덱스(index)를 구축하는 인덱스 구축부;

상기 인덱스가 저장되는 인덱스 저장부; 및

상기 인덱스 저장부에 저장되어 있는 인덱스를 기초로 질의를 처리하여, 상기 질의와 관련된 의견을 기술하고 있는 문서 또는 상기 질의와 관련된 의견이 기술하고 있는 대상을 검색하여 반환하는 질의 처리부;를 포함하는 감상 기반 질의 처리 시스템.

청구항 2

제 1 항에 있어서,

상기 세그먼트는 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장을 하나 이상 포함하도록 분할되는 감상 기반 질의 처리 시스템.

청구항 3

제 1 항에 있어서,

상기 질의 처리부는

상기 질의에 대한 파싱(parsing)을 수행하는 질의 파싱부; 및

상기 인덱스에 포함되어 있는 각 세그먼트에 대해 상기 세그먼트의 내용 및 상기 속성-감상 쌍에 기초하여 상기 질의와의 관련성을 심사하여 세그먼트 점수를 산출하는 세그먼트 심사부;를 포함하는 감상 기반 질의 처리 시스템.

청구항 4

제 3 항에 있어서,

상기 질의 처리부는

상기 세그먼트 심사부가 산출한 세그먼트 점수를 합산함으로써 상기 세그먼트가 포함되어 있는 문서 또는 상기 세그먼트가 기술하고 있는 대상에 대해 상기 질의와의 관련성을 심사하는 감상 기반 질의 처리 시스템.

청구항 5

제 3 항에 있어서,

상기 질의 파싱부는

상기 질의에서 감상을 나타내는 키워드에 기초하여 질의의 극성 부호를 산출하고, 상기 감상을 나타내는 키워드 중 감상의 극성만을 나타내는 키워드를 제거하는 감상 기반 질의 처리 시스템.

청구항 6

제 3 항에 있어서,

상기 질의 파싱부는

상기 질의에서 상기 대상이 속하는 분야(domain)를 나타내는 키워드를 제거하는 감상 기반 질의 처리 시스템.

청구항 7

제 3 항에 있어서,

상기 질의 파싱부는

상기 질의를 속성을 나타내는 키워드를 기준으로 하나 이상의 의미 단위로 분할하며,

상기 세그먼트 심사부는

상기 분할된 의미 단위에 각각에 대해 세그먼트 점수를 산출하는 감상 기반 질의 처리 시스템.

청구항 8

제 3 항에 있어서,

상기 세그먼트 심사부는

세그먼트의 내용이 상기 파싱된 질의에 포함되어 있는 키워드를 포함하고 있는 세그먼트를 검색한 후, 상기 검색된 세그먼트에서 상기 파싱된 질의에 포함된 속성 키워드에 대응하는 속성-감상 쌍을 찾고, 상기 검색된 속성-감상 쌍의 감상 점수에 기초하여 산출한 속성-감상 쌍 점수에 상기 극성 부호를 곱하여 상기 검색된 세그먼트의 세그먼트 점수를 산출하는 감상 기반 질의 처리 시스템.

청구항 9

제 3 항에 있어서,

상기 세그먼트 심사부는

상기 파싱된 질의에 속성 키워드가 포함되어 있지 않을 경우, 상기 검색된 세그먼트가 포함하고 있는 모든 속성-감상 쌍의 감상 점수에 기초하여 산출한 속성-감상 쌍 점수를 산출하는 감상 기반 질의 처리 시스템.

청구항 10

제 3 항에 있어서,

상기 시스템은 각 감상 키워드별로 기 지정된 극성 가중치 점수를 저장하는 감상 점수 사전;을 더 포함하며,

상기 세그먼트 심사부는 속성-감상 쌍이 포함하고 있는 감상의 극성 가중치 점수를 상기 감상 점수 사전에서 검색함으로써, 상기 속성-감상 쌍의 감상 점수를 산출하는 감상 기반 질의 처리 시스템.

청구항 11

감상 기반 질의 처리 시스템을 사용한 감상 기반 질의 처리 방법에 있어서,

(a) 하나 이상의 문서를 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장을 하나 이상 포함하는 하나 이상의 세그먼트(segment)로 분할하는 단계;

(b) 상기 세그먼트에 기술된 의견의 대상(object)이 갖는 하나의 속성(aspect)을 나타내는 속성 키워드 및 상기 속성에 대한 문서 작성자의 감상(sentiment)을 나타내는 감상 키워드를 상기 세그먼트에서 추출하여 속성-감상 쌍(pair)을 생성하는 단계;

(c) 상기 세그먼트의 내용 및 상기 속성-감상 쌍을 포함하는 인덱스(index)를 구축하는 단계;

(d) 수신한 질의에 대한 파싱(parsing)을 수행하여, 상기 질의에서 감상을 나타내는 키워드에 기초하여 질의의 극성 부호를 산출하고, 상기 감상을 나타내는 키워드 중 감상의 극성만을 나타내는 키워드를 제거하는 단계;

(e) 상기 인덱스에 포함되어 있는 각 세그먼트에 대해 상기 세그먼트의 내용 및 상기 속성-감상 쌍에 기초하여 상기 질의와의 관련성을 심사하여 세그먼트 점수를 산출하는 단계; 및

(f) 상기 세그먼트 심사부가 산출한 세그먼트 점수를 합산하여 상기 문서 또는 상기 대상에 대해 상기 질의에 대한 관련성을 심사하는 단계;를 포함하는 감상 기반 질의 처리 방법.

청구항 12

제 11 항에 있어서,

상기 (d) 단계는

상기 질의에서 상기 대상이 속하는 분야(domain)를 나타내는 키워드를 제거하는 감상 기반 질의 처리 방법.

청구항 13

제 11 항에 있어서,

상기 (d) 단계는

상기 질의를 속성을 나타내는 키워드를 기준으로 하나 이상의 의미 단위로 분할하며,

상기 (e) 단계는

상기 분할된 의미 단위에 각각에 대해 세그먼트 점수를 산출하는 감상 기반 질의 처리 방법.

청구항 14

제 11 항에 있어서,

상기 (e) 단계는

(e1) 세그먼트의 내용이 상기 파싱된 질의에 포함되어 있는 키워드를 포함하고 있는 세그먼트를 검색하는 단계;

(e2) 상기 검색된 세그먼트에서 상기 파싱된 질의에 포함된 속성 키워드에 대응하는 속성-감상 쌍을 찾고, 상기 검색된 속성-감상 쌍의 감상 점수를 기초로 상기 검색된 세그먼트의 속성-감상 쌍 점수를 산출하는 단계;

(e3) 상기 극성 부호와 상기 속성-감상 쌍 점수를 기초로 상기 세그먼트 점수를 산출하는 단계;를 포함하는 감상 기반 질의 처리 방법.

청구항 15

제 14 항에 있어서,

상기 (e2) 단계는

상기 파싱된 질의에 속성 키워드가 포함되어 있지 않을 경우, 상기 검색된 세그먼트가 포함하고 있는 모든 속성-감상 쌍의 감상 점수를 기초로 속성-감상 쌍 점수를 산출하는 감상 기반 질의 처리 방법.

청구항 16

제 14 항에 있어서,

상기 (e2) 단계는 속성-감상 쌍이 포함하고 있는 감상의 극성 가중치 점수를 상기 감상 점수 사전에서 검색함으로써, 상기 속성-감상 쌍의 감상 점수를 산출하는 감상 기반 질의 처리 방법.

명 세 서

기술 분야

[0001] 본 발명은 감상 기반 질의 처리 시스템 및 방법에 관한 것이다.

배 경 기 술

[0002] 사용자의 질의를 처리하는 기술은 최근 가장 많은 주목을 받고 있는 분야 중 하나이다. 특히, 질의 대상(object)의 객관적인 속성(aspect) 뿐 아니라 해당 속성에 대한 감상(sentiment)까지 처리할 수 있도록 하기 위한 연구가 많이 이루어지고 있다.

[0003] 예를 들어, 질의 대상이 영화일 때, 즉, 사용자가 영화를 검색하려고 할 때, 영화의 연출, 각본, 주연이 누구인가 등의 객관적인 속성에 대한 질의 뿐 아니라, 연출이 좋았는지, 각본은 흥미진진했는지 등 해당 속성에 대한 주관적인 감상에 대한 질의까지 처리할 수 있도록 하려는 것이다.

[0004] 종래 기술은 주관적인 의견이나 감상에 대해 질의하였을 때 반환하는 검색 결과의 정확도가 떨어지는 문제가 있

다. 예를 들어, "연기가 좋은 영화"라는 질의에 대해 "각본은 좋았는데 배우들의 연기는 별로였다"라는 의견이 기술된 문서가 검색될 수 있다. 이에 따라 사용자는 질의와의 연관성이 낮은 검색 결과를 탐색하며 스스로 결과를 걸러내거나 새로운 질의를 가지고 재시도하는 등의 불편을 겪어야 한다.

[0005] 따라서, 이러한 주관적인 감상 및 의견을 반영하여 질의를 처리하고 정확한 검색 결과를 반환할 수 있는 감상 기반 질의 처리 시스템 및 방법이 필요하다. 주관적인 감상이 포함된, 질의 범위가 비교적 모호한 질의에 대해서도 질의와의 관련성이 높은 결과만이 반환되므로, 사용자의 검색 편의성이 크게 향상될 것이다.

