



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2012년03월07일
(11) 등록번호 10-1117339
(24) 등록일자 2012년02월09일

(51) 국제특허분류(Int. Cl.)
G06F 17/30 (2006.01)
(21) 출원번호 10-2008-7009060(분할)
(22) 출원일자(국제) 2004년09월30일
심사청구일자 2009년06월26일
(85) 번역문제출일자 2008년04월16일
(65) 공개번호 10-2008-0040055
(43) 공개일자 2008년05월07일
(62) 원출원 특허 10-2006-7006205
원출원일자(국제) 2004년09월30일
(86) 국제출원번호 PCT/US2004/032258
(87) 국제공개번호 WO 2005/033890
국제공개일자 2005년04월14일
(30) 우선권주장
10/675,234 2003년09월30일 미국(US)
(56) 선행기술조사문헌
JP2002032408 A
JP2002063206 A
전체 청구항 수 : 총 26 항

(73) 특허권자
야후! 인크.
미국, 94089 캘리포니아, 쉐니배일, 퍼스트 애브뉴 701
(72) 발명자
왕, 쑤예준
미국 95120 캘리포니아주 산 호세 마운틴 콰일 서클 1164
액튼, 브라이언
미국 95070 캘리포니아주 사라토가 크라이사이드 레인 12452
판카파케산, 벤카트
미국 94560 캘리포니아주 네워크 넘버235 자비스 애비뉴 6167
(74) 대리인
백만기, 이중희, 주성민

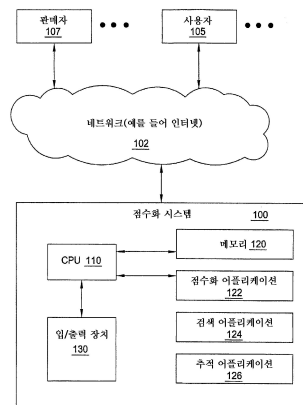
심사관 : 석상문

(54) 발명의 명칭 검색을 점수화하기 위한 방법 및 장치

(57) 요약

더 높은 관련성을 갖는 검색 결과를 생성하기 위한 방법과 장치가 제공된다. 본 발명은 각각의 주어진 검색어에 대한 사용자의 선택은 몇몇 판매자로부터의 몇몇 상품에 집중되며, 모든 결과는 검색어와 상당히 관련이 있다는 사실을 이용한다. 일 실시예에서, 이러한 결과가 각각의 검색어에 대한 판매자의 순위를 결정하기 위해 사용된다. 특히 구매 및/또는 클릭 정보로부터 사용자의 선택을 습득함으로써, 높은 관련성이 있고 가장 인기있는 상품이 텍스트 관련성만 있는 상품보다 더 높은 점수 또는 순위를 할당받을 수 있다.

대표도 - 도1



특허청구의 범위

청구항 1

검색 엔진 방법으로서,

검색에 응답하여 생성되며 복수의 멤버(member)들을 가지는 검색 결과 집합 중, 하나의 멤버와 관련된 클릭 정보를 수집하는 단계;

상기 클릭 정보, 및 상기 검색 결과 집합의 표시를 위한 순위(ordering) 내에서의 상기 멤버의 위치에 기초하여, 상기 멤버의 상대적 사용자 반응성(relative user responsiveness)을 결정하는 단계;

상기 검색에 대한 상기 멤버의 상기 상대적 사용자 반응성을 반영하는, 상기 멤버와 관련된 점수를 생성하는 단계; 및

후속 검색에 대한 응답에 영향을 미치도록 상기 점수를 사용하는 단계를 포함하는, 검색 엔진 방법.

청구항 2

제1항에 있어서,

상기 점수는 상기 검색 결과 집합의 상기 순위 내에서 상기 멤버의 배치에 영향을 주도록 적용되는 검색 엔진 방법.

청구항 3

제1항에 있어서, 상기 점수를 생성하는 단계 이후에

시간의 경과를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는 검색 엔진 방법.

청구항 4

제1항에 있어서, 상기 점수를 생성하는 단계 이후에

상기 멤버의 상세 정보를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는 검색 엔진 방법.

청구항 5

제1항에 있어서, 상기 점수를 생성하는 단계 이후에

상기 검색에 사용되는 검색어의 상세 정보를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는 검색 엔진 방법.

청구항 6

제1항에 있어서, 상기 점수를 사용하는 단계는

상기 점수를 텍스트 관련도 점수, 유료 검색 등록 점수, 또는 유료 스폰서 등록 점수와 공동으로 적용하는 단계를 포함하는

검색 엔진 방법.

청구항 7

제1항에 있어서,

상기 멤버는 상품 페이지인

검색 엔진 방법.

청구항 8

제1항에 있어서,

상기 멤버는 카탈로그 페이지인

검색 엔진 방법.

청구항 9

제8항에 있어서,

상기 카탈로그 페이지는 상품을 제공하는 다수의 판매자를 표시하는, 상기 상품에 대한 디스플레이 페이지를 나타내는

검색 엔진 방법.

청구항 10

제9항에 있어서,

상기 카탈로그 페이지는 상기 다수의 판매자에 의한 상기 상품에 대한 가격 정보도 함께 표시하는

검색 엔진 방법.

청구항 11

제1항에 있어서,

상기 생성하는 단계는 상기 클릭 정보의 클릭 유형에 기초하여 상기 멤버의 점수를 생성하는

검색 엔진 방법.

청구항 12

제1항에 있어서,

상기 클릭 정보는 선호 판매자에 의해 이루어진 클릭을 나타내는 선호 판매자 클릭 유형으로 카테고리화될 수 있는 클릭을 포함하는

검색 엔진 방법.

청구항 13

제1항에 있어서,

상기 클릭 정보는 카탈로그 페이지에 관련된 상품 페이지 상의 클릭에 대한 맵핑된 카탈로그 클릭 유형의 클릭을 포함하는

검색 엔진 방법.

청구항 14

제1항에 있어서, 상기 점수를 생성하는 단계 이후에

상기 점수에 사용 함수를 적용하는 단계를 더 포함하는

검색 엔진 방법.

청구항 15

제1항에 있어서, 상기 점수를 생성하는 단계 이전에

상기 점수를 생성하기 위한 설정(configuration)을 선택하는 단계를 더 포함하는

검색 엔진 방법.

청구항 16

제15항에 있어서, 상기 설정을 선택하는 단계 이후에

상기 설정에서 정의된 변수(parameter)를 계산하는 단계를 더 포함하며,

상기 점수는 상기 설정 변수와 클릭 유형에 따라 생성되는

검색 엔진 방법.

청구항 17

제1항에 있어서, 상기 점수는 다음에 따라 생성되며,

$$Hotscore_{k,p} = \sum (\alpha_{k,t,T(t)} C_{k,p,t})$$

$C_{k,p,t}$ 는 멤버 p를 리턴한 검색에 사용된 검색어 k에 대한 적어도 하나의 클릭 유형 t 내의 클릭 횟수이며,

$\alpha_{k,t,T(t)}$ 는 설정 변수인

검색 엔진 방법.

청구항 18

제1항에 있어서, 상기 점수를 생성하는 단계는

각각 서로 다른 점수화 전략을 강조하는 다수의 공식으로부터, 하나의 점수화 전략을 강조하는 공식을 선택하는 단계를 포함하는

검색 엔진 방법.

청구항 19

컴퓨터 판독 가능한 매체로서, 컴퓨터가

검색에 응답하여 생성되며 복수의 멤버(member)들을 가지는 검색 결과 집합 중, 하나의 멤버와 관련된 클릭 정보를 수집하는 단계;

상기 클릭 정보, 및 상기 검색 결과 집합의 표시를 위한 순위(ordering) 내에서의 상기 멤버의 위치에 기초하여, 상기 멤버의 상대적 사용자 반응성(relative user responsiveness)을 결정하는 단계;

상기 검색에 대한 상기 멤버의 상기 상대적 사용자 반응성을 반영하는, 상기 멤버와 관련된 점수를 생성하는 단계; 및

후속 검색에 대한 응답에 영향을 미치도록 상기 점수를 사용하는 단계

를 포함하는 방법을 수행하도록 하기 위한 프로그램을 저장한, 컴퓨터 판독 가능한 매체.

청구항 20

제19항에 있어서, 상기 점수를 생성하는 단계는

각각 서로 다른 점수화 전략을 강조하는 다수의 공식으로부터, 하나의 점수화 전략을 강조하는 공식을 선택하는 단계를 포함하는

컴퓨터 판독 가능한 매체.

청구항 21

제19항에 있어서, 상기 방법은 상기 점수를 생성하는 단계 이전에

상기 점수를 생성하기 위한 설정(configuration)을 선택하는 단계를 더 포함하는

컴퓨터 판독 가능한 매체.

청구항 22

제19항에 있어서,

상기 점수는 상기 검색 결과 집합의 상기 순위 내에서 상기 멤버의 배치에 영향을 주도록 적용되는

컴퓨터 판독 가능한 매체.

청구항 23

제19항에 있어서, 상기 방법은 상기 점수를 생성하는 단계 이후에

시간의 경과를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는

컴퓨터 판독 가능한 매체.

청구항 24

제19항에 있어서, 상기 방법은 상기 점수를 생성하는 단계 이후에

상기 멤버의 상세 정보를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는

컴퓨터 판독 가능한 매체.

청구항 25

제19항에 있어서, 상기 방법은 상기 점수를 생성하는 단계 이후에

상기 검색에 사용되는 검색어의 상세 정보를 고려하도록 상기 점수를 조정하는 단계를 더 포함하는

컴퓨터 판독 가능한 매체.

청구항 26

제19항에 있어서, 상기 점수를 사용하는 단계는

상기 점수를 텍스트 관련도 점수, 유료 검색 등록 점수, 또는 유료 스폰서 등록 점수와 공동으로 적용하는 단계를 포함하는

컴퓨터 판독 가능한 매체.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 검색 결과를 점수화하고 순위화하는 방법 및 장치에 관한 것이다. 더욱 상세하게는, 본 발명은 상거래(transaction) 및/또는 클릭 이력에 기반한 점수화 접근 방법에 관한 것이다.

배경 기술

[0002] 인터넷 상에서 정보가 광대한 양으로 확산됨에 따라, 관계없는 상당수의 검색 결과를 먼저 읽는데에 상당한 시간을 소비해야 할 필요 없이 관련된 정보를 검색하여 나열하기는 상당히 어렵다. 검색되는 자료에 따라, 사용자는 종종 관련없는 수많은 검색 결과를 열람해야 하는 낭패를 겪는다.

