



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2013년01월10일
 (11) 등록번호 10-1220080
 (24) 등록일자 2013년01월02일

(51) 국제특허분류(Int. Cl.)
 G06F 17/30 (2006.01) G06F 17/27 (2006.01)
 (21) 출원번호 10-2010-0140403
 (22) 출원일자 2010년12월31일
 심사청구일자 2010년12월31일
 (65) 공개번호 10-2012-0089963
 (43) 공개일자 2012년08월16일
 (56) 선행기술조사문헌
 JP2009134700 A*
 KR1020040029895 A*
 KR1020080074116 A*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 엔에이치엔(주)
 경기도 성남시 분당구 불정로 6, 그린팩토리 (정자동)
 (72) 발명자
 신재승
 경기도 성남시 분당구 불정로 6, NHN 그린팩토리 (정자동)
 박영섭
 경기도 성남시 분당구 불정로 6, NHN 그린팩토리 (정자동)
 (뒷면에 계속)
 (74) 대리인
 특허법인무한

전체 청구항 수 : 총 30 항

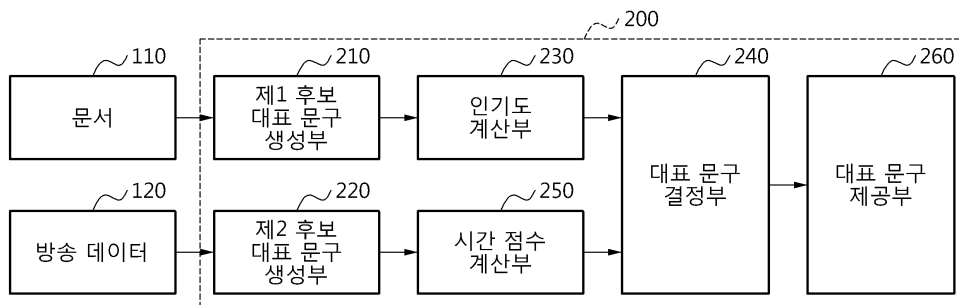
심사관 : 이명진

(54) 발명의 명칭 **실시간 인기 키워드에 대한 대표 문구 제공 방법 및 시스템**

(57) 요약

실시간 인기 키워드에 대한 대표 문구를 제공하는 방법 및 시스템이 개시된다. 대표 문구 제공 방법 및 시스템은, 문서를 기초로 제1 후보 대표 문구를 생성하고, 방송 데이터를 기초로 제2 후보 대표 문구를 생성할 수 있다. 그리고, 제1 및 제2 후보 대표 문구 중 적합한 대표 문구를 인기 키워드와 결합하여 제공할 수 있다.

대표도 - 도2



(72) 발명자

최재걸

경기도 성남시 분당구 불정로 6, NHN 그린팩토리
(정자동)

노원숙

경기도 성남시 분당구 불정로 6, NHN 그린팩토리
(정자동)

특허청구의 범위

청구항 1

제1 후보 대표 문구 생성부, 제2 후보 대표 문구 생성부 및 대표 문구 결정부를 포함하는 대표 문구 제공 시스템이 수행하는 대표 문구 제공 방법에 있어서,

상기 제1 후보 대표 문구 생성부가 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계;

상기 제2 후보 대표 문구 생성부가 방송 데이터에서 상기 인기 키워드를 포함하는 방송 프로그램을 결정하고 상기 결정된 방송 프로그램의 방송 시간을 이용하여 상기 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계; 및

상기 대표 문구 결정부가 상기 제1 후보 대표 문구 및 상기 제2 후보 대표 문구를 이용하여 상기 인기 키워드에 대해 최종적인 대표 문구를 결정하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 2

제1항에 있어서,

상기 최종적인 대표 문구를 결정하는 단계는,

상기 대표 문구 결정부가 상기 제1 후보 대표 문구와 상기 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 방법.

청구항 3

제1항에 있어서,

상기 대표 문구 제공 시스템은 인기도 계산부를 더 포함하고,

상기 인기도 계산부가 상기 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 단계를 더 포함하는 대표 문구 제공 방법.

청구항 4

제3항에 있어서,

상기 대표 문구를 결정하는 단계는,

상기 대표 문구 결정부가 상기 문서의 인기도와 기설정된 기준 인기도를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제1 후보 대표 문구를 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 방법.

청구항 5

제2항에 있어서,

상기 대표 문구 제공 시스템은 시간 점수 계산부를 더 포함하고,

상기 시간 점수 계산부는 상기 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 단계

를 더 포함하는 대표 문구 제공 방법.

청구항 6

제5항에 있어서,

상기 대표 문구를 결정하는 단계는,

상기 대표 문구 결정부가 상기 제2 후보 대표 문구의 시간 점수와 기설정된 기준 시간 점수를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제2 후보 대표 문구를 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 방법.