[0006] 질의 처리와 관련하여 한국공개특허 제10-2009-0048997호("키워드를 통한 여론 자료 수집 방법 및 시스템과 이를 위한 기록 매체")에는 키워드를 기반으로 여론 자료를 수집하는 구성이 개시되어 있다.

[0007] 또한, 한국공개특허 제10-2011-0038247호("키워드 추출 장치 및 방법")는 게시물 및 확장된 유사 문서로부터 키워드를 추출하는 구성이 개시되어 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 감상 기반 질의 처리 문제를 해결하기 위한 것으로서, 그 목적은 주관적인 감상에 대한 질의를 처리하여 정확한 검색 결과를 반환할 수 있는 감상 기반 질의 처리 시스템 및 방법을 제공하는 것이다.

과제의 해결 수단

[0009] 상기와 같은 목적을 달성하기 위한 본 발명의 제 1 측면에 따른 감상 기반 질의 처리 시스템은, 하나 이상의 문서를 하나 이상의 세그먼트(segment)로 분할하고, 상기 세그먼트에 기술된 의견의 대상(object)이 갖는 속성(aspect)을 나타내는 속성 키워드 및 상기 속성에 대한 문서 작성자의 감상(sentiment)을 나타내는 감상 키워드를 상기 세그먼트에서 추출하여 속성-감상 쌍(pair)을 생성하고, 상기 세그먼트의 내용 및 상기 속성-감상 쌍을 포함하는 인덱스(index)를 구축하는 인덱스 구축부; 상기 인덱스가 저장되는 인덱스 저장부; 및 상기 인덱스 저장부에 저장되어 있는 인덱스를 기초로 질의를 처리하여, 상기 질의와 관련된 의견을 기술하고 있는 문서 또는 상기 질의와 관련된 의견이 기술하고 있는 대상을 검색하여 반환하는 질의 처리부;를 포함하는 것을 특징으로 한다.

[0010] 상기와 같은 목적을 달성하기 위한 본 발명의 제 2 측면에 따른 감상 기반 질의 처리 방법은 (a) 하나 이상의 문서를 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장을 하나 이상 포함하는 하나 이상의 세그먼트(segment)로 분할하는 단계; (b) 상기 세그먼트에 기술된 의견의 대상(object)이 갖는 하나의 속성(aspect)을 나타내는 속성 키워드 및 상기 속성에 대한 문서 작성자의 감상(sentiment)을 나타내는 감상 키워드를 상기 세그먼트에서 추출하여 속성-감상 쌍(pair)을 생성하는 단계; (c) 상기 세그먼트의 내용 및 상기 속성-감상 쌍을 포함하는 인덱스(index)를 구축하는 단계; (d) 수신한 질의에 대한 파싱(parsing)을 수행하여, 상기 질의에서 감상을 나타내는 키워드에 기초하여 질의의 극성 부호를 산출하고, 상기 감상을 나타내는 키워드 중 감상의 극성만을 나타내는 키워드를 제거하는 단계; (e) 상기 인덱스에 포함되어 있는 각 세그먼트에 대해 상기 세그먼트의 내용 및 상기 속성-감상 쌍에 기초하여 상기 질의와의 관련성을 심사하여 세그먼트 점수를 산출하는 단계; 및 (f) 상기 세그먼트 심사부가 산출한 세그먼트 점수를 합산하여 상기 문서 또는 상기 대상에 대해 상기 질의에 대한 관련성을 심사하는 단계;를 포함하는 것을 특징으로 한다.

발명의 효과

[0011] 본 발명은 감상 기반 질의 처리 시스템 및 방법에 있어, 정확한 검색 결과를 반환할 수 있다는 효과를 얻는다.

[0012] 주관적인 감상이 포함된, 질의 범위가 비교적 모호한 질의에 대해서도 질의와의 관련성이 높은 결과만이 반환되므로, 사용자의 검색 편의성이 크게 향상된다. 예를 들어, 사용자는 질의와의 연관성이 낮은 검색 결과를 탐색하며 스스로 결과를 걸러낼 필요가 없다. 더욱 중요하게는, 사용자는 자신이 원하는 결과를 얻기 위하여 질의 키워드와 표현을 신중하게 선택할 필요가 없다. 질의 키워드를 객관적인 속성에 대한 명확한 범위의 값만으로 한정시키지 않아도 되므로, 사용자는 검색하기 원하는 모호한 개념을 명확한 질의어로 정제하는 과정을 거치지 않고 그대로 사용하여도 된다.

[0013] 따라서, 사용자의 의사 결정을 돕는 도구로 사용될 수 있다. 사용자는 다른 사람들의 의견을 효과적으로 검색할 수 있으므로, 자신의 의사 결정에 다른 많은 사람들의 경험 및 의견을 참조할 수 있다.

[0014] 또한, 본 발명은 질의 처리 과정이 단순하고 효과적이다. 예를 들어, 질의에 포함된 감상의 극성만을 나타내는 키워드를 동의어 및 유의어로 확장하지 않고, 감상의 극성 부호만을 고려하므로, 질의 처리 속도가 빠르면서도, 질의와 관련한 의견을 빠뜨리지 않고 검색해낼 수 있다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일실시예에 따른 감상 기반 질의 처리 시스템의 구조를 도시함.
 도 2는 본 발명의 일실시예에 따른 감상의 극성 가중치 점수를 도시함.
 도 3은 본 발명의 일실시예에 따른 의견이 표현된 문서를 도시함.
 도 4는 도 3의 세그먼트가 포함하는 세그먼트 내용 및 속성-감상 쌍을 도시함.
 도 5는 본 발명의 일실시예에 따른 파싱된 질의를 도시함.
 도 6은 본 발명의 또다른 실시예에 따른 파싱된 질의를 도시함.
 도 7은 본 발명의 또다른 실시예에 따른 파싱된 질의를 도시함.
 도 8은 도 5의 질의에 대해 도 4의 세그먼트를 심사하는 실시예를 도시함.
 도 9는 도 6의 질의에 대해 도 4의 세그먼트를 심사하는 실시예를 도시함.
 도 10은 도 7의 질의에 대해 도 4의 세그먼트를 심사하는 실시예를 도시함.
 도 11은 본 발명의 일실시예에 따른 인덱스 구축 방법의 흐름을 도시함.
 도 12는 본 발명의 일실시예에 따른 질의 파싱 방법의 흐름을 도시함.
 도 13은 본 발명의 일실시예에 따른 세그먼트 심사 방법의 흐름을 도시함.

발명을 실시하기 위한 구체적인 내용

[0016] 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.

[0017] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0018] 도 1은 본 발명의 일실시예에 따른 감상 기반 질의 처리 시스템(10)을 나타낸 블록도이다.

[0019] 먼저 도 1을 참조하면, 본 발명에 따른 감상 기반 질의 처리 시스템(10)은 감상 점수 사전(200), 인덱스 저장부(100), 인덱스 구축부(300), 질의 처리부(400)를 포함한다. 간략하게 설명하자면, 인덱스 구축부(300)가 의견이 기술된 하나 이상의 문서를 기초로 질의 처리에 사용할 인덱스를 구축하여 인덱스 저장부(100)에 저장하면, 질의 처리부(400)가 인덱스 저장부(100)에 있는 인덱스 및 감상 점수 사전(200)에 정의되어 있는 감상의 극성 가중치 점수를 기초로 질의를 처리한다. 본 발명의 일실시예에 따른 인덱스는 의미 단위를 기초로 문서를 세그먼트로 분할하되, 속성-감상 쌍을 세그먼트 내용과 함께 포함할 수 있다.

[0020] 이에 대해 자세히 기술하기 전에 속성(aspect)와 감상(sentiment)이 무엇인지를 먼저 설명한다.

[0021] 속성은 질의 대상(object)이 갖는 여러 가지 특징이다. 책을 예로 들어 설명하자면, 제목, 작가, 분야, 가격 등이 책이라는 질의 대상이 갖는 속성이다. 번역서라면 번역가 등의 속성을 추가적으로 더 가질 것이다. 사용자는 질의에 속성을 사용하여 원하는 대상을 검색할 수 있다. 예를 들어, 사용자는 제목에 "홈즈"가 포함되어 있는 책을 검색하거나, 작가가 "코난 도일"인 책을 검색할 수 있다. 여기에서 "홈즈", "코난 도일"은 속성이 가지는 객관적인 값이다. 이러한 속성의 객관적인 값을 이용하여 검색하는 것은 종래의 질의 처리 기술로도 수행할 수 있다.

[0022] 그러나 이러한 객관적인 질의는 사용자가 명확한 정보를 가지고 있어야 한다. 예를 들어, 사용자는 작가가 코난

도일이라는 정보를 가지고 있어야 한다. 하지만 사용자는 "가장 매력적인 주인공을 창조한 추리 소설 작가"라는 지극히 주관적인 질의를 사용하고 싶어할 수도 있다. 이러한 질의는 사용자가 명확한 정보를 가지고 있지 않거나, 다른 사용자들의 의견(opinion)을 알고 싶을 때 사용할 수 있다.