[0003] 점수화 또는 순위화는 검색, 예컨대 특히 쇼핑/상품 검색에 있어 핵심적인 문제들 중 하나이다. 검색이 검색 결과 리스트의 상위 근처에 가장 관련 있는 문서를 제공할 수 없다면, 그 검색은 관련도가 낮다고 인식되곤 한다. 사용자의 목적은 관련된 하나의 결과만을 찾는 것이 아니기 때문에, 사용자는 통상의 웹 검색보다 쇼핑/상품 검색과 같은 검색에 더 높은 관련도를 요구하는 경향이 있다. 사용자는 종종 가장 관련 있는 상품을 보고 다른 상품 및 다른 판매자들과 비교할 수 있기를 원한다.

[0004] 단순 텍스트(text) 관련성에 기반한 점수화는 몇몇 검색 기술의 기초이다. 기본적인 아이디어는 문서의 제목, 설명 및 기타 영역에서 일치하는 텍스트를 찾는 것이다. 예컨대 제목과 같은 몇몇 영역에 더 높은 가중치를 부여하거나, 문구가 일치하는 경우에 더 높은 가중치를 부여하는 등의 추가적인 개선 작업이 더해질 수 있다. 그러나 이러한 단순 텍스트 관련 점수화의 접근 방법은 사용자가 정확히 무엇을 검색하고자 하는지 결정할 수 없기 때문에 가장 관련 있는 검색 결과를 생성하는 데에 문제가 있다.

[0005] 예컨대 단순 텍스트 관련 검색에 있어서, "컴퓨터(computer)"라는 단어를 검색할 때, "소니 바이오 FX340(Sony VAIO FX340)"과 같은 제목을 가진 문서는 제목에 "컴퓨터"라는 단어가 없기 때문에 텍스트 일치도가 좋은 것으로 인식되지 않을 것이며, "컴퓨터 케이스(computer case)"와 같은 제목을 가진 문서가 텍스트 일치도가 좋은 것으로 인식될 것이다. 이러한 예는 컴퓨터에 대한 검색이 많은 관계 없는 아이템이 포함된 검색 결과를 생성할 가능성이 높을 것임을 입증한다.

[0006] 모든 검색 결과가 관련이 있다고 인식될 때에도, 더 인기도가 높은 상품에 더 높은 점수 또는 순위를 부여하는 것이 여전히 바람직하다. 그러나 단순 텍스트 관련 검색은 이러한 중요한 개선 작업을 제공할 수 없을 것이다.

발명의 내용

해결 하고자하는 과제

[0007] 따라서 검색 결과에 더 높은 관련성을 부여하는 방법 및 장치에 대한 기술이 필요하다.

과제 해결수단

[0008] 일 실시예에서, 본 발명은 더 높은 관련성이 있는 검색 결과를 생성하는 방법 및 장치를 제공한다. 예컨대, 본 발명은 쇼핑/상품 검색에 대한 더 높은 관련성이 있는 검색 결과를 생성하기 위한 방법 및 장치를 제공한다.

[0009] 본 발명의 한가지 전제는, 사용자는 선호하는 상품을 구매 및/또는 클릭함으로써 인기 검색어에 대한 선호 상품에 관하여 사용자의 기호를 드러낸다는 점이다. 사용자가 쇼핑/상품 검색 사이트에서 어떤 단어를 검색할 때, 사이트가 많은 관계 없는 검색 결과를 보여준다 할지라도, 많은 사용자들은 그들이 관심 있는 결과, 즉 관련 있는 결과를 간단하게 선택함으로써 관계 없는 결과들을 걸러낼 것이다. 이는 사용자가 실제로 검색 결과 목록으로부터 상품을 구매할 때 특히 정확하며, 따라서 검색어에 대한 결과의 관련성 뿐만 아니라 구매한 상품의 가격 관련성 및/또는 구매한 상품을 판매하는 판매자의 관련성을 나타낸다.

효 과

[0010] 본 발명은 각각의 주어진 검색어에서 사용자의 선택이 몇몇 판매자로부터의 몇몇 상품들에 집중되는 경향이 있으며, 모든 결과가 그 검색어와 상당히 관련 있다는 사실을 이용한다. 일 실시예에서, 이러한 결과는 각각의 검색어에 대한 판매자의 순위를 결정하기 위해 사용된다. 특히 구매 및/또는 클릭 정보로부터 사용자의 선택을 습득함으로써, 높은 관련성이 있고 가장 인기있는 상품이, 텍스트만 관련 있는 상품보다 더 높은 점수 또는 순위를 할당받을 수 있다.

발명의 실시를 위한 구체적인 내용

[0011] 도 1은 본 발명의 점수화 시스템(100)을 도시한 블록도이다. 점수화 시스템(100)은 검색어에 관련하여 생성된 검색 결과 내에서, 예컨대 상품과 같은 문서의 점수화를 수행한다.

[0012] 더욱 상세하게는, 도 1은 다수의 사용자(105)가 검색을 수행할 수 있는 네트워크, 예컨대 인터넷(102)과 연결된 점수화 시스템(100)을 도시한다. 검색은 주로 하나 이상의 검색어, 예컨대 "랩탑 컴퓨터(laptop computer)", "디브이디(DVD)", "가스 그릴(gar grill)" 등등을 입력할 사용자에게 의해 시작된다. 검색은 사용자가 필요로 하는 상품 및 서비스에 대한 검색을 포함할 수 있다. 상품 및 서비스는 점수화 시스템(100)을 유지하는 실체, 예컨대 다량의 상품 및 서비스를 제공하는 기업, 예컨대 월마트(Walmart) 등에 의해 제공될 수 있다. 이외에 상품 및 서비스는 다수의 판매자(107)에 의해 제공될 수 있으며, 여기서 점수화 시스템(100)은 제3자에 의해 설치되고 사용자에게 의해 제공되는 검색어에 관한 검색 결과 생성, 예컨대 검색 엔진 어플리케이션만을 수행한다. 요컨대, 본 발명의 점수화 시스템(100)은 설치되는 방법에 있어서 한정되지 않는다.

[0013] 일 실시예에서, 점수화 시스템(100)은 범용 컴퓨터 또는 기타 하드웨어 장치를 사용하여 실시된다. 더 상세하

게는, 점수화 시스템(100)은 프로세서(CPU; 110), 예컨대 임의 접근 기억장치(random access memory; RAM) 및/또는 판독 전용 기억장치(read only memory; ROM)와 같은 메모리(120), 점수화 엔진 또는 어플리케이션(122), 검색 엔진 또는 어플리케이션(124), 추적(tracking) 엔진 또는 어플리케이션(126) 및 다양한 입/출력 장치(130)(예컨대 테이프 드라이브, 플로피 드라이브, 하드디스크 드라이브 또는 콤팩트 디스크 드라이브, 수신기(receiver), 전송기(transmitter), 스피커, 디스플레이, 출력 포트, 사용자 입력 장치(키보드, 키패드, 마우스 등), 또는 음성 명령어를 포착하기 위한 마이크로폰을 포함하되 이에 한정되지 않는 저장 장치)를 포함한다.

[0014] 점수화 엔진 또는 어플리케이션(122), 검색 엔진 또는 어플리케이션(124), 및 추적 엔진 또는 어플리케이션(126)은 통신 채널을 통해 CPU(110)에 연결되는 물리적인 장치 또는 시스템으로서 실시될 수 있다. 이외에, 점수화 엔진 또는 어플리케이션(122), 검색 엔진 또는 어플리케이션(124), 그리고 추적 엔진 또는 어플리케이션(126)은 하나 이상의 소프트웨어 어플리케이션(또는 소프트웨어와 하드웨어의 조합, 예컨대 주문형 반도체(application specific integrated circuit; ASIC))을 의미할 수 있으며, 소프트웨어는 저장 매체(예컨대 자기 또는 광학 드라이브 또는 디스켓)로부터 불러와서 컴퓨터의 메모리(120)에서 CPU에 의해 작동된다. 이와 같이 본 발명의 점수화 엔진 또는 어플리케이션(122), 검색 엔진 또는 어플리케이션(124), 및 추적 엔진 또는 어플리케이션(126)(관련 데이터 구조를 포함함)은 컴퓨터 판독가능한 기록매체, 예컨대 램(RAM) 메모리, 자기 또는 광학 드라이브 또는 디스켓 등에 저장될 수 있다.

[0015] 요컨대, 점수화 시스템은 검색 관련도의 극대화를 이루기 위해 설계된다. 본 발명은 사용자가 그들이 선호하는 상품을 구매 또는 클릭함으로써 인기 검색어에 대한 선호 상품에 관한 사용자의 기호를 드러낸다는 사실을 이용한다. 사용자가 쇼핑/상품 검색 사이트에서 어떠한 단어를 검색할 때, 사이트는 종종 다수의 관련 없는 결과를, 그것도 검색 결과의 상위에 표시할 것이다. 사용자는 종종 잘못된 결과를 간단히 걸러내고, 사용자가 관심있는 결과, 즉 관련 있는 결과만을 선택할 것이다. 검색 결과의 관련성은 사용자가 실제로 검색 결과로부터 선택된 상품을 구매할 때 현저히 입증된다. 즉 사용자가 상품을 구매하기로 결정하면, 사용자가 선택한 상품은 상품 가격 및/또는 상품을 판매하는 판매자와 관련하여 검색어와 상당히 관련 있음이 틀림없다.

[0016] 추적 데이터의 양이 충분히 크다면, 각각의 주어진 검색어에 대한 사용자의 선택은 몇몇 판매자로부터의 몇몇 상품에 집중되는 경향이 있으며, 모든 결과는 검색어와 상당히 관련 있음이 인정되어왔다. 특히 구매 및/또는 클릭으로부터 사용자의 선택을 습득하고 적용함으로써, 관련성이 높은 상품이 텍스트 관련성만 있는 상품보다 더 높은 점수/순위를 부여받을 수 있다. 이러한 새로운 접근 방법은 검색어에 대해 관련성이 높은 검색 결과를 생성할 것이다. 실제로도, 개선 또는 표준화가 추가적으로 적용될 수 있는데, 예컨대 각각의 검색어에 대한 판매자의 순위를 매길 수 있다. 이러한 추가적인 조정은 아래에서 더 설명된다.