[0023] 이러한 질의는 상술한 예와 같이, 작가라는 속성에 대한 질의에 "가장 매력적인"이라는 주관적인 감상이 포함된다. 본 발명은 이러한 감상 기반의 주관적인 질의를 처리하기 위해, 의견이 기술된 문서에서 속성과 해당 속성에 대한 문서 작성자의 감상을 추출하여 생성한 속성-감상 쌍을 사용한다.

[0024] 상기한 질의에 대한 결과로 "애거서 크리스티의 추리 소설은 흥미진진하고 매력적이지만, 주인공인 포와로는 그다지 매력적이지 않은 것 같다. 애거서 크리스티라는 작가는 조금 우스꽝스러운 주인공을 창조해냈다."라는 의견이 포함되어 있는 문서가 반환되었다면, 이 결과는 질의와의 관련성이 낮은 의견이므로 부정확한 결과이다. 소설 자체는 매력적이지만, 주인공인 포와로는 매력적이지 않다는 의견을 기술하고 있는데도, "매력적이다"와 "주인공"이 포함되어 있기 때문에 반환되었다.

[0025] 종래 기술은 이 예에서 볼 수 있는 바와 같이, 주관적인 감상이 포함된 질의에 대해 정확하지 않은 검색 결과를 반환하는 경우가 많다. 반면, 본 발명은 속성-감상 쌍을 사용하여 주관적인 감상 기반의 질의에 대해서도 정확한 검색 결과를 반환한다. 따라서, 본 발명은 전술한 바와 같이, 사용자의 검색 편의성을 향상시킨다.

[0026] 감상 기반 질의에 대해 정확한 검색 결과를 반환하기 위해, 문서를 동일한 의미적 관계를 갖는 최소의 구, 절, 또는 문장 단위로 분할하고, 분할된 각 세그먼트(segment)를 각각 인덱싱하는 방법도 생각해볼 수 있다.

[0027] 예를 들어, 질의 대상이 영화일 때, "지난 주말에 여자 친구와 보러 갔었다. 각본은 좋았는데 배우들의 연기는 별로였다. 그러나 전체적으로는 괜찮은 영화라고 생각한다. 재미있게 봤다."라는 영화에 대한 의견이 기술된 문서를 고려해보자.

[0028] "연기가 좋은 영화"라는 질의에 대해, 앞서 기술한 예와 마찬가지로, 위 문서는 "연기"와 "좋다"를 포함하기 때문에 검색 결과로 반환될 수 있다. 이를 막기 위해 위 문서를 "지난 주말에 여자 친구와 보러 갔었다.", "각본은 좋았는데", "배우들의 연기는 별로였다.", "그러나 전체적으로는 괜찮은 영화라고 생각한다.", "재미있게 봤다."와 같이 문장의 최소 의미 단위로 분할하여 분할된 각 세그먼트를 각각 인덱싱할 수 있다. 그러면 질의에 대해 어느 세그먼트도 매칭되지 않으므로, 이 문서는 검색 결과로 반환되지 않는다.

[0029] 그러나 이러한 접근 방법은 정확도를 높이는 데에는 도움이 되지만, 세그먼트의 단위가 너무 작아서 생기는 문제를 갖고 있다. 예를 들어, "여자 친구와 보기 좋은 영화"라는 질의에 대해, 이 문서는 질의와 유관함에도 불구하고 검색 결과로 반환되지 못한다. 첫번째 세그먼트가 "여자 친구"와 "보다"를 포함하고 있어 매칭이 되기는 하지만, 이 세그먼트의 내용은 감상을 포함하고 있지 않으므로 이것만으로는 좋았는지 나빴는지를 판단할 수 없다. 이러한 종류의 질의를 처리하기 위해서는, 의견을 하나의 세그먼트로 처리하는 것이 바람직할 것이다.

[0030] 따라서, 본 발명은 세그먼트를 복수개의 최소 의미 단위를 포함하도록 토픽(topic) 단위로 분할한다. 토픽 단위로 나누는 방법에는 제한이 없다. 자연어 처리 연구를 통해 알려진 종래의 기술들을 사용할 수도 있고, 단순히 몇 개의 문장 단위로 쪼개는 것도 가능하다. 예를 들어, 기지정된 단위가 5라면, 문서를 5개의 문장씩 잘라 분할할 수 있다.

[0031] 본 발명은 이렇게 세그먼트의 단위를 크게 하면서도 전술한 예에서와 같은 부정확한 검색 결과가 반환되는 것을 방지하기 위해, 전술한 바와 같이 세그먼트 내용과 함께 속성-감상 쌍을 함께 포함하도록 인덱스를 구축한다.

[0032] 따라서, 본 발명의 일실시예에 따른 인덱스 구축부(300)는 하나 이상의 문서를 하나 이상의 세그먼트로 분할하고, 세그먼트에 기술된 의견의 대상이 갖는 속성을 나타내는 속성 키워드 및 속성에 대한 문서 작성자의 감상을 나타내는 감상 키워드를 세그먼트에서 추출하여 속성-감상 쌍을 생성하고, 세그먼트의 내용 및 속성-감상 쌍을 포함하는 인덱스를 구축하여 인덱스 저장부(100)에 저장한다. 또한, 본 발명의 일 실시예에 따른 질의 처리부(400)는 인덱스 저장부(100)에 저장되어 있는 인덱스를 기초로 질의를 처리하여, 질의와 관련된 의견을 기술하고 있는 문서 또는 질의와 관련된 의견이 기술하고 있는 대상을 검색하여 반환한다.

[0033] 일실시예에서, 본 발명에 따른 감상 기반 질의 처리 시스템(10)은 분야(domain)별로 인덱스를 구축할 수 있다. 예를 들어, 질의의 대상이 영화라면, 의견의 대상이 영화인 문서, 즉, 영화에 대한 의견을 기술한 문서에 대해 구축된 인덱스를 기초로 질의를 수행할 수 있다. 또다른 예로, 질의의 대상이 책이라면, 의견의 대상이 책인 문서, 즉, 책에 대한 의견을 기술한 문서에 대해 구축된 인덱스를 기초로 질의를 수행할 수 있다. 이러한 실시예에서는 질의에서 분야를 나타내는 키워드를 제거하고 질의를 수행할 수 있으므로, 검색해야 할 인덱스가 적어

질의 속도가 향상될 수 있다. 그러나 분야를 단지 속성의 하나로 취급하여 인덱싱하고 질의하는 것도 가능하다. 자세한 내용은 도 5 내지 도 7을 통해 후술한다.

- [0034] 전술한 바와 같이, 일실시예에서, 질의 처리부(400)는 질의와 관련된 의견을 기술하고 있는 문서를 반환할 수 있다. 예를 들어, "연기가 좋은 영화"에 대해 "마지막 장면에서 주연 배우의 화면을 응시하는 연기에 전율을 느꼈다. 그는 정말 훌륭한 배우이다."라고 기술되어 있는 문서가 반환될 수 있다. 이때 문서가 반환되는 방법에는 제한이 없다. 문서의 전체 내용 또는 문서의 해당 부분의 내용이 반환될 수도 있고, 문서의 URL이 반환될 수도 있다. 특히 문서가 온라인 리뷰라면 해당 부분의 내용과 문서의 URL이 함께 반환되는 것이 바람직할 것이다. 또는 질의 처리부(400)는 질의와 관련된 의견이 기술하고 있는 대상 자체를 반환할 수도 있다. 예를 들어, 상기한 문서가 기술하고 있는 의견의 대상이 영화 "살인의 추억"이라면, 살인의 추억에 대한 정보가 반환될 수 있을 것이다.
- [0035] 질의 처리부(400)는 질의에 대한 파싱(parsing)을 수행하는 질의 파싱부(410), 및 인덱스에 포함되어 있는 각 세그먼트의 세그먼트 내용 및 속성-감상 쌍에 기초하여 질의와의 관련성을 심사하여 세그먼트 점수를 산출하는 세그먼트 심사부(420)를 포함한다. 세그먼트 심사부(420)가 산출한 각 세그먼트 점수는 합산되어, 해당 세그먼트가 포함되어 있는 각 문서 또는 해당 세그먼트가 기술하고 있는 대상의 질의와의 관련성을 심사하는 데 사용된다.
- [0036] 질의 파싱부(410)는 불용어 제거 등의 전처리를 수행할 수 있으나, 이는 종래 기술이므로 자세한 설명을 생략한다. 질의 파싱부(410)는 질의를 파싱하여 속성을 나타내는 키워드, 감상을 나타내는 키워드, 및 분야를 나타내는 키워드를 추출한다. 전술한 바와 같이 일실시예에서 분야를 나타내는 키워드를 제거할 수 있다.
- [0037] 속성을 나타내는 키워드가 두개 이상 포함되어 있을 경우, 질의 파싱부(410)는 질의를 속성을 나타내는 키워드를 기준으로 하나 이상의 의미 단위로 분할하며, 세그먼트 심사부(420)는 분할된 의미 단위 각각에 대해 세그먼트 점수를 산출한다. 예를 들어, "연기가 좋고, 각본이 좋은 영화"라는 질의는 "연기가 좋다"와 "각본이 좋다"라는 두개의 의미 단위로 분할되어 각각 처리한 후, 각 의미 단위에 대해 산출된 세그먼트 또는 문서 또는 대상 점수를 합산함으로써, 전체 질의에 대한 세그먼트 또는 문서 또는 대상 점수를 산출할 수 있을 것이다.
- [0038] 질의 파싱부(410)는 감상을 나타내는 키워드에 기초하여 질의의 극성 부호를 산출한 후, 그 중에서 감상의 극성만을 나타내는 키워드를 제거한다. 이에 대해 설명하기 위해 먼저 도 2를 참조한다.
- [0039] 도 2는 본 발명의 일실시예에 따른 감상의 극성 가중치 점수를 도시하고 있다.
- [0040] 도면은 설명의 편의를 위해 감상의 극성 및 가중치를 수직선 상에 표현하였다. 대상에 대한 긍정적인 감상은 + 극성을, 부정적인 감상은 - 극성을 가진다. 또한, 긍정적 또는 부정적인 강도는 가중치로 표현할 수 있다. 예를 들어, 본실시예에서 "좋다"와 "나쁘다"는 각각 "+2"와 "-2"로, 그보다 강도가 센 "환상적이다"와 "형편없다"는 각각 "+4"와 "-4"로 정의되었다. 긍정적인 감상 및 부정적인 감상을 나타내는 표현은 매우 다양하므로, 당업자는 본 실시예가 단지 설명의 편의를 위해 몇가지의 예를 들고 있을 뿐임을 쉽게 이해할 수 있을 것이다.
- [0041] 이러한 감상의 극성 가중치 점수는 전술한 바와 같이 감상 점수 사전에 기정의될 수 있으며, 질의 처리부(400)에 의해 참조된다. 예를 들어, 질의 파싱부(410)는 이를 기초로 질의의 극성 부호를 산출하며, 세그먼트 심사부(420)는 이를 기초로 속성-감상 쌍의 감상 점수를 산출한다.
- [0042] 다시 도 1로 돌아가서, 본 발명의 일실시예에 따른 질의 파싱부(410)가 감상을 나타내는 키워드에 기초하여 질의의 극성 부호를 산출한 후 그 중 감상의 극성만을 나타내는 키워드를 제거하는 이유는 긍정 또는 부정을 나타내는 표현이 너무 다양하기 때문이다.
- [0043] 예를 들어, "연기가 좋은 영화"와 같이 질의에 "좋다"라는 긍정적인 감상을 나타내는 키워드가 포함되어 있을 경우, "연기가 괜찮다"나 "연기가 환상적이다"와 같은 의견이 기술되어 있는 문서도 검색될 수 있게 하는 것이 바람직할 것이다. 이를 위해 "좋다"의 동의어, 유의어 등으로 질의를 확장하는 방안을 고려해볼 수 있으나, 확장할 수 있는 "좋다"의 동의어, 유의어가 너무 많기 때문에, 이를 모두 포함하도록 질의를 확장하는 것은 매우 비효율적이다. 또한 모든 동의어, 유의어를 포함하도록 질의를 확장할 수 없기 때문에 해당 확장 키워드를 포함하고 있는 의견이 기술된 문서가 검색되지 못할 수 있다.
- [0044] 본 발명은 감상의 극성만을 나타내는 키워드를 제거하는 대신 해당 키워드가 나타내는 극성 부호만을 고려함으로써, 이러한 문제를 해결하였다. 예를 들어, "좋다", "괜찮다", "환상적이다"와 같은 긍정적인 감상 키워드에 대해서는 + 부호, 즉 +1을, "나쁘다", "별로다", "형편없다"와 같은 부정적인 감상 키워드에 대해서는 - 부호,

즉 -1을 산출한다.

- [0045] 단, "재미있다"와 "재미없다"는 감상의 극성만이 아니라 추가적인 감상 정보를 포함하고 있으므로, 극성 부호를 산출하되 제거하지 않는다. 추가적인 예를 들자면, "짱이다", "후지다"은 감상의 극성만을 나타내므로 극성 부호를 산출하고 제거하며, "흥미롭다", "감동적이다", "신난다" 등은 감상의 극성 이외에도 추가적인 감상 정보를 포함하고 있으므로 극성 부호를 산출한 후 제거되지 않는다.
- [0046] 이러한 키워드를 남겨두면 검색 결과의 정확성을 더욱 높일 수 있다. 예를 들어, "신난다"라는 감상 키워드에 대해, 추출된 극성 부호인 +1만을 인덱스 검색에 사용하면, "신난다"라는 감상은 아니나 긍정적인 감상을 기술하고 있는 문서가 "신난다"라는 감상을 기술한 문서보다 높은 점수를 받아 검색 결과로 우선 반환되는 경우가 있을 것이다. 본 발명의 일실시예에 따른 질의 파싱부(410)은 "신난다"라는 질의에 감상 키워드를 남겨둠으로써 이를 방지할 수 있다.
- [0047] 이때, 감상의 극성 이외에도 추가적인 감상 정보를 추가하고 있어 질의에 남겨진 감상 키워드에 대해서는 동의어, 유의어 등으로 확장하는 것이 바람직할 것이다. 이는 전술한 바와 같이, 예를 들어, 질의에 "재미있다"라는 감상 키워드가 포함되어 있다면, "재미있다"라는 의견이 기술되어 있는 문서 뿐 아니라 "흥미롭다"라는 의견이 기술되어 있는 문서도 검색할 수 있도록 하기 위함이다.
- [0048] 한편, 감상 점수 사전에서 해당 키워드가 갖는 극성 가중치 점수의 부호를 검색하면 되므로, 감상을 나타내는 키워드에서 극성 부호를 산출하는 것은 매우 간단하게 수행된다. 또한 의견이 기술된 문서에 긍정적인 감상이 포함되어 있는지 부정적인 감상이 포함되어 있는지를 기준으로 질의를 검색하므로, 감상 키워드 및 유의어, 동의어 각각을 인덱스와 비교하는 과정이 필요없어 검색 속도가 빠르며, 감상의 구체적인 표현과 상관없이 다양한 동의어, 유의어를 모두 처리할 수 있다. 즉, 질의 처리 속도가 빠르면서도, 질의와 관련한 의견을 빠뜨리지 않고 검색해낼 수 있다. 따라서 이러한 방법은 매우 효율적이며 검색 결과의 정확도를 높여준다.
- [0049] 산출된 극성 부호는 세그먼트 심사부(420)에 의해 검색된 세그먼트의 순위를 뒤집기 위해 사용될 수 있다. 산출된 각 세그먼트의 점수에 곱하면 손쉽게 결과의 순서를 뒤집을 수 있을 것임을 당업자는 쉽게 이해할 수 있을 것이다.
- [0050] 사용자는 자신의 의사 결정에 도움을 받기 위해 다른 사람의 의견을 검색하는 경우가 많으므로, 대부분의 경우 질의에 포함되는 감상 키워드의 극성은 긍정적(+)일 것이다. 예를 들어, 무슨 영화를 볼지 선택하기 위해 다른 사람의 의견을 참조하려는 사용자는 일반적으로 "연기가 나쁜 영화"가 아니라 "연기가 좋은 영화"를 검색할 것이다. 따라서 일실시예에서 극성 부호의 기본값을 +1로 두고, 긍정적인 감상을 포함하는 문서를 검색할 수 있다. 사용자가 "연기가 나쁜 영화"를 검색할 경우에는 긍정적인 감상에 대해 검색한 결과에 -1을 곱하면 쉽게 순위를 뒤집을 수 있다.
- [0051] "연기가 나쁜 영화"라는 질의에 질의 파싱부(410)는 "나쁘다"라는 부정적인 감상 키워드 대신 극성 부호로 "-1"을 산출하고, 세그먼트 심사부(420)는 "연기가 좋은 영화"라는 질의를 수신했을 때처럼 긍정적인 감상을 기술한 세그먼트에 대해 검색한다. 예를 들어, 그 결과, 세그먼트 1, 세그먼트 2, 세그먼트 3의 점수가 각각 +0.2, +2, -1이라면, 극성 부호인 -1을 각각의 점수에 곱한 결과는 -0.2, -2, +1이 될 것이므로, 세그먼트 3이 "연기가 나쁜 영화"라는 질의에 대해 가장 관련성이 높은 의견을 기술하고 있는 세그먼트로서 반환된다. 이는 "연기가 좋은 영화"에 대해 세그먼트 2가 가장 관련성이 높은 검색 결과로 반환되는 것만큼이나 정확도가 높은 결과일 것임은 당업자에게 쉽게 이해될 것이다.
- [0052] 세그먼트 심사부(420)는 세그먼트의 내용이 파싱된 질의에 포함되어 있는 키워드를 포함하고 있는 세그먼트를 검색한 후, 검색된 세그먼트에서 파싱된 질의에 포함된 속성 키워드에 대응하는 속성-감상 쌍을 찾고, 검색된 속성-감상 쌍의 감상 점수를 합산 또는 평균 또는 기타 다른 연산을 수행하여, 검색된 세그먼트의 속성-감상 쌍 점수를 산출한다. 파싱된 질의에 속성 키워드가 포함되어 있지 않을 경우에는, 검색된 세그먼트가 포함하고 있는 모든 속성-감상 쌍의 감상 점수를 합산 또는 평균 또는 기타 다른 연산을 수행함으로써 속성-감상 쌍 점수를 산출한다. 속성-감상 쌍의 감상 점수는 속성-감상 쌍이 포함하고 있는 감상의 극성 가중치 점수를 감상 점수 사전에서 검색함으로써 산출할 수 있다. 전술한 바와 같이, 이렇게 산출된 속성-감상 쌍 점수에 극성 부호를 곱하면 해당 세그먼트의 세그먼트 점수가 최종 산출된다.
- [0053] 도 3 내지 도 10의 실시예를 통하여 본 발명의 일실시예에 따른 감상 기반 질의 처리 시스템(10) 및 방법을 좀 더 자세히 설명한다.
- [0054] 도 3은 본 발명의 일실시예에 따른 의견이 표현된 문서를 도시하고 있으며, 도 4는 도 3의 세그먼트가 포함하는

세그먼트 내용 및 속성-감상 쌍을 도시하고 있다.