[0017] 본 발명의 일 실시예에서, 사용자 구매 및/또는 클릭 정보에 기반하여 검색어에 대해 상품에 할당된 점수는 "최신점수(hotscore)"라 불린다. 본 발명의 최신점수는 검색어에 대한 검색 결과를 생성하는 검색 엔진에 의해 사용될 수 있다. 이러한 최신점수는 검색 결과를 생성함에 있어서 우선 순위(가중치를 더 부여받는) 변수로서 사용될 수 있으며, 또는 일반적으로 유료 검색 등록(paid inclusion), 유료 스폰서 등록(paid sponsorship), 텍스트 관련성과 같은 다른 변수들을 우선 순위 변수로 사용하는 검색 엔진을 보충하기 위해 사용됨을 주목해야 한다.

[0018] 도 2는 더 높은 관련성이 있는 검색 결과의 문서 목록을 이루기 위한 본 발명의 점수화 방법을 적용하는 관계를 도시한다. 도 2는 특정 검색어에 대해 생성되고 사용자에게 제공되는 제1 결과 집합(220)을 도시한다. 상기 예에서, 검색 결과 설정 내의 아이템은 넓게는 문서로 정의되며, 쇼핑의 경우에 있어서, 문서는 상품 또는 상품-판매자 쌍일 것이다. 그러나 문서는 넓게는 웹사이트, 텍스트 문서, 이미지 등을 포함한다.

[0019] 도 2는 제1 검색 결과 집합 내의 다양한 문서들에 대한 구매 및/또는 클릭(210)을 추적하여 제1 결과 집합(220)에 대한 사용자 응답의 추적을 도시한다. 이러한 구매 및/또는 클릭 정보가 추적되면 문서들 중 하나에 관한 각각의 점수가 있는 다수의 점수(최신점수)(240)를 생성하는 점수화 처리(230)에 사용된다. 그 다음, 제1 결과 집합을 생성한 동일한 검색어에 대해 제2 검색 결과 집합(260)을 생성하기 위해, 최신점수(240)는 텍스트 점수(252) 및 기타 점수들(254)(예컨대 유료 검색 등록 점수)과 함께 최신점수를 적용할 수 있는 또다른 점수화 시스템(250)에 의해 선택적으로 사용된다. 도 2는 최신점수의 적용이 현재 문서의 순위에 영향을 미쳐왔으며, 어찌면 제2 결과 집합 내에 문서의 추가 또는 삭제에도 영향을 미칠 수 있고, 따라서 제2 검색 결과 집합에 더 우수한 관련성을 제공함을 도시한다.

[0020] 일 실시예에서, 각각의 검색어에 대해, 본 발명은 사용자가 클릭하고 최종적으로 구매한 판매자/상품-식별자를 추적한다. 클릭/구매가 발생할 때 검색 결과 내의 상품의 위치, 상기 클릭/구매가 일어난 시간, 및 상기 클릭/

구매가 발생할 때 상품이 할당된 매장을 포함하는 더 상세한 정보가 추적될 수도 있다.

- [0021] 도 3은 다수의 상품에 대한 최신점수를 생성하기 위한 예시적인 방법(300)의 흐름도를 도시한다. 방법(300)은 단계(305)에서 시작하여 단계(310)로 진행된다.
- [0022] 단계(310)에서, 방법(300)은 특정의 검색어에 따른 각각의 상품에 대한 판매 및/또는 클릭 데이터를 전처리한다. 예컨대 본 발명은 각각의 튜플(tuple)<k,t,p>에 대한 데이터를 생성하며, k는 검색어, p는 상품, t는 유형이다. 즉, 방법(300)은 $C_{k,p,t}$ 를 생성할 것이며, 이는 검색어 k에 대해 일정 기간(time period) "tp" 동안 발생한 유형 t의 이벤트의 수 또는 횟수이다. 유형 t의 이벤트는 구매 이벤트 및 또는 클릭 이벤트(예컨대 선호하는 판매자로부터 상품을 구매하거나 검색 결과 상의 문서를 클릭한 경우)의 특정 유형을 정의할 수 있다. 다수의 예시적인 유형의 이벤트가 아래에 소개된다.
- [0023] 특히 설정 화일에서 정의되고 조정될 수 있는 주어진 시간 범위 동안, 각각의 검색어에 대한 모든 판매자/상품-식별자가 $C_{k,p,t}$ 에 기반하여 여러 유형으로 카테고리화 되고 카운트된다. 추가적으로 신뢰도가 낮은 결과는 소멸된다. 신뢰도가 낮은 결과에는 스팸성(spamming) 결과 및 산발적인(scattered) 결과가 포함될 수 있다. 산발적인 결과는, 예컨대 우연히 접근되었고 실질적으로 관련 링크를 가르키지 않는 링크와 같이, 주어진 시작지점 하에서 반복되는 결과이다.
- [0024] 단계(320)에서, 방법(300)은 선택적으로 시간 및/또는 위치를 나타내는 데이터를 표준화한다. 특히 상품이 검색 결과 집합에서 "더 상위"에 위치할수록, 상품이 사용자에게 의해 클릭/구매되는 확률이 더 높아지는 것으로 관찰되어왔다. 더 상세하게는 클릭수는 위치에 의해 큰 영향을 받으며(예컨대 더 상위에 위치한 상품이 자주 클릭된다), 반면에 구매는 영향을 적게 받는다는 점(예컨대 구매자는 관련 상품의 위치에는 영향을 적게 받는다)도 관찰된다. 따라서 사용자는 더 상위에 위치한 상품을 클릭할 수 있지만 결국엔 관련성으로 인해 훨씬 하위에 위치한 상품을 구매할 수 있다.
- [0025] 검색 결과 집합의 제1 상위는 검색 결과 집합 내에서 최상위에 위치하는 것으로 간주된다. 더 관련된 결과를 찾기 위해, 판매자/상품-식별자의 신뢰도가 클릭/구매가 일어날 때의 위치에 기반하여 표준화될 수 있다. 예컨대, 검색 결과 집합 내에 상당히 하위인 문서에서의 구매 또는 클릭은 상기 문서가 검색어에 관해 높은 관련성이 있음을 나타낼 것이다.
- [0026] 선택적으로, 시간("발생한 시간" 또는 "일어난 시간")을 나타내는 데이터가 표준화될 수 있다. 즉, 문서에서 판매 및/또는 클릭이 얼마나 최근에 일어났는지를 의미한다. 판매자/상품-식별자의 "발생 시간"이 상기 쌍의 관련성에 영향을 미치지 않는다는 할지라도, 시장에 있어 새로운 경향을 반영할 수도 있다. 이러한 경향을 알아내고 가장 인기있는 결과를 먼저 보여주는 것이 본 발명의 점수화 시스템의 목적 중 하나이다. 바꾸어 말하면, 상품의 인기도 또는 "시간 관련성"을 나타내는 순위로 관련 상품이 나열될 수 있다. 위치 및 시간의 표준화에 대한 다양한 표준화 함수가 사용될 수 있다.
- [0027] 단계(330)에서, 방법(300)은 설정 변수 α 를 계산한다. 더 상세하게는, 방법(300)은 각각의 <k,t> 쌍에 대한 $\alpha_{k,p,\text{MAX}}$ 와 $\alpha_{k,p,\text{MIN}}$ 을 계산한다. 설정 변수는 다양한 유형의 구매 및/또는 클릭의 효과를 정의하기 위해 사용된다. 예컨대, 스토어(store)(예컨대 선호되지 않는 소규모 판매자로 간주됨)를 통해 이루어진 구매는 카탈로그(선호되는 대규모 판매자로 간주됨)를 통해 이루어진 구매와 다르다. 마찬가지로, "선호되는 판매자"를 통해 이루어진 구매는 "일반적인 판매자"를 통해 이루어진 구매와 다르다. 이러한 차이는 구매 및 클릭 유형에 관한 정보가 아래에 개시될 바와 같이 검색 결과의 관련성을 더 개선하기 위해 사용될 수 있기 때문에 본 발명의 점수화 시스템의 운영자에게 중요하다.
- [0028] 단계(340)에서, 방법(300)은 구매 및/또는 클릭 정보에 기반하여 각각의 검색어에 대한 각각의 상품에 대한 점수(최신점수)를 생성한다. 상기 점수는 아래에 더 개시되는 다수의 다른 접근 방법에서 생성될 수 있다. 즉, 기업의 전략에 따라 다양한 공식들이 적용될 수 있다. 따라서 한 공식에서 계산된 판매자/상품-식별자에 대한 최신점수가 제2 공식에서 계산될 때에는 달라질 수 있다.
- [0029] 단계(350)에서, 방법(300)은 최신점수의 조정이 필요한지를 질의한다. 특히 조정은 예컨대, 검색어에 대한 특정 정보, 판매자-상품의 성과에 관한 정보, 구매자 행위 정보, 구매자 연령 정보, 구매자의 성별 정보 등과 같은 다른 정보를 나타내기 위해 선택적으로 적용될 수 있다. 상기 정보가 유효하다면, 최신점수가 그에 따라 조정될 수 있다.
- [0030] 예컨대, 인기 검색어에 기반하여 최신점수에 대한 조정이 이루어질 수 있다. 정보 기반으로 포함된 몇가지 인

기 검색어에 대해, 본 발명은 검색어에 대한 판매 정보를 추가할 수 있다. 예컨대 일 실시예에서, 검색어 "델(de11)"은 "제조자=델"로 해석될 수 있으며, 본 발명은 검색어 "델"에 대해서는 "제조자=델"이라는 판매 정보를 적용할 수 있다.

[0031] 이외에 관련 검색어에 대한 사용자 행위에 기반하여 최신점수에 대한 조정이 이루어질 수 있다. 관련 검색에 대한 사용자의 행위는, 포괄적인 검색어와 그에 관련된 더 협의의 검색어 간의 실제 관련성을 생성하는데 도움이 될 수 있다. 즉, 이것은 사용자가 포괄적인 검색어에 의한 검색을 좁혀갈 수 있도록 도울 것이다. 일 실시예에서, 본 발명은 포괄적인 검색어에 대한 판매자/상품에 대해 관련 검색어의 최신점수를 추가하며, 그에 따라 범위가 확장된다.

[0032] 이외에, 데이터가 판매자-상품 쌍의 매칭이 성과를 내지 못하고 있음을 나타낸다면 최신점수의 조정이 이루어질 수 있으며, 즉 부정확한 또는 선호되지 않는 문서에 대한 점수의 영향을 감소시키도록 최신점수를 조정할 수 있다. 예컨대 본 발명의 시스템은 최신점수가 판매자-상품 쌍에 할당된 후에 계속하여 결과를 평가한다. 성과를 잘 내지 못하는 쌍은 검색 결과 집합에서 잘못 선택된 문서 또는 선호되지 않는 문서로 간주되며, 최신점수가 감소될 것이다. 예컨대 검색 결과가 다수의 관련 문서(예컨대 검색어에 높은 관련이 있는 판매자-상품 쌍)를 제공할 수 있지만, 한가지 혹은 그밖의 이유로 인해 구매자는 특정 집합의 판매자-상품 쌍에 관심을 갖지 않는다. 이러한 경우, 관련성이 있지만 선호되지 않는 판매자-상품 쌍은 "처벌되며", 그 결과 최신점수는 더 낮아지거나, 심지어 음수가 될 것이다.

[0033] 단계(350)으로 돌아가면, 질의에 대해 부정의 응답을 받는다면, 방법(300)은 단계(375)에서 종료된다. 질의에 대해 긍정의 응답을 받는다면 방법(300)은 최신점수가 조정되는 단계(360)로 진행한다.

[0034] 단계(370)에서, 방법(300)은 최신점수에 대한 추가적인 조정이 필요한지 질의한다. 질의가 긍정의 응답을 받는다면, 방법(300)은 최신점수가 재조정되는 단계(360)로 진행한다. 질의가 부정의 응답을 받는다면, 방법(300)은 단계(375)에서 종료된다.

[0035] 일단 최신점수가 생성되면, 검색 엔진(124)은 즉시 최신점수를 적용하여 쇼핑/상품 검색을 수행한다. 일 실시예에서, 임의의 검색 방법에 기반하여 점수화된 검색은 즉시 본 발명의 최신점수로 조정된다. 예컨대, 사용자가 검색어를 타이핑하면, 쇼핑/상품 검색 시스템은 최신점수 증폭 비율과 함께 검색어를 검색 엔진에 보낼 것이다. 상기 비율은 아주 높을 수 있으며, 이는 최신점수가 있는 모든 상품이 최신점수가 없는 상품의 앞에 위치할 것임을 의미한다. 상기 비율은 또한 아주 낮을 수 있으며, 이는 최신점수가 검색 결과의 순위에 최소의 영향만을 미칠 것임을 의미한다.

[0036] 도 4는 판매 및 클릭 데이터를 전처리하는 방법(400)의 흐름도를 도시한다. 방법(400)은 단계(405)에서 시작하여 단계(410)로 진행한다.

[0037] 단계(410)에서, 방법(400)은 클릭 정보가 상품의 실제 구매를 포함하는지 질의한다. 질의가 긍정의 응답을 받으면 방법(400)은 원래의 클릭 정보가 사용되는 단계(492)로 진행한다. 즉, 상품의 구매는 검색 결과의 관련성의 관점에서 가장 높은 신뢰도를 부여한다. 따라서 구매에 관한 클릭 정보가 유지되며 사용된다. 질의가 부정의 응답을 받으면, 방법(400)은 단계(420)로 진행한다.

[0038] 단계(420)에서, 방법(400)은 클릭 정보가 기정의된 기준치보다 작은지 질의한다. 질의가 긍정의 응답을 받으면, 방법(400)은 단계(430)로 진행한다. 질의가 부정의 응답을 받으면, 방법(400)은 클릭 정보가 폐기되는 단계(494)로 진행한다. 즉, 단계(420)는 예컨대 검색 결과 내의 특정 문서에 대한 접근을 인위적으로 부풀리는 플러딩 공격(flooding attack)과 같은 부정 클릭 데이터를 제거하려는 것이다.

[0039] 단계(430)에서, 방법(400)은 클릭 정보가 신뢰성 있는 사이트로부터 온 것인지 질의한다. 질의가 긍정의 응답을 받으면, 방법(400)은 원래의 클릭 정보가 사용되는 단계(492)로 진행한다. 즉, 신뢰성 있는 사이트로부터 온 상품의 클릭 정보는 검색 결과의 관련성의 관점에 있어서 어느 정도의 신뢰도를 부여한다. 따라서 클릭 정보가 유지되며 사용된다. 질의가 부정의 응답을 받으면, 방법(400)은 단계(440)로 진행한다.

[0040] 단계(440)에서, 방법(400)은 특정 아이피(IP) 주소로부터 온 클릭 정보가 다른 IP 주소보다 많은지 질의한다. 바꿔 말하면, 통계적으로 특정 IP 주소에 관련된 클릭 정보가 다른 IP 주소로부터 온 클릭 정보에 비해 비정상적으로 많은지 질의한다. 질의가 긍정의 응답을 받으면, 방법(400)은 상기 특정 IP 주소로부터 온 클릭 정보를 폐기하는 단계(450)로 진행한다. 즉, 상기 특정 IP 주소로부터 온 클릭 정보는 의심되는 클릭 정보이다. 질의가 부정의 응답을 받으면, 방법(400)은 단계(460)로 진행한다.

[0041] 단계(460)에서, 방법(400)은 클릭과 페이지 조회수의 비율이 평균 비율보다 현저하게 높은지 질의한다. 질의가 긍정의 응답을 받으면, 방법(400)은 클릭 정보가 폐기되는 단계(470)로 진행한다. 즉, 클릭과 페이지 조회수의 비율 또는 빈도가 상당히 높다면, 다시 말해 사용자가 원래 클릭된 페이지를 조회하는데에 극히 짧은 시간을 보내면서 문서를 클릭한 다음 즉시 다른 문서를 클릭한다면, 클릭 정보는 의심되는 클릭 정보이다. 질의가 부정의 응답을 받으면, 방법(400)은 단계(480)로 진행한다.

[0042] 단계(480)에 있어서, 방법(400)은 검색 결과 집합 내의 문서의 클릭 횟수가 동일한 검색어의 동일한 검색 결과 집합 내의 다른 문서들의 클릭 횟수보다 현저하게 많은지를 질의한다. 예컨대 검색 결과 집합 내의 한 특정 문서가 동일한 검색 결과 집합 내의 다른 문서들보다 현저하게 많이 반복적으로 접근된다면, 클릭 정보는 의심되는 클릭 정보이다. 사용자가 한 문서를 동일한 검색 결과 집합 내의 다른 문서들보다 현저하게 높은 빈도로 반복 클릭하는 것은 비정상일 것이라는 점을 전제로 한다. 질의가 부정의 응답을 받는다면, 방법(400)은 원래의 클릭 정보가 사용되는 단계(492)로 진행한다. 질의가 긍정의 응답을 받는다면, 방법(400)은 클릭 정보의 평균치가 사용되는 단계(490)로 진행한다. 방법(400)은 단계(495)에서 종료한다.

[0043] 도 5는 한 유형에 대한 설정 변수 α 를 계산하는 방법(500)의 흐름도를 도시한다. 더 상세하게는 방법(500)은 각각의 $\langle k, t \rangle$ 쌍에 대한 $\alpha_{k,p,MAX}$ 및 $\alpha_{k,p,MIN}$ 를 계산한다. 설정 변수는 여러 유형의 구매 및/또는 클릭의 효과를 기술하기 위해 사용된다. 방법(500)은 단계(505)에서 시작하여 단계(510)로 진행한다.

[0044] 방법(500)은 단계(510)에서 튜플 $\langle k, t \rangle$ 를 선택하며, k 는 검색어이고 t 는 유형이다. 그다음 단계(520)에서 방법(500)은 $\langle k, t \rangle$ 에 대한 $C_{k,t,p}$ 를 선택하며, k 는 검색어, p 는 상품이고 t 는 유형이다. 즉 $C_{k,t,p}$ 는 상품 p 의 검색어 k 에 대해 일정 기간동안 발생한 유형 t 의 이벤트의 횟수 또는 수이다.

[0045] 단계 (530)에서, 방법(500)은 설정 변수 α 를 계산한다. 더 상세하게는, α 는 다음과 같이 표현될 수 있다.

$$\alpha_{k,t,MIN} = m_t \quad \text{(수식 1)}$$

$$\alpha_{k,t,MAX} = m_t / \text{MAX}(C_{k,1,t}, C_{k,2,t}, \dots, C_{k,n,t}) \quad \text{(수식 2)}$$

[0048] m_t 는 아래의 표 1 및 2에 도시된 바와 같이 유형 t 의 이벤트의 기본 점수이며, 이는 두개의 서로 다른 사업상 필요조건에 기반하여 정의된다. 각각의 유형 t 의 이벤트에 대해 수식 1 및 2의 "min(최소값)" 또는 "max(최대값)" 함수는 아래에 개시된 바와 같이 사용될 수 있다는 점을 주목해야 한다.

유형	m_t
선호 판매자 판매량 최소값	150
관련 검색 선호 판매자 판매량 최소값	120
선호 판매자 클릭수 최대값	100
비선호(스토어) 판매량 최대값	80
카탈로그 판매량 최소값	600
맵핑된 검색 카탈로그 판매량 최소값	500
맵핑된 카탈로그 클릭수 최대값	550
관련 검색 맵핑된 카탈로그 판매량 최소값	450
맵핑된 카탈로그 클릭수 최대값	160
정보 기반 판매량 최소값	580

[0050] <표 1>

유형	m_t
선호 판매자 판매량 최소값	110
관련 검색 선호 판매자 판매량 최소값	105
선호 판매자 클릭수 최대값	100
비선호(스토어) 클릭수 최소값	105

카탈로그 판매량 최소값	600
관련 검색 카탈로그 판매량 최소값	500
맵핑된 카탈로그 판매량 최소값	550
관련 검색 맵핑된 카탈로그 판매량 최소값	450
맵핑된 카탈로그 클릭수 최대값	160
정보 기반 판매량 최소값	550

[0052] <표 2>

[0053] 다양한 유형의 판매량 및 클릭수에 할당된 m_i 값은 각각의 실행을 나타내기 위해 조정될 수 있다. 다음의 유형들은 아래와 같이 정의된다:

[0054] 선호 판매자 판매량은 선호 판매자에 의해 이루어지는 판매량으로 정의된다. 한 판매자를 선호 판매자로 정의하는 기준은, 예컨대 수수료를 검색 개체에 제공하는 판매자는 선호 판매자로 간주될 수 있는 것과 같이, 적용예에 따라 특정된다.