- [0055] 도시되어 있는 문서 1은 영화에 대한 의견을 기술하고 있으며, "지난 주말에 여자 친구와 보러 갔었다. 각본은 좋았는데 배우들의 연기는 별로였다. 그러나 전체적으로는 괜찮은 영화라고 생각한다. 재미있게 봤다."라는 내용을 담고 있는 세그먼트 1을 포함하고 있다. 세그먼트 1은 전술한 바와 같이 토픽 단위로 분할된 세그먼트이다. 편의상 이후 내용은 세그먼트 1에 대해서만 설명하겠다. 그러나 전술한 바와 같이, 생략된 세그먼트들에 대해서도 세그먼트 점수를 산출하고, 해당 세그먼트들의 점수는 문서 1의 점수를 산출하는 데 사용될 것이다.
- [0056] 세그먼트 1은 "영화"라는 분야 키워드(D)를 포함하고 있다. 전술한 바와 같이 분야를 나타내는 키워드는 실시예에 따라 속성을 나타내는 키워드와 동일하게 취급할 수도 있다. "각본", "연기"는 속성 키워드(A)이며, "좋다", "별로다", "괜찮다", "재미있다"는 감상 키워드(S)이다. 이때, 감상을 나타내는 키워드(S)가 기본형으로 표현된 것은 질의 파싱부(410)가 전처리를 수행하는 것과 마찬가지로 인덱스 구축부(300)도 필요한 전처리를 수행하기 때문이다.
- [0057] 질의를 파싱할 때와 마찬가지로, 일실시예에서 분야 키워드(D)는 인덱스 구축시에도 제외될 수 있다. 또다른 실시예에서는 속성 키워드(A)처럼 취급될 수도 있다. 도 4의 실시예는 "영화"라는 분야 키워드(D)를 제거한 인덱스의 예를 보여주고 있다.
- [0058] 세그먼트 1에서 추출된 각 속성 키워드(A)와 대응하는 감상 키워드(S)로 이루어진 속성-감상 쌍이 생성되어, 세그먼트 내용과 함께 인덱스에 포함되어 있다. 이때, 설명의 편의를 위해 세그먼트 내용이 인덱스에 포함되도록 도시한 것일 뿐, 인덱스의 구성 방법에는 제한이 없다. 예를 들어, 인덱스는 세그먼트 ID 등 해당 세그먼트를 접근하기 위한 정보만을 포함하고, 필요시 해당 정보를 사용하여 세그먼트 1이 포함된 문서를 접근함으로써 세그먼트 1의 내용을 참조하도록 구성할 수 있다. 속성-감상 쌍의 구성 방법에도 제한이 없다.
- [0059] 단, 속성 키워드(A)와 대응되는 감상 키워드(S)가 정확하게 짝을 이루어 저장되도록 하는 것이 바람직하다. 예를 들어, "좋다"라는 감상 키워드(S)는 "연기"라는 속성 키워드(A)가 아니라, "각본"이라는 속성 키워드(A)에 대응되어야 한다.
- [0060] 그러나 속성 키워드(A) 없이 감상 키워드(S)만 있는 속성-감상 쌍이 생성되는 것은 가능하다. 예를 들어, 도시되어 있는 바와 같이, 본실시예에서는 "재미있다"라는 감상 키워드(S)가 대응되는 속성 키워드(A) 없이 추출되었다.
- [0061] 도 5 내지 도 7은 본 발명의 일실시예에 따라 파싱된 질의의 세가지 실시예를 도시하고 있다.
- [0062] 도 5는 긍정적인 감상 키워드(S)를 포함한 질의이고, 도 6은 부정적인 감상 키워드(S)를 포함한 질의이다. 또한 도 7은 긍정적인 감상 키워드(S)를 포함하되, 특정 속성 키워드(A)를 포함하지 않은 질의이다.
- [0063] 전술한 바와 같이, 질의 파싱부(410)는 전처리를 수행하고, 키워드들을 추출한 후, 감상 키워드(S)에 기초하여 극성 부호를 산출하고, 분야 키워드(D) 및 극성만을 나타내는 감상 키워드(S)를 제거한다.
- [0064] 도 5에서는 "좋다"에 의해 극성부호 +, 즉 +1이 추출되었고, 분야 키워드(D) 및 감상 키워드(S)를 제거한 결과 파싱된 질의는 "연기"이다.
- [0065] 도 6에서는 "형편없다"에 의해 극성부호 -, 즉 -1이 추출되었고, 분야 키워드(D) 및 감상 키워드(S)를 제거한 결과 파싱된 질의는 "각본"이다. 이때 도면에는 "평가받다"라는 속성에 대한 감상에 큰 영향을 미치지 않는 키워드도 제외된 것으로 도시되어 있으나, 실시예에 따라 이를 제거하지 않는 것도 가능하다.
- [0066] 도 7에서는 "좋다"에 의해 극성부호 +, 즉 +1이 추출되었고, 분야 키워드(D) 및 감상 키워드(S)를 제거한 결과 파싱된 질의는 "여자친구, 보다"이다.
- [0067] 도 8 내지 도 10은 도 5 내지 도 7의 세가지 질의에 대해 도 4의 세그먼트를 심사하는 실시예를 도시하고 있다.
- [0068] 전술한 바와 같이, 세그먼트 심사부(420)는 파싱된 질의에 포함되어 있는 키워드를 포함하고 있는 세그먼트를 추출한다. 즉, 도 5의 질의에 대해서는 세그먼트의 내용에 "연기"가 포함되어 있는 세그먼트를, 도 6의 질의에 대해서는 세그먼트의 내용에 "각본"이 포함되어 있는 세그먼트를, 도 7의 질의에 대해서는 세그먼트의 내용에 "여자친구, 보다"가 포함되어 있는 세그먼트를 검색한다. 세 경우 모두 도 4의 세그먼트 1이 검색된다.
- [0069] 검색된 세그먼트에 대해 세그먼트 심사부(420)는 파싱된 질의에 포함된 속성 키워드(A)에 대응하는 속성-감상

쌍을 찾는다. 도 5의 질의에 대해서는 "연기"라는 속성 키워드(A)에 대응하는 속성-감상 쌍으로 "연기-별로다"가 검색되었고, 도 6의 질의에 대해서는 "각본"이라는 속성 키워드(A)에 대응하는 속성-감상 쌍으로 "각본-좋다"가 검색되었으며, 도 7의 질의에 대해서는 질의에 포함되어 있는 속성 키워드(A)가 없으므로, 검색된 속성-감상 쌍이 없다.

[0070] 세그먼트 심사부(420)는 검색된 속성-감상 쌍의 감상 점수를 기초로 속성-감상 쌍 점수를 산출한다. 이때, 산출 방법에는 제한이 없으며, 예를 들어 합산 또는 평균 또는 기타 다른 연산이 사용될 수 있다. 예를 들어, 도 5의 질의에 대해 "연기-별로다"라는 하나의 속성-감상 쌍만이 검색되었으므로, 이 속성-감상 쌍의 감상 점수가 속성-감상 쌍 점수로 산출되었지만, 두개 이상의 속성-감상 쌍이 검색되었다면 각 속성-감상 쌍의 감상 점수를 합산 또는 평균한 값이 속성-감상 쌍 점수로 산출될 수 있다.

[0071] 각 속성-감상 쌍의 감상 점수는 전술한 바와 같이, 감상 점수 사전(200)에 기정의된 극성 및 가중치를 기초로 산출될 수 있다. 도 5의 질의에 대해 검색된 "연기-별로다"에 포함되어 있는 "별로다"에 대해 감상 점수 사전(200)에서 검색한 극성 및 가중치로 산출한 감상 점수는 -1이고, 도 6의 질의에 대해 검색된 "각본-좋다"에 포함되어 있는 "좋다"에 대해 감상 점수 사전(200)에서 검색한 극성 및 가중치로 산출한 감상 점수는 +2이다.

[0072] 도 7의 질의는 검색된 속성-감상 쌍이 없다. 이런 경우에는 검색된 세그먼트가 포함하고 있는 모든 속성-감상 쌍의 감상 점수를 기초로 속성-감상 쌍 점수를 산출하면 된다. 이때, 산출 방법에는 제한이 없으며, 예를 들어 합산 또는 평균 또는 기타 다른 연산이 사용될 수 있다. 따라서 도 10에 도시되어 있는 실시예는 평균을 사용하였으며, 그 결과 속성-감상 쌍 점수합으로 +1.25가 산출되었다.