[0055] 관련 검색 선호 판매자 판매량은 검색어에 관련되지만 선호 판매자의 이이 포함된 검색어로 이루어지는 판매량으로 정의된다. 설명을 위해, 두가지 검색어 "디지털 카메라(digital camera)"와 "소니 디지털 카메라(Sony digital camera)"가 있다고 가정하자. 검색어 "소니 디지털 카메라"로부터 생성된 검색 결과에서 상품 "A"를 구매하면 표 1의 m_i 값 120에 상품 "A"의 점수가 추가될 것이며, 검색어 "디지털 카메라"로부터 생성된 검색 결과에서 상품 "A"를 구매하면 표 1의 m_i 값 150에 상품 "A"의 점수가 추가될 것이다. 이러한 접근은 더 협의인 검색어인 "소니 디지털 카메라"를 더 광의이고 일반적인 검색어인 "디지털 카메라"와 연관시키는 것이다.

[0056] 선호 판매자 클릭수는 선호 판매자에 관한 검색 결과 집합 내의 문서의 클릭수로 정의된다.

[0057] 비선호 판매량은 예컨대 소규모 판매자와 같은 비선호 판매자에 의해 이루어지는 판매량으로 정의된다. 판매자를 비선호 판매자로 적용하는 기준은 예컨대 검색 개체에 대해 수수료가 적거나 수수료를 내지 않는 소규모 판매자는 비선호 판매자로 간주될 수 있는 것과 같이, 적용예에 따라 특정된다.

[0058] 카탈로그 판매량은 카탈로그 페이지 또는 상품 가이드 페이지에서 이루어진 판매량으로 정의된다. 카탈로그 페이지는 다음의 정보 중 하나 이상을 표시하는 특정 상품에 대한 디스플레이 페이지로 정의된다: 판매자 목록, 판매자-가격 쌍 목록(예컨대 특정 가격에 상품을 판매하는 판매자), 상품 리뷰 목록, 상품 상세 설명 및 기타 사항. 상기 카탈로그 페이지로부터 이루어진 구매는 검색어와 상당히 관련 있는 것으로 추정된다.

[0059] 관련 카탈로그 판매량은 관련 카탈로그 페이지 또는 상품 가이드 페이지로부터 이루어진 판매량으로 정의된다. 설명을 위해, 두가지 검색어 "디지털 카메라"와 "소니 디지털 카메라"가 있다고 가정해보자. 검색어 "소니 디지털 카메라"로부터 생성된 카탈로그 페이지에서 상품 "A"를 구매하면 표 1의 m_i 값 500에 검색어 "디지털 카메라"에 대한 상품 "A"의 점수가 추가될 것이며, 검색어 "디지털 카메라"로부터 생성된 카탈로그 페이지에서 상품 "A"를 구매하면 표 1의 m_i 값 600에 상품 "A"의 점수가 추가될 것이다.

[0060] 맵핑된(mapped) 카탈로그 판매량은 맵핑된 카탈로그 페이지 또는 상품 가이드 페이지와 연관된 판매량으로 정의된다. 즉, 구매가 카탈로그 페이지로부터 이루어지지 않지만, 대신에 판매자의 페이지를 통해 직접 구매가 이루어진다. 예컨대, 특정 검색어에 대한 검색 결과가 다수의 카탈로그 페이지와 다수의 판매자 페이지를 포함한다. 그러면 사용자는 특정 판매자 페이지로의 접근을 선택하면 상품의 구매가 판매자와 직접 이루어진다. 따라서 상품의 구매가 특정 판매자로부터 직접 구매되었는지 검출되고, 시스템도 구매한 상품이 특정 카탈로그 페이지 또는 상품 가이드 페이지에 "맵핑된" 것인지를 검출하면, 구매 정보는 표 1의 m_i 값 550에 카탈로그 페이지의 점수가 추가될 것이다. 최신점수는 문서들에 대해 폭넓게 생성되며, 문서는 상품, 판매자-상품 쌍 또는 카탈로그 페이지를 포함할 수 있음을 주의해야 한다. 관련 카탈로그 페이지에 높은 점수를 할당하는 것은 사용자가 동일한 상품을 판매하는 판매자들 간의 비교 결과를 제공받을 수 있기 때문에 바람직하다. 다시 말해서, 카탈로그 페이지에서 상품을 구매하는 것이 이상적인 쇼핑 환경이며, 여기서 높은 최신점수를 할당함으로써 카탈로그 페이지가 사용자에게 빈번하게 제공될 것이다.

[0061] 관련 검색 맵핑된 카탈로그 판매량은 관련하여 맵핑된 카탈로그 페이지 또는 관련하여 맵핑된 상품 가이드 페이지에 관련된 판매량으로 정의된다.

- [0062] 맵핑된 카탈로그 클릭수는 카탈로그 페이지 또는 상품 가이드 페이지에 맵핑될 수 있는 판매자 페이지의 클릭수로 정의된다. 즉, 클릭이 카탈로그 페이지에서 일어나지 않지만, 대신에 판매자 페이지에서 직접 클릭이 발생한다. 예컨대, 특정 검색어에 대한 검색 결과는 다수의 카탈로그 페이지와 다수의 판매자 페이지를 포함한다. 그러면 사용자는 상품에 대한 특정 판매자 페이지의 클릭을 선택한다. 시스템도 클릭된 상품이 특정 카탈로그 페이지 또는 상품 가이드 페이지에 "맵핑된" 것으로 검출하면, 클릭 정보는 표 1의 m_i 값 160에 카탈로그 페이지의 점수가 추가될 것이다.
- [0063] 정보 기반 판매량은 검색어의 몇가지 정보에 기반하여 조정된 결과로 이루어진 판매량으로 정의된다. 예컨대, 검색어가 "소니(Sony)"라면, 검색어는 "상표(brand)=소니(Sony)"로 조정된다. 상기 검색 결과로부터의 상품 판매량에 의해, 구매된 상품은 표 1의 m_i 값 580을 부여받을 것이다.
- [0064] 도 5로 돌아가서, 단계(540)에서, 방법(500)은 모든 $C_{k,p,t}$ 가 예컨대 상기 수식 2와 관련하여 계산되었는지를 질의한다. 질의가 부정의 응답을 받으면, 방법(500)은 단계(520)로 돌아간다. 질의가 긍정의 응답을 받으면, 방법(500)은 단계(550)으로 진행한다.
- [0065] 단계(550)에서, 방법(500)은 모든 토플 <k,t>가 요약되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(500)은 단계(510)로 돌아간다. 질의가 긍정의 응답을 받으면, 방법(500)은 단계(555)에서 종결된다.
- [0066] 도 6은 본 발명의 최신점수를 생성하는 방법(600)의 흐름도를 도시한다. 방법(600)은 단계(605)에서 시작하여 단계(610)로 진행한다.
- [0067] 단계(610)에서, 방법(600)은 상세 설정이 최신점수를 생성하기 위해 선정되었는지 선택적으로 질의한다. 즉, 일 실시예에서, 다수의 설정 또는 공식이 시스템의 다양한 필요사항을 처리하기 위해 사용될 수 있다. 예컨대, 몇몇 시스템은 최신점수의 사용을 선호할 수 있으며, 따라서 최대값(MAX) 설정이 선택되고, 최신점수는 검색 결과 집합에 나열된 문서들에 큰 영향을 미칠 것이다. 이외에, 몇몇 시스템은 최신점수의 사용을 경감시키기를 원할 수 있으며, 따라서 최소값(MIN) 설정이 선택되고, 최신점수는 검색 결과 집합 내에 나열된 문서에 더 적은 영향을 미칠 것이다.
- [0068] 그러나 다수의 설정이 고려되지 않는다면, 단계(610)은 생략되고 표준 설정이 선택된다. 질의가 부정의 응답을 받으면, 방법(600)은 단계(615)로 진행하며, 설정이 선택된다. 질의가 긍정의 응답을 받으면, 방법(600)은 단계(620)로 진행한다.
- [0069] 방법(600)은 단계(620)에서 토플 <k,p>를 선택하며, k는 검색어이고 p는 상품이다. 그러면 방법(600)은 단계(630)에서 유형 t를 선택한다.
- [0070] 단계(640)에서, 방법(600)은 <k,p,t>에 대한 $C_{k,p,t}$ 가 존재하는지 질의하며, k는 검색어, p는 상품이고 t는 유형이다. $C_{k,p,t}$ 는 상품 p에 대한 검색어 k에 대해 일정 기간 동안 발생한 유형 t인 이벤트의 횟수 또는 수이다. 질의가 부정의 응답을 받으면, 방법(600)은 단계(630)로 돌아가며, 다른 유형이 선택된다. 질의가 긍정의 응답을 받으면, 방법(600)은 단계(650)로 진행한다.
- [0071] 단계(650)에서, 방법(600)이 선택된 설정에 관한 설정 변수 α 를 계산한다. 일 실시예에서, 검색어 k에 대한 판매자/상품 쌍 p의 최신점수는 다음과 같이 정의된다.

$$HotScore_{k,p} = \sum (\alpha_{k,t,T(t)} C_{k,p,t}) \quad (\text{수식 3})$$

- [0072]
- [0073] $C_{k,p,t}$ 는 상품 p의 검색어 k에 대한 유형 t의 이벤트가 발생한 수이다. $\alpha_{k,t,T(t)}$ 는 상기 수식 2 및 수식 3에서 정의된 설정 변수이다.
- [0074] 일 실시예에서, T(t) 함수는 예컨대 T(t)가 최대값(MAX) 함수 또는 최소값(MIN) 함수가 될 수 있는 것으로 정의될 수 있다. 함수값의 예는 상기 표 1 및 2에 개시되어 있다. T(t) 함수에 대한 값은 점수와 시스템의 설정에서 미리 정의될 수 있다. 본 발명이 두개의 설정 함수인 최대값(MAX)과 최소값(MIN)을 개시하고 있지만, 본 발명이 이에 한정되는 것은 아니다. 즉 임의의 갯수의 설정이 특정 점수와 시스템의 필요사항을 처리하기 위해 사용될 수 있다.
- [0075] 단계(660)에서, 방법(600)은 모든 유형 t가 처리되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(600)

0)은 단계(630)로 돌아간다. 질의가 긍정의 응답을 받으면, 방법(600)은 단계(670)로 진행하며, 선택된 투플 <k,p>에 대한 최신점수를 생성하기 위해 수식 3이 사용된다.

[0076] 단계(680)에서, 방법(600)은 모든 투플 <k,p>가 처리되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(600)은 단계(620)로 돌아가며, 다른 투플이 선택된다. 질의가 긍정의 응답을 받으면, 방법(600)은 단계(685)에서 종료된다.

[0077] 일 실시예에서, 본 발명의 최신점수는 기존의 검색 점수화 시스템에서 사용된다. 설명을 위해, 검색어 k, 판매자/상품 쌍 p는 다음과 같이 점수(score_{k,p})를 획득한다.

$$\text{Score}_{k,p} = \text{BT}_{k,p} + H(\text{hotscore}_{k,p}) + \text{OB}_{k,p} \quad (\text{수식 } 4)$$

[0079] BT_{k,p}는 검색어 k에 대해 상품 p가 획득한 기본 텍스트 관련도 점수이며, 최신점수(hotscore_{k,p})는 검색어 k에 대한 p의 최신점수이며, H는 필요하다면 검색 점수화 체계에 따라 최신점수를 조정하는 사용 함수이며, OB_{k,p}는 검색어 k에 대해 추가적으로 증폭된 나머지 점수의 총합이다. H는 최신점수가 아래에 개시될 전체 점수에 어떻게 사용될 것인지를 설명하는 함수임을 주목해야 한다.

[0080] 다수의 표준화 함수가 적용될 수 있다. 다양한 유형의 함수가 아래에 기술된다.

[0081] 일 실시예에서, 원래의 최신점수는 다음과 같이 표현되는 "작용 인자(affect factor)"로 표준화된다.

$$H(\text{hotscore}_{k,p}) = \text{hotscore}_{k,p} * af \quad (\text{수식 } 5)$$

[0083] af는 작용 인자이며, 다음과 같이 정의될 수 있다.

$$af = \text{standard_hotscore} / \text{standard_score_for_hotscore_in_whole_score} \quad (\text{수식 } 6)$$

[0085] 상기 함수는 최신점수 중에 표준에 따라 한 점수를 선택하며, 최신점수 부분의 표준 점수에 따라 전체 점수 중에 한 점수를 선택한다. 그다음 작용 인자를 사용하여 최신점수가 전체 점수에 적용된다. 이러한 접근에는, 최신점수의 사용에 대한 상한선 또는 하한선의 설정이 없다. 따라서 상당히 높은 신뢰도의 상품은 높은 순위를 보장받을 것이다.

[0086] 제2 실시예에서, 최신점수는 다음과 같이 표준화될 수 있다.

$$\begin{aligned} &\text{If } \text{hotscore}_{k,p} = 0, \text{ then } H(\text{hotscore}_{k,p}) = 0; \\ &\text{Otherwise,} \\ &H(h_{k,p}) = H_L + (H_U - H_L) * (h_{k,p} - \text{MIN}(h_{k,1}, h_{k,2}, \dots, h_{k,n}) / (\text{MAX}(h_{k,1}, h_{k,2}, \dots, h_{k,n}) - \text{MIN}(h_{k,1}, h_{k,2}, \dots, h_{k,n}))) \end{aligned} \quad (\text{수식 } 7)$$

[0088] H_L은 총 점수 중에 최신점수의 하한선이며, H_U는 총 점수 중 최신점수의 상한선이다. 함수 H는 검색 점수화에 있어서 최신점수가 어느 정도의 큰 비중을 차지해야 하는지를 결정한다. H_U는 점수에 있어서 최신점수가 갖는 최대치의 영향을 정의하며, H_L은 점수에 있어서 최신점수가 갖는 최소치의 영향을 정의한다.

[0089] 한 극단적인 체계에서는 H_U와 H_L에 상당히 큰 값을 부여하며, 그 결과 최신점수가 전체 점수에 큰 영향을 미칠 것이다. 이외에 다른 극단적인 체계에서는 H_U와 H_L에 상당히 작은 값을 부여하며, 그 결과 최신점수는 수식 4에서 동일한 BT_{k,p}와 OB_{k,p}를 갖는 상품의 순위에만 영향을 줄 것이다. 전자의 접근법은 폐쇄형 시스템(closed system)에 적합하며, 상기 시스템에서는 모든 상거래 정보가 입수가 가능하다. 몇몇의 판매 정보만이 입수가 가능한

개방형 시스템(open system)에서는, 높은 신뢰도의 최신점수가 점수에 큰 영향을 미치게 하고, 반면에 낮은 신뢰도의 최신점수는 아주 한정된 작용만을 하고 다른 점수 효과 요소들과 혼합되도록 H_0 에만 높은 값을 부여하는 것이 더 적절할 수 있다.

[0090] 제3 실시예에서, 최신점수는 표준화되어 놓여질 수 있다. 특히, AC_i 는 위치 i 에서의 총 클릭수이고, $C_{k,p,i}$ 가 위치 i 에서 검색어 k 에 대한 상품 p 의 클릭수이며, $NC_{k,p,i}$ 가 위치 i 에서 검색어 k 에 대한 상품 p 의 표준화된 클릭수라 하면, 다음과 같이 된다.

$$NC_{k,p,i} = C_{k,p,i} * AC_0 / AC_i \quad (수식 8)$$

[0092] AC_0/AC_i 는 위치 i 에 대한 일반적인 증폭 인자이다. 검색 결과 집합 내에 상당히 상위에 있는 문서의 클릭수의 영향을 감소시키기 위해, 본 발명의 접근법은 상위에서 하나의 부정 클릭이 불균형하게 전체 점수와 시스템에 영향을 주지 않도록 AC_0 과 같이 일정한 숫자로 AC_i 를 제한할 수 있다.

[0093] 추가적으로, $\langle k, p \rangle$ 쌍에서의 클릭 위치는 여러 날에 걸쳐 달라질 수 있기때문에, i 는 주어진 기간동안 $\langle k, p \rangle$ 의 평균 클릭 위치를 계산하여 결정된다.

[0094] 상기 함수는 하나의 $\langle k, p \rangle$ 쌍에 대한 한 위치에서의 클릭수와 평균 클릭수를 비교한다. 평균 클릭률보다 더 높아야만 표준화된 후에 더 높은 수를 가질 수 있으며, 즉 실제로 $C_{k,p,0}/C_{k,p,i}$ 와 AC_0/AC_i 를 비교한다. 따라서 이러한 접근법은 자가-증폭의 가능성을 최소화할 것이다. 동일한 함수가 판매 위치 표준화에 적용될 수 있음을 주의해야 한다.

[0095] 제4 실시예에서, 최신점수는 시간상 표준화될 수 있다. 특히 E 는 이벤트가 발생한 횟수이며, NE 는 이벤트에 대한 표준화된 횟수이며, age 는 이벤트가 발생한 당시 시간으로부터 지금까지의 일수이고, ff 는 망각 인자(forget factor), 즉 시스템이 이벤트를 망각하는 경향의 비율이다. 망각 인자는 본 발명의 시스템이 망각 인자를 적절히 조정할 수 있도록 설정 화일에서 정의된다. E 는 다음에 따라서 표준화된다.

$$NE = E * (1 - ff)^{age}, \quad (0 \leq age \leq n). \quad (수식 9)$$

[0097] 수식 9에서 "age"의 상한선 범위(n)는 특정 어플리케이션의 필요사항을 충족시키기 위해, 또는 다양한 상품에 따라서 조정될 수 있다.

[0098] 도 7은 정보 변수에 기반하여 본 발명의 최신점수를 조정하는 방법(700)의 흐름도를 도시한다. 방법(700)은 단계(705)에서 시작하여 단계(710)로 진행한다.

[0099] 단계(710)에서, 방법(700)은 지식 기반으로부터 검색어 k 를 선택한다. 즉, 정보 KN_k 를 검색한다. 예컨대, 검색어가 "델(dell)"이라면, 정보 KN_k 는 "제조사(manufacturer)=델(dell)"로 표시될 수 있다.

[0100] 단계(720)에서, 방법(700)은 정보 KN_k 를 적용하기 위한 설정 인자 또는 수식이 존재하는지 질의한다. 예컨대 설정 인자는 델(Dell)의 모든 상품이 델의 모든 상품의 판매량을 나타내도록 그 최신점수를 조정할 것을 지시할 수 있다. 대안적으로, 설정 인자는 모든 델 컴퓨터 상품이 모든 델 컴퓨터 상품의 판매량과 기타 등등을 나타내도록 그 최신점수를 조정할 것을 지시할 수 있다. 질의가 부정의 응답을 받으면, 방법(700)은 단계(710)로 돌아가고 다른 검색어가 선택된다. 질의가 긍정의 응답을 받으면, 방법(700)은 단계(730)로 진행한다.

[0101] 단계(730)에서, 방법(700)은 각각의 상품($P_{KNk1}, \dots, P_{KNkn}$)에 대한 정보 KN_k 에 관한 모든 판매량 정보를 검색한다. 예컨대, 데스크탑 컴퓨터, 랩탑, PDA, 프린터, 모니터, 스피커 및 기타에 대한 판매량 정보가 수집된다. 상기 정보는 아래와 같이 적용될 수 있다.

[0102] 단계(740)에서, 방법(700)은 선택적으로 전술한 시간과 위치의 표준화를 적용할 수 있다.

[0103] 단계(750)에서, 방법(700)은 단계(730)에서 언급된 상품들 중에 한 상품 p 를 선택한다. 예컨대 델(Dell) 데스크탑 컴퓨터가 선택된다.