[0073] 검색된 속성-감상 쌍이 없는 경우는 도 7의 질의에서와 같이 질의가 속성 키워드(A)를 포함하고 있지 않을 때 외에도, 질의는 속성 키워드(A)를 포함하고 있는데 검색된 세그먼트에 속성 키워드(A)가 포함되어 있지 않을 때 도 있을 것이다(미도시). 해당 세그먼트가 질의의 속성 키워드(A) 이외의 다른 키워드에 매칭되어 검색된 경우이다. 이 경우 역시 검색된 세그먼트가 포함하고 있는 모든 속성-감상 쌍의 감상 점수를 기초로 속성-감상 쌍 점수를 산출하면 된다. 즉, 세그먼트가 포함하고 있는 모든 감상 키워드(S)의 감상 점수를 기초로 속성-감상 쌍 점수를 산출하는 것이다. 이때도 역시 산출 방법에는 제한이 없으며, 예를 들어 합산 또는 평균 또는 기타 다른 연산이 사용될 수 있다.

[0074] 전술한 바와 같이, 세그먼트 심사부(420)는 이렇게 산출된 속성-감상 쌍 점수 합에 질의 과성부(410)가 산출한 극성 부호를 곱하여 최종 세그먼트 점수를 산출한다. 도 5의 질의에 대해서는 도 8에서 산출한 속성-감상 쌍 점수 합은 -1이고, 극성 부호는 +1이므로, 세그먼트 점수는 -1이 산출되며, 도 6의 질의에 대해서는 도 9에서 산출한 속성-감상 쌍 점수 합은 +2이고, 극성 부호는 -1이므로, 세그먼트 점수는 -2이 산출되고, 도 7의 질의에 대해서는 도 10에서 산출한 속성-감상 쌍 점수 합은 +1.25이고, 극성 부호는 +1이므로, 세그먼트 점수는 +1.25이 산출된다.

[0075] 따라서, 도 8 및 도 9의 실시예에서는 최종 세그먼트 점수가 부정적인 극성을 가지고 있으므로, 세그먼트 1은 각각 도 5 및 도 6의 질의에 대한 검색 결과로 반환되지 않을 것이다. 도 10의 실시예에서는 긍정적인 극성을 갖는 세그먼트 점수가 산출되었으므로, 다른 세그먼트들의 점수와 비교한 결과에 따라 세그먼트 1이 검색 결과로 반환될 수도 있고 반환되지 않을 수도 있다. 예를 들어, 세그먼트 2의 점수가 +3이라면 세그먼트 2가 세그먼트 1에 비해 질의와의 관련성이 더 높은 것이므로, 세그먼트 2가 세그먼트 1보다 우선적으로 반환될 것이다.

[0076] 이때, 세그먼트 1이 검색 결과로 반환되는 것처럼 기술했지만, 실제로는 전술한 바와 같이, 질의 처리부(400)는 세그먼트 심사부(420)에 의해 산출된 세그먼트 점수에 기초하여 세그먼트가 포함되어 있는 문서 또는 세그먼트가 기술하고 있는 대상의 질의와의 관련성을 심사한다. 예를 들어, 문서 1이 세그먼트 1, 세그먼트 2로 분할되었고, 세그먼트 1의 점수가 +1.25, 세그먼트 2의 점수가 +3이라면 문서 1의 점수는 두 점수를 평균한 +2.125가 될 수 있다. 또는 두 점수를 합산한 +4.25가 될 수도 있다. 또는 기타 다른 연산을 수행한 값이 산출될 수도 있다. 이렇게 최종 산출된 문서 1의 점수가 다른 문서들의 점수와 비교되어 가장 높은 점수를 갖는 문서 또는 해당 문서가 기술하고 있는 대상이 검색 결과로 반환된다.

[0077] 이때, 질의와 관련된 대상을 검색 결과로 반환하는 방법에는 전술한 바와 같이 세그먼트 점수를 문서별로 묶어 취합하고 해당 문서가 기술하고 있는 대상을 반환하는 방법 이외에도, 문서에 대한 고려없이 세그먼트 점수를 대상별로 묶어 취합하는 방법도 있다. 예를 들어, 문서 1의 세그먼트 1이 영화 1에 대한 의견을 기술하고 있고, 문서 2의 세그먼트 3이 영화 1에 대한 의견을 기술하고 있어, 세그먼트 심사 결과 산출된 세그먼트 점수가 각각 +1, +2라면, 영화 1의 점수는 두 점수를 평균한 +1.5가 될 수 있다. 이 경우에도 마찬가지로 취합 연산에는 합

산, 평균, 또는 기타 다른 연산이 사용될 수 있다.

- [0078] 이상 기술한 실시예를 통해, 본 발명에 따른 감상 기반 질의 처리 시스템(10)은 속성 키워드(A)에 대해 긍정적인 감상 키워드(S)를 포함한 질의, 속성 키워드(A)에 대해 부정적인 감상 키워드(S)를 포함하는 질의, 특정 속성 키워드(A)를 포함하지 않은 질의 등 다양한 경우의 질의에 대해 효율적이고 정확하게 처리할 수 있음을 알 수 있다. 또한, 도면으로 실시예를 설명하지는 않았으나, 전술한 바와 같이, 두 개 이상의 속성 키워드(A)를 포함하는 질의에 대해서도 효율적으로 처리하는 것이 가능하다.
- [0079] 이하 도 11 내지 도 13을 통해 본 발명의 일실시예에 따른 감상 기반 질의 처리 방법의 흐름을 설명한다.
- [0080] 먼저, 도 11은 본 발명의 일실시예에 따른 인덱스 구축 방법의 흐름을 도시하고 있다.
- [0081] 온라인 리뷰 등 특정 대상에 대한 의견이 기술되어 있는 하나 이상의 문서를 토픽 단위의 세그먼트로 분할한다(S1110). 세그먼트로 분할하는 방법은 전술한 바와 같이 제한이 없다. 예를 들어, 자연어 처리 분야에서 도출된 기술을 사용하거나, 단순히 문서를 일정 개수의 문장 단위로 분할할 수 있다.
- [0082] 각 세그먼트별 속성-감상 쌍을 추출한다(S1120). 대상이 지니고 있는 속성에 대한 의견 작성자의 감상을 추출하여, 속성 키워드(A) 및 감상 키워드(S)가 짝을 이루도록 하는 것이다. 이때, 전술한 바와 같이, 속성 키워드(A)와 감상 키워드(S)의 대응 관계는 정확해야 한다. 이를 위해 전술한 바와 같이 인덱스 구축부(300)는 각 세그먼트를 파싱하며, 그 전에 필요한 전처리를 수행한다. 예를 들어, "좋은"이라는 표현은 "좋다"라는 기본형의 감상 키워드(S)로 추출된다.
- [0083] 세그먼트 내용과 속성-감상 쌍을 포함하는 인덱스를 구축하여 인덱스 저장부(100)에 저장한다(S1130).
- [0084] 도 12는 본 발명의 일실시예에 따른 질의 파싱 방법의 흐름을 도시하고 있다.
- [0085] 질의를 수신하면, 먼저 질의에서 분야 키워드(D), 속성 키워드(A), 및 감상 키워드(S)를 추출한다(S1210). 이를 위해 전술한 바와 같이 질의 파싱부(410)는 질의를 파싱하며, 그 전에 필요한 전처리를 수행한다. 예를 들어, "좋은"이라는 표현은 "좋다"라는 기본형의 감상 키워드(S)로 추출된다.
- [0086] 다음, 속성 키워드를 기준으로 질의를 의미 단위로 분리한다(S1220). 예를 들어, "각본이 좋고, 연기가 괜찮은 영화"는 "각본이 좋다"와 "연기가 괜찮다"라는 두가지 의미를 포함하고 있으므로, 전술한 바와 같이 각각을 분리하여 따로 처리한 후 이후 결과를 통합할 것이다.
- [0087] 다음, 분야 키워드(D)를 제거한다(S1230). 이는 전술한 바와 같이, 인덱스가 특정 분야의 대상에 대해서만 구축되어 있을 때이다. 예를 들어, 영화에 대한 의견을 기술한 문서만을 대상으로 인덱스를 구축했을 때는 "영화"라는 키워드는 필요없으므로 질의에서 생략된다. 그러나, 인덱스가 영화, 책, TV 프로그램 등 다양한 분야에 대해 구축되어 있을 때는 "영화"라는 키워드는 속성 키워드(A)로 간주되어 처리된다.
- [0088] 다음, 극성 부호 산출 후 극성만을 나타내는 감상 키워드(S)를 제거한다(S1240). 감상 키워드(S)에 대해 극성 부호만을 고려한 후 극성만을 나타내는 감상 키워드(S)를 질의에서 제거하는 것은 전술한 바와 같이, 동의어, 유사어로 확장하지 않고도 모든 동의어, 유사어를 처리할 수 있도록 하기 위함이다. 전술한 바와 같이 단지 극성 부호만을 고려하면 되므로, 질의 처리 과정을 크게 단순화시키면서도 질의와 연관함에도 누락되는 문서가 없게 하므로 검색 결과의 정확성을 더욱 높일 수 있다. 단, 전술한 바와 같이, 극성 이외의 추가적인 감상 정보를 나타내는 감상 키워드(S)는 질의에서 제거되지 않고, 이러한 키워드에 대해서는 동의어, 유사어로 확장하는 것이 바람직하다. 예를 들어, "좋다"는 극성만을 나타내므로 제거되지만, "재미있다"는 극성 이외의 추가적인 감상 정보를 포함하므로 제거되지 않으며, "흥미롭다" 등으로 확장될 수 있다.
- [0089] 각 의미 단위가 모두 처리될 때까지(S1250) 상기 단계(S1230, S1240)를 반복하고, 완료되면 세그먼트 심사 단계로 간다.
- [0090] 도 13은 본 발명의 일실시예에 따른 세그먼트 심사 방법의 흐름을 도시하고 있다.
- [0091] 파싱된 질의 키워드를 포함하는 세그먼트를 검색한다(S1310). 인덱스를 검색하여 세그먼트 내용에 해당 키워드가 포함되어 있는 세그먼트를 추출하면 된다.
- [0092] 다음, 질의에 속성 키워드가 포함되어 있다면(S1320), 해당 속성-감상 쌍들의 감상 점수를 합산하고(S1330), 그렇지 않다면 모든 속성-감상 쌍의 감상 점수를 평균함으로써(S1340), 검색된 세그먼트의 속성-감상 쌍 점수를 산출한다. 이때, 전술한 바와 같이, 상기 단계들(S1330 및 S1340)에는 합산 또는 평균 이외의 기타 다른 연산이

사용될 수도 있다. 또한, 속성-감상 쌍의 감상 점수는 속성-감상 쌍이 포함하고 있는 감상 키워드(S)를 사용하여 감상 점수 사전(200)에서 해당 감상 키워드(S)의 극성 가중치를 검색함으로써 산출할 수 있다.