- [0104] 단계(760)에서, 방법(700)은 단계(720)에서 언급된 설정 인자 또는 수식에 기반하여 최신점수(hotscore_{k,p})를 조정한다. 예컨대, 델(De11) 데스크탑 컴퓨터에 대한 최신점수는, 델(De11) 랩탑에 대한 판매량 정보가 델(De11) 데스크탑 컴퓨터에 대한 최신점수를 증폭하기 위해 사용되도록 조정된다. 상기 조정의 원리는 델(De11)이 선호 판매자라는 점 또는 델(De11) 랩탑을 선호하는 구매자는 마찬가지로 델(De11) 데스크탑을 선호할 것이라는 정보가 있다는 점이 될 수 있다. 상기 방법에서, 최신점수를 더 개선하기 위해 상세 정보가 이용될 수 있다.
- [0105] 단계(770)에서, 방법(700)은 모든 관련 상품이 조정되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(700)은 단계(750)로 돌아가며 다른 상품이 선택된다. 질의가 긍정의 응답을 받으면, 방법(700)은 단계(780)로 진행한다.
- [0106] 단계(780)에서, 방법(700)은 모든 관련 정보가 처리되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(700)은 단계(710)로 돌아가서 다른 검색어가 선택된다. 질의가 긍정의 응답을 받으면, 방법(700)은 단계(785)에서 종료된다.
- [0107] 도 8은 관련된 협의의 검색에 기반하여 본 발명의 최신점수를 조정하는 방법(800)의 흐름도를 도시한다. 방법(800)은 단계(805)에서 시작하여 단계(810)로 진행한다.
- [0108] 단계(810)에서, 방법(800)은 관련된 협의의 검색을 적용하기 위한 설정 인자 또는 수식이 존재하는지 질의한다. 예컨대, 검색어 "SDRAM 장착 컴퓨터"는 "컴퓨터"에 대한 협의의 검색어로 간주될 것이다. 질의가 부정의 응답을 받으면, 방법(800)은 단계(890)에서 종료된다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(820)로 진행한다.
- [0109] 단계(820)에서, 방법(800)은 검색어 k를 선택한다. 그 다음, 방법(800)은 단계(830)에서 관련된 협의의 검색어 k₁을 선택한다.
- [0110] 단계(840)에서, 방법(800)은 관련된 협의의 검색어 k₁에 관한 판매량 및/또는 클릭 정보가 있는지 질의한다. 예컨대, 방법(800)은 "SDRAM 장착 컴퓨터"라는 검색어에 관한 어떤 판매량 정보가 있는지를 결정할 수 있다. 질의가 부정의 응답을 받으면, 방법(800)은 단계(830)로 돌아가서 다른 관련 검색어 k_n을 선택한다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(850)로 진행한다.
- [0111] 단계(850)에서, 방법(800)은 관련 검색어에 대한 판매량 정보가 기준치보다 높은지를 질의한다. 다시 말하면, 방법(800)은 판매량 정보가 검색어 k에 대한 최신점수의 조정에 사용하기에 신뢰할 수 있는 것인지 결정한다. 일 실시예에서, 판매량 정보가 실제로 광의이며 더 일반적인 검색어에 영향을 주도록 적용되기 전에 관련된 협의의 검색어에 대한 상당한 판매량이 있는지를 입증하는 것이 바람직할 수 있다. 따라서 질의가 부정의 응답을 받으면, 방법(800)은 단계(830)로 돌아가서 다른 관련 검색어 k_n을 선택한다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(860)로 진행한다.
- [0112] 단계(860)에서, 방법(800)은 검색어 k로부터 도출된 검색 결과 집합에 나열된 상품으로부터 최신점수를 선택한다. 그 다음, 최신점수(hotscore_{k,p})가 검색어 k₁에 관한 판매량 및/또는 클릭수 정보와 관련하여 조정된다. 실제로, 최신점수(hotscore_{k,p})는 최신점수(hotscore_{k₁,p})와 관련하여 직접 조정될 수 있다.
- [0113] 단계(870)에서, 방법(800)은 검색어 k로부터 도출된 검색 결과 집합의 상품들의 모든 최신점수가 조정되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(800)은 단계(860)로 돌아가서 다른 상품을 선택한다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(880)로 진행한다.
- [0114] 단계(880)에서, 방법(800)은 모든 관련된 협의의 검색어가 처리되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(800)은 단계(830)로 돌아가서 다른 검색어를 선택한다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(885)로 진행한다.
- [0115] 단계(885)에서, 방법(800)은 모든 일반 검색어가 처리되었는지 질의한다. 질의가 부정의 응답을 받으면, 방법(800)은 단계(820)로 돌아가서 다른 일반 검색어를 선택한다. 질의가 긍정의 응답을 받으면, 방법(800)은 단계(890)에서 종료된다.
- [0116] 상기 개시된 사항들은 쇼핑 환경에서의 본 발명을 설명함을 주의해야 한다. 그러나 기술 분야의 당업자라면 본 발명이 그에 한정되지 않음을 인식할 것이다. 즉, 일 실시예에서, 본 발명은 예컨대 클릭 정보와 관련하여

점수를 생성하는 것과 같은 일반 검색에 이용될 수 있다.

[0117] 다양한 실시예가 전술되었지만, 상기 실시예들은 예시적으로 제공된 것이며 이에 제한되지 않는 것으로 이해되어야 한다. 따라서 바람직한 실시예의 영역과 범위는 전술한 예시적인 실시예들 중 어느 것에 의해서도 제한되지 않아야 하며, 후술할 특허청구범위와 그 균등물에 의해서만 정의되어야 한다.

도면의 간단한 설명

[0118] 전술한 사항 및 다른 사항과 장점들은 도면을 참조하여 후술할 본 발명의 바람직한 실시예의 상세한 설명으로부터 더 잘 이해될 수 있다.

[0119] 도 1은 본 발명의 점수화 시스템을 도시한 블록도(block diagram).

[0120] 도 2는 본 발명의 점수화 방법의 적용과 검색 결과의 문서 순위 목록의 결과 관계를 도시한 도면.

[0121] 도 3은 다수의 상품에 대한 최신점수(hotscores)를 생성하는 방법의 흐름도를 도시한 도면.

[0122] 도 4는 판매 및 클릭 데이터를 전처리하는 방법의 흐름도를 도시한 도면.

[0123] 도 5는 설정 변수(configuration parameter) α 를 계산하는 방법의 흐름도를 도시한 도면.

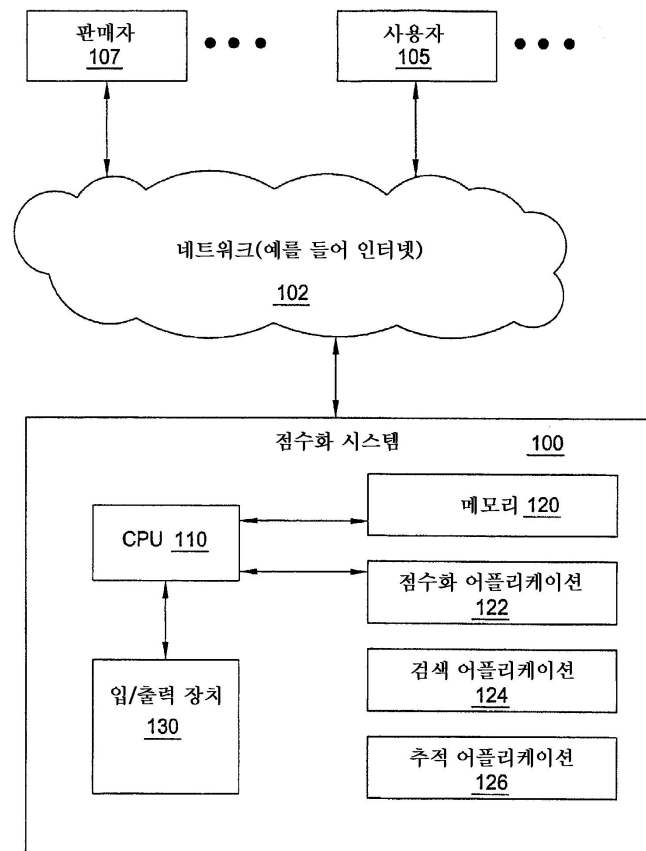
[0124] 도 6은 본 발명의 최신점수를 생성하는 방법의 흐름도를 도시한 도면.

[0125] 도 7은 본 발명의 최신점수를 조정하는 방법의 흐름도를 도시한 도면.

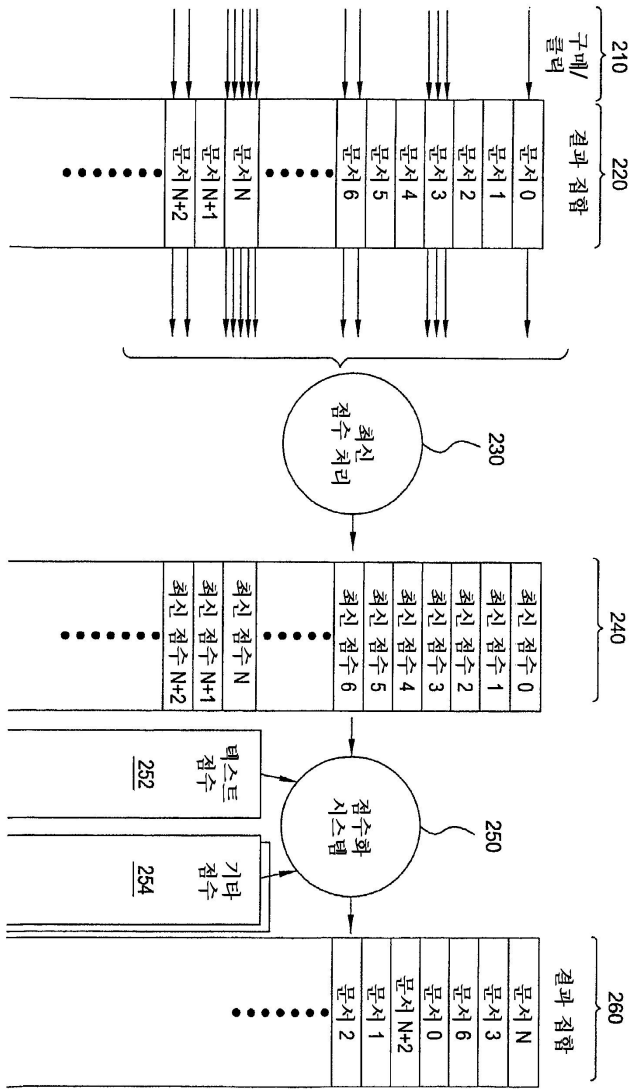
[0126] 도 8은 본 발명의 최신점수를 조정하는 제2 방법의 흐름도를 도시한 도면.