[0093] 다음, 극성 부호를 곱하여 해당 세그먼트의 점수를 산출한다(S1350). 전술한 바와 같이, 극성 부호를 곱함으로써 검색 결과의 순위를 뒤집는 단순한 단계를 통해, 본 발명의 일 실시예에 따른 감상 기반 질의 처리 방법은 긍정적인 감상을 갖는 의견을 검색한 결과를 쉽게 부정적인 감상을 갖는 의견을 검색한 결과에 활용할 수 있다.

[0094] 검색된 세그먼트가 모두 처리될 때까지(S1360), 상기 단계(S1320 내지 S1350)를 반복한다. 완료되면, 각 질의의 의미 단위가 모두 처리될 때까지(S1370), 상기 단계(S1310 내지 S1360)를 반복한다.

[0095] 도시하지 않았으나, 전술한 바와 같이, 이렇게 해서 산출된 세그먼트 점수는 문서 점수 또는 대상 점수를 산출하는 데 활용된다. 문서 점수가 높은 문서 또는 해당 문서가 기술하고 있는 대상이 검색 결과로 반환된다.

[0096] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체 및 통신 매체를 모두 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다. 통신 매체는 전형적으로 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈, 또는 반송파와 같은 변조된 데이터 신호의 기타 데이터, 또는 기타 전송 매커니즘을 포함하며, 임의의 정보 전달 매체를 포함한다.

[0097] 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. 도 1은 본 발명의 일 실시예의 하나 이상의 구성 요소 또는 동작을 실시하기 위하여 사용될 수 있는 컴퓨터 시스템 아키텍처의 일례를 도시한 도면이다. 하드웨어 시스템(800)은 프로세서(810), 캐쉬(815), 메모리(815) 및 상술한 기능에 관련된 하나 이상의 소프트웨어 어플리케이션 및 드라이버를 포함할 수 있다.

[0098] 추가적으로, 하드웨어 시스템(800)은 고성능 입/출력(I/O) 버스(840) 및 표준 I/O 버스(870)를 포함한다. 호스트 브릿지(820)는 프로세서(810)를 고성능 I/O 버스(840)에 결합시키고, I/O 버스 브릿지(850)는 두 개의 버스(840 및 870)를 상호 결합시킨다. 시스템 메모리(860) 및 네트워크/통신 인터페이스(830)는 고성능 I/O 버스(840)에 결합된다. 하드웨어 시스템(800)은 비디오 메모리(도시 생략) 및 그 비디오 메모리에 결합된 디스플레이 장치를 더 포함할 수 있다. 대용량 기억장치(830) 및 I/O 포트(890)는 표준 I/O 버스(870)에 결합된다. 하드웨어 시스템(800)은 키보드 및 포인팅 디바이스(pointing device), 그리고 표준 I/O 버스(870)에 결합된 디스플레이 장치(도시 생략)를 선택적으로 포함할 수 있다. 총괄적으로, 이러한 요소들은 넓은 범위의 컴퓨터 하드웨어 시스템을 표현하도록 의도되며, 인텔사에 의해 제조된 펜티엄 프로세서뿐만 아니라 다른 적절한 프로세서에 기반하는 범용 컴퓨터 시스템을 포함하나 이에 제한되지는 않는다.

[0099] 하드웨어 시스템(800)의 구성 요소들은 아래에서 더 상세하게 설명된다. 보다 구체적으로, 네트워크 인터페이스(830)는 하드웨어 시스템과 이더넷(예를 들어, IEEE 802.3) 네트워크 등과 같은 임의의 넓은 범위의 네트워크 사이의 통신을 제공한다. 본 발명의 일 실시예의 경우, 네트워크 인터페이스(830)는 하드웨어 시스템(800)과 네트워크 사이를 접속하여 하드웨어 시스템(800)이 그것들의 데이터베이스들을 관리하도록 한다. 대용량 기억장치(830)는 데이터와 프로그래밍 명령을 위한 영구 기억장치를 제공하여 본 발명의 일 실시예에서 구현되는 상술한 기능을 수행하고, 시스템 메모리(860)(예를 들어, DRAM)는 프로세서(810)에 의하여 수행될 때 데이터 및 프로그래밍 명령을 위한 임시 기억장치를 제공한다. I/O 포트(890)는 하드웨어 시스템(800)에 결합될 수 있는 추가적인 주변 장치 간의 통신을 제공하는 하나 이상의 직렬 및/또는 병렬 통신 포트이다.

[0100] 하드웨어 시스템(800)은 다양한 종류의 시스템 아키텍처를 포함할 수 있고, 하드웨어 시스템(800)의 다양한 구성요소가 재배치될 수 있다. 예를 들어, 캐쉬(815)는 프로세서(810)에 내장될 수 있다. 선택적으로, 캐쉬(815) 및 프로세서(810)는 "프로세서 모듈"로써 함께 묶여질 수 있고, 이 때 프로세서(810)는 "프로세서 코어"로서 지칭될 수 있다. 또한, 본 발명의 특정한 실시예는 상술한 구성 요소의 전부를 요구하거나 포함하지 않을 수 있다. 예를 들어, 표준 I/O 버스(870)에 결합되는 것으로 도시된 주변 장치들은 고성능 I/O 버스(840)에 결합될 수 있다. 추가적으로, 임의의 실시예에서 단지 하나의 버스만 존재할 수 있고, 하드웨어 시스템(800)의 구성 요소들은 그 하나의 버스에 결합될 수 있다. 더욱이, 하드웨어 시스템(800)은 추가적인 프로세서, 기억 장치 또는 메모리와 같은 추가적인 구성요소를 포함할 수 있다. 이하에서 논의되는 것처럼, 본 발명의 일 실시예의 동작은

하드웨어 시스템(800)에 의해 구동되는 일련의 소프트웨어 루틴으로서 실시될 수 있다. 이러한 소프트웨어 루틴들은 프로세서(810)와 같은 하드웨어 시스템에서 프로세서에 의하여 실행될 수 있는 복수의 또는 일련의 명령들을 포함한다. 우선, 일련의 명령들은 대용량 기억 장치(830)와 같은 기억 장치에 저장된다. 그러나 일련의 명령들은 디스켓, CD-ROM, ROM, EEPROM 등과 같은 임의의 적절한 기억 매체에 저장될 수 있다. 더욱이, 일련의 명령들은 국소적으로 저장될 필요가 없고, 네트워크/통신 인터페이스(830)를 통하여 네트워크 상의 서버 등과 같은 원격 기억 장치로부터 수신될 수 있다. 그 명령들은 대용량 기억 장치(830)와 같은 기억 장치로부터 시스템 메모리(860)로 복사되고, 프로세서(810)에 의하여 액세스되고 실행된다.

[0101] 운영 시스템은, 소프트웨어 어플리케이션(도시 생략)과의 데이터 입/출력을 포함하는 하드웨어 시스템(800)의 동작을 관리하고 제어한다. 운영 시스템은 시스템 상에서 실행되는 소프트웨어 어플리케이션과 시스템의 하드웨어 구성 요소 사이의 인터페이스를 제공한다. 본 발명의 일 실시예에 따른 운영 시스템은 마이크로소프트사의 Windows 95/98/NT/XP/VISTA 운영 시스템일 수 있다. 그러나, 본 발명은 애플 컴퓨터사의 애플 맥킨토시 운영 시스템, 유닉스(UNIX) 운영 시스템, 리눅스(LINUX) 운영 시스템 등과 같은 다른 적절한 운영 시스템에서도 사용될 수 있다.

[0102] 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.

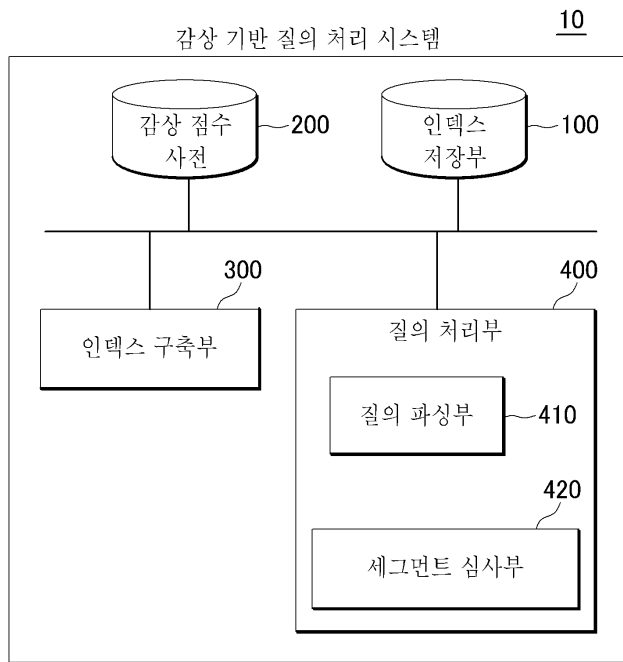
[0103] 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

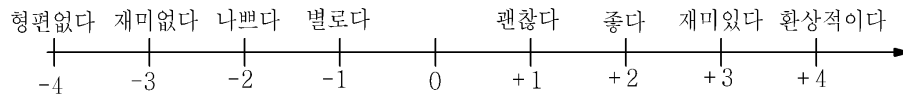
[0104] 10: 감상 기반 질의 처리 시스템
100: 인덱스 저장부
200: 감상 점수 사전
300: 인덱스 구축부
400: 질의 처리부
410: 질의 파싱부
420: 세그먼트 심사부

도면

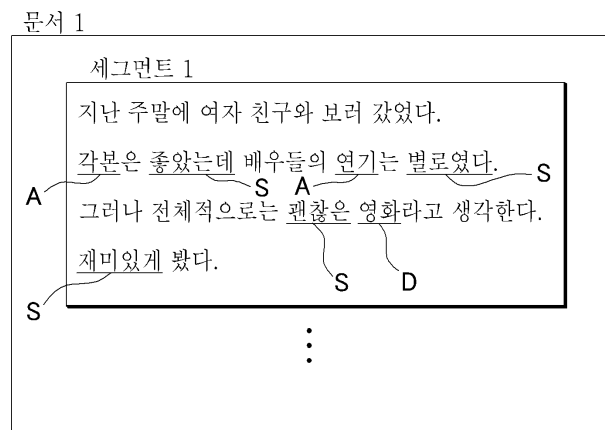
도면1



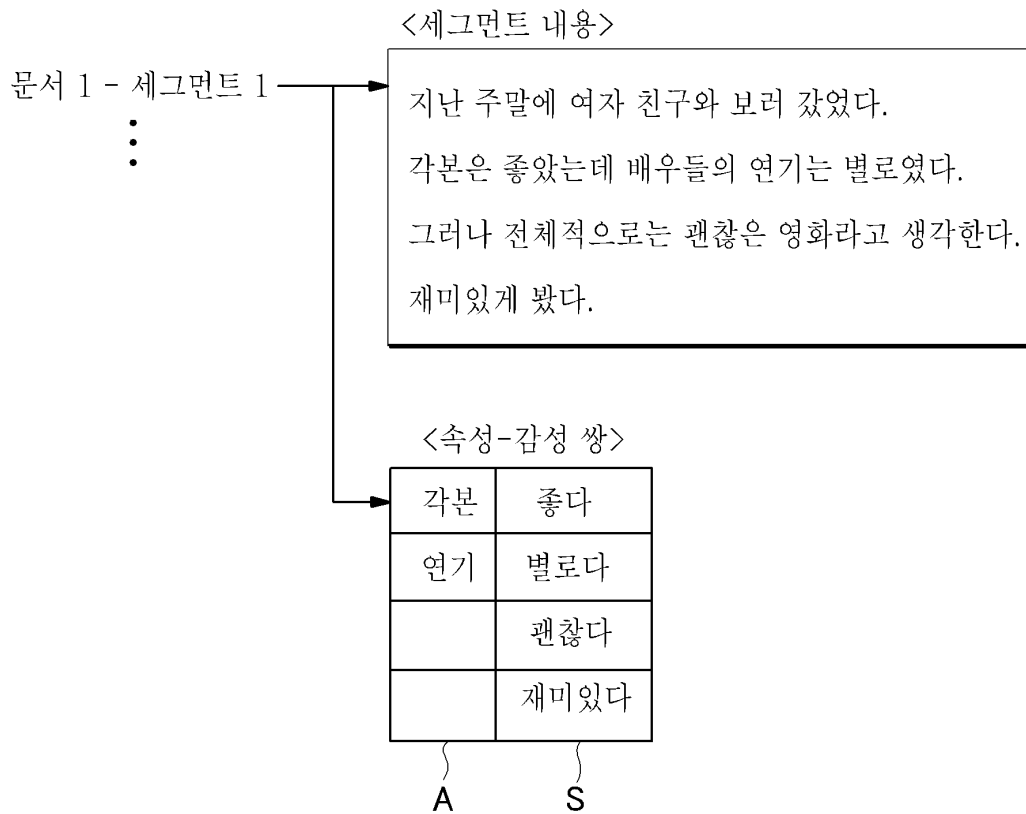
도면2



도면3



도면4



도면5

사용자 질의: 연기가 좋은 영화
A S D
파싱된 질의: 연기 (극성부호: +)

도면6

사용자 질의: 각본이 형편없다고 평가받은 영화
A S D
파싱된 질의: 각본 (극성부호: -)

도면7

사용자 질의: 여자 친구와 보기 좋은 영화

S D

파싱된 질의: 여자 친구, 보다 (극성부호: +)

도면8

질의: 연기가 좋은 영화

검색된 세그먼트: 세그먼트 1

검색된 속성-감상 쌍: 연기-별로다 (점수=-1)

속성-감상 쌍 점수 합= -1

세그먼트 점수= -1 x (+1) = -1

결론: 세그먼트 1은 검색 결과로 반환되지 않음.

도면9

질의: 각본이 형편없다고 평가받은 영화

검색된 세그먼트: 세그먼트 1

검색된 속성-감상 쌍: 각본-좋다 (점수=+2)

속성-감상 쌍 점수 합= +2

세그먼트 점수= +2 x (-1) = -2

결론: 세그먼트 1은 검색 결과로 반환되지 않음.

도면10

질의: 여자 친구와 보기 좋은 영화

검색된 세그먼트: 세그먼트 1

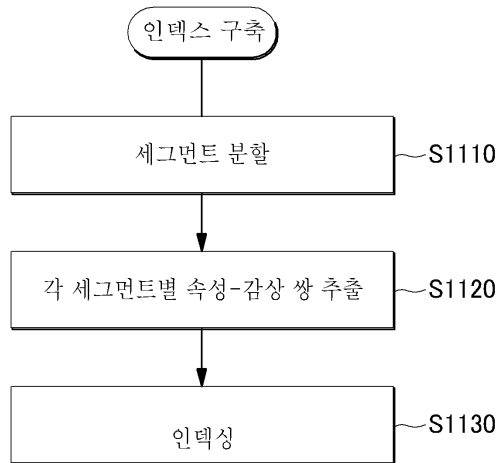
검색된 속성-감상 쌍: 없음

속성-감상 쌍 점수 합= $\frac{(+2) + (-1) + (+1) + (+3)}{4} = +1.25$

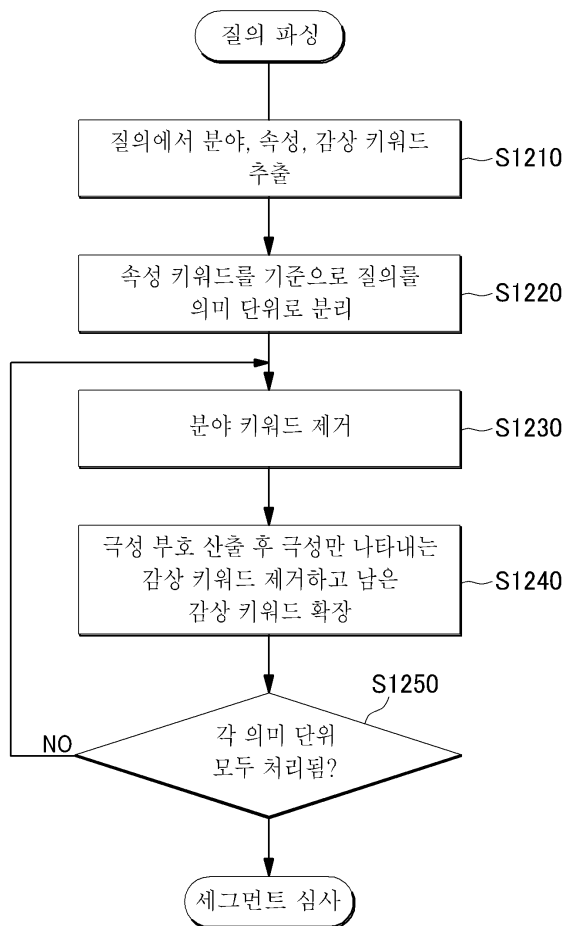
세그먼트 점수= +1.25 x (+1) = +1.25

결론: 세그먼트 1은 검색 결과로 반환될 가능성이 있음.

도면11



도면12



도면13