도면

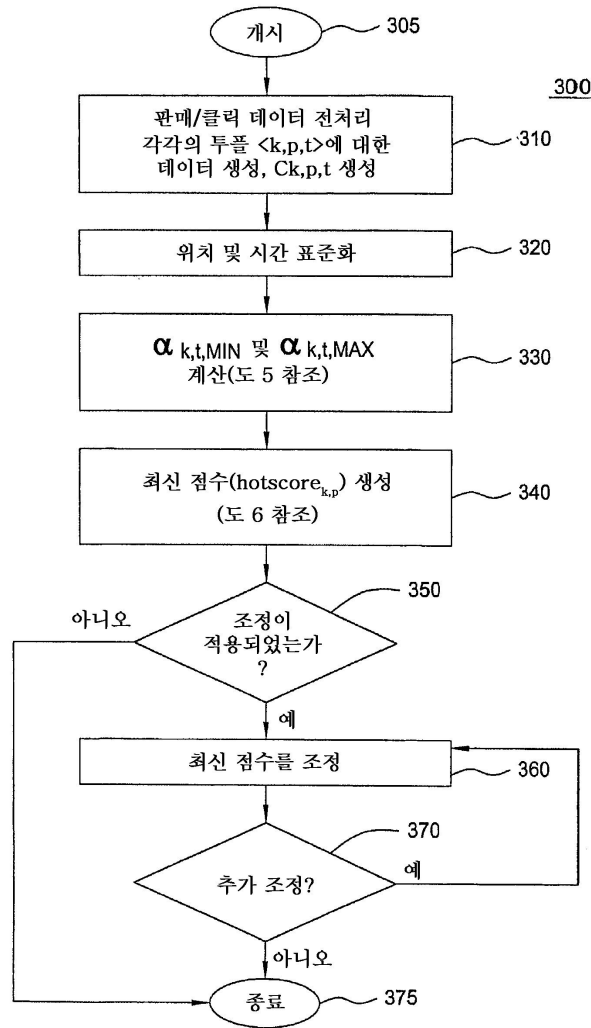
도면1



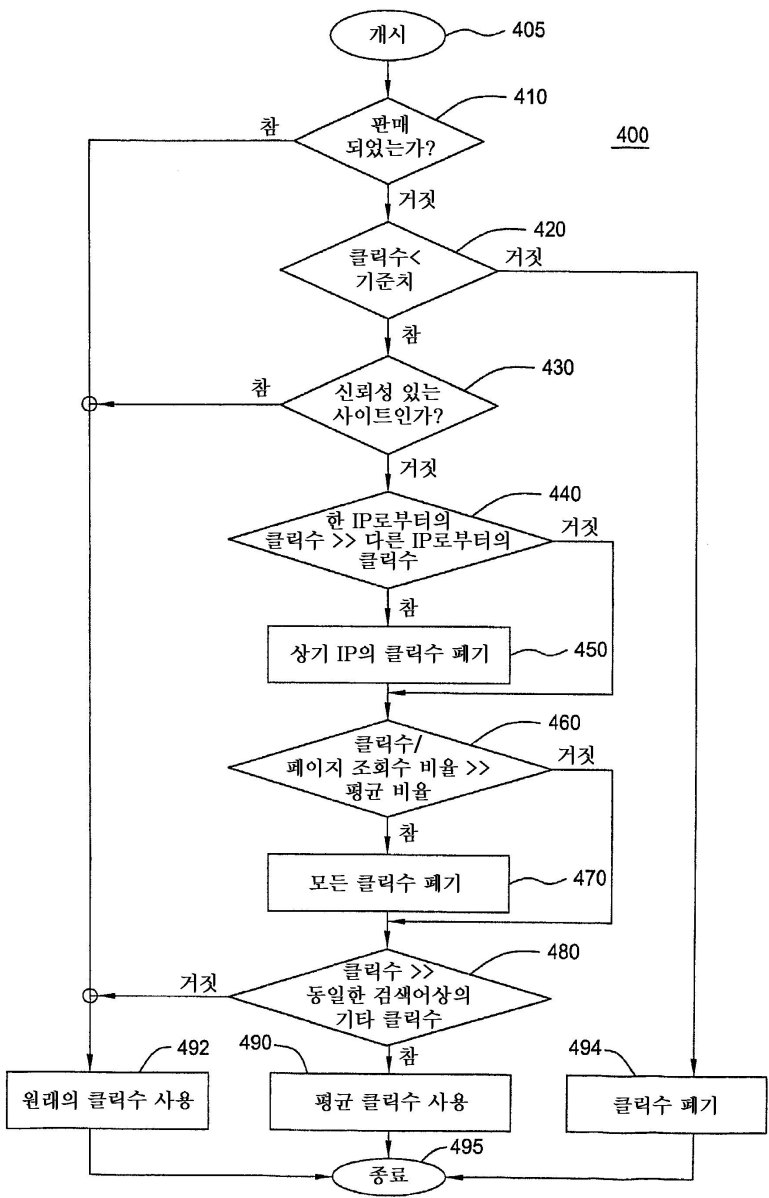
도면2



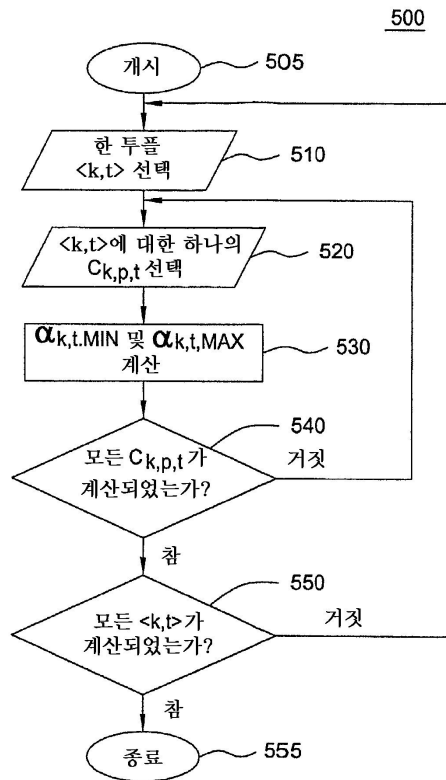
도면3



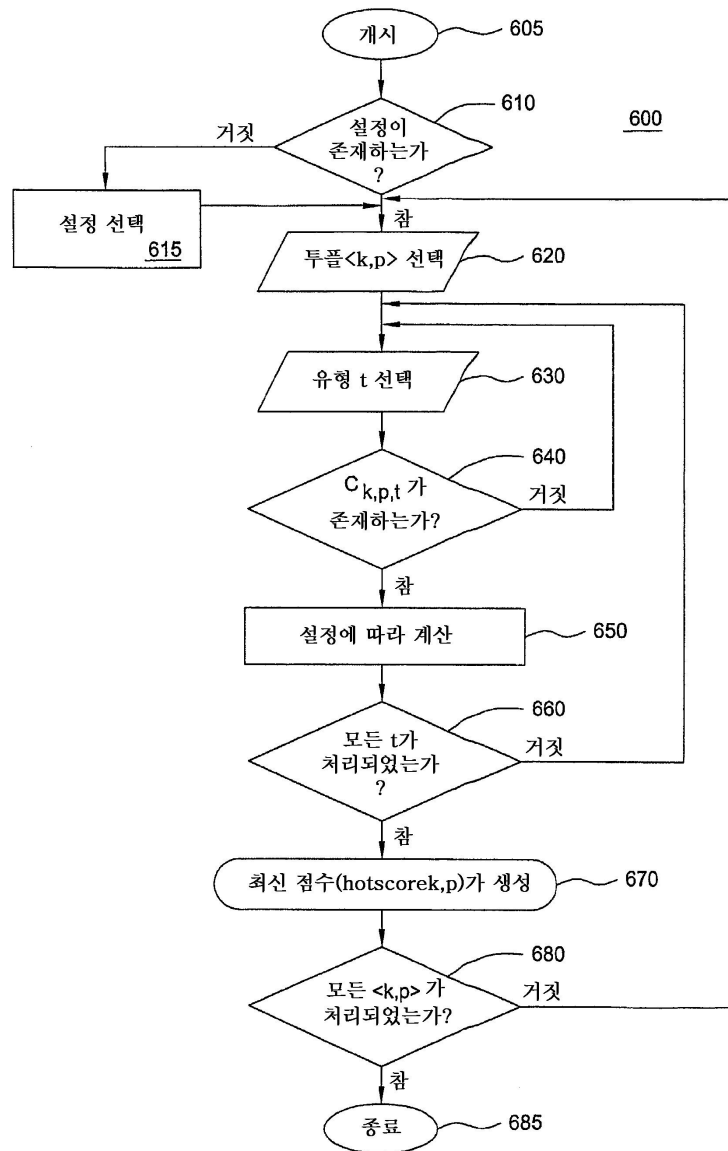
도면4



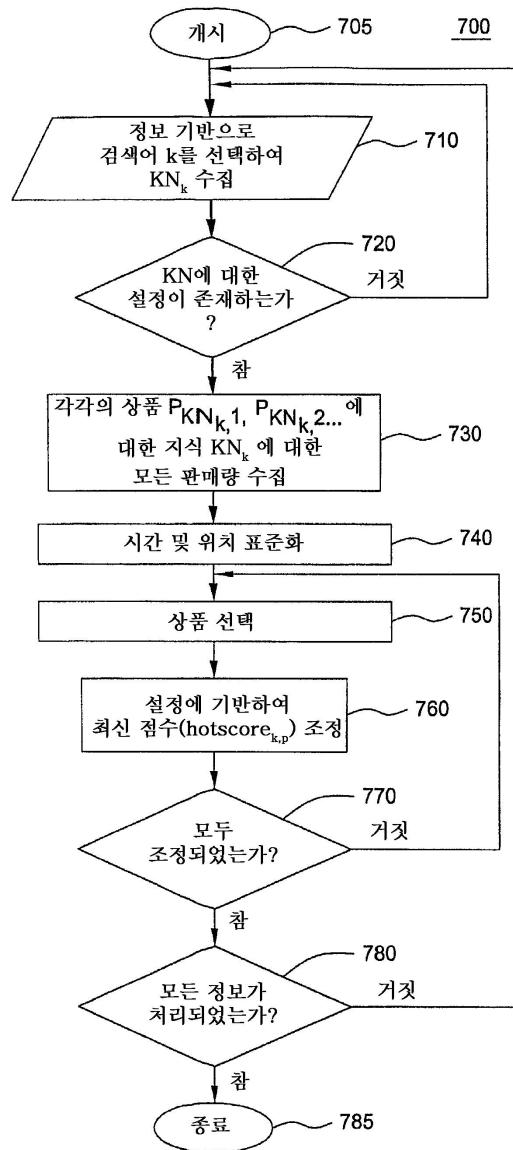
도면5



도면6



도면7



도면8