청구항 7

제1항에 있어서,

상기 제1 후보 대표 문구를 생성하는 단계는,

상기 제1 후보 대표 문구 생성부의 기준 단어 결정부가 상기 인기 키워드를 포함하는 문서들의 기준 단어를 결정하는 단계;

상기 제1 후보 대표 문구 생성부의 기준 단어 결정부가 상기 기준 단어 중에서 대표 기준 단어를 결정하는 단계;

상기 제1 후보 대표 문구 생성부의 기준 단어 확장부가 대표 기준 단어의 이전 또는 이후에 연속하는 단어와 상기 대표 기준 단어를 조합하여 상기 대표 기준 단어를 확장하는 단계; 및

상기 제1 후보 대표 문구 생성부의 기준 단어 확장부가 확장된 대표 기준 단어를 이용하여 상기 제1 후보 대표 문구를 생성하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 8

제7항에 있어서,

상기 대표 기준 단어를 확장하는 단계는,

상기 제1 후보 대표 문구 생성부의 기준 단어 확장부가 상기 인기 키워드를 포함하는 문서들에서 상기 대표 기준 단어의 이전 또는 이후에 연속하는 단어가 포함되는 조건부 확률을 계산하는 단계; 및

상기 제1 후보 대표 문구 생성부의 기준 단어 확장부가 상기 조건부 확률에 기초하여 상기 대표 기준 단어를 확장하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 9

제7항에 있어서,

상기 대표 기준 단어를 결정하는 단계는,

상기 제1 후보 대표 문구 생성부의 기준 단어 결정부가 상기 인기 키워드를 포함하는 문서들의 형태소를 분석하여 상기 기준 단어의 빈도수를 카운트하는 단계; 및

상기 제1 후보 대표 문구 생성부의 기준 단어 결정부가 카운트된 빈도수에 기초하여 상기 대표 기준 단어를 결정하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 10

제1항에 있어서,

상기 대표 문구 제공 시스템은 문서 수집부 및 클러스터 수행부를 더 포함하고,

상기 문서 수집부가 상기 인기 키워드를 포함하는 문서들을 수집하는 단계; 및

상기 클러스터 수행부가 수집된 문서들에 대해 클러스터(Cluster)를 수행하는 단계

를 더 포함하고,

상기 제1 후보 대표 문구를 생성하는 단계는,

상기 제1 후보 대표 문구 생성부가 상기 클러스터에 속하는 문서들을 기초로 대표 기준 단어를 확장하여 상기 제1 후보 대표 문구를 생성하는 것을 특징으로 하는 대표 문구 제공 방법.

청구항 11

제10항에 있어서,

상기 제1 후보 대표 문구를 생성하는 단계는,

상기 대표 문구 제공부가 상기 수집된 문서들에 가중치를 추가하는 단계; 및

상기 대표 문구 제공부가 상기 가중치를 이용하여 클러스터의 노출 우선 순위를 결정하는 단계를 더 포함하는 대표 문구 제공 방법.

청구항 12

삭제

청구항 13

제1항에 있어서,

상기 제2 후보 대표 문구를 생성하는 단계는,

상기 제2 후보 대표 문구 생성부의 방송 프로그램 결정부가 상기 방송 프로그램이 복수 개인 경우, 복수의 방송 프로그램 중에서 방송 프로그램의 방송 시작 시간 및 방송 종료 시간이 상기 인기 키워드가 노출된 시간과 가장 인접한 어느 하나의 방송 프로그램을 선택하는 단계

를 더 포함하고,

상기 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계는,

상기 제2 후보 대표 문구 생성부의 생성부가 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 것을 특징으로 하는 대표 문구 제공 방법.

청구항 14

제1항에 있어서,

상기 제2 대표문구 생성부는 방송 프로그램 결정부 및 생성부를 더 포함하고,

상기 제2 후보 대표 문구를 생성하는 단계는,

상기 제2 후보 대표 문구 생성부의 방송 프로그램 결정부가 상기 인기 키워드가 여러 단어로 조합된 형태인 경우, 상기 인기 키워드를 형태소 분석하여 복수의 단어들로 분리하는 단계;

상기 제2 후보 대표 문구 생성부의 방송 프로그램 결정부가 상기 인기 키워드를 포함하는 방송 프로그램 중에서 분리된 복수의 단어들에 기초하여 어느 하나의 방송 프로그램을 선택하는 단계; 및

상기 제2 후보 대표 문구 생성부의 생성부가 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 15

제1항에 있어서,

상기 제2 대표문구 생성부는 매칭 점수 계산부, 방송 프로그램 결정부 및 생성부를 더 포함하고,

상기 제2 후보 대표 문구를 생성하는 단계는,

상기 제2 대표문구 생성부의 매칭 점수 계산부가 상기 인기 키워드를 포함하는 방송 프로그램에 매칭 점수, 시간 점수, 및 방송국 가중치를 추가하는 단계;

상기 제2 대표문구 생성부의 매칭 점수 계산부가 상기 매칭 점수, 시간 점수, 방송국 가중치 및 방송 프로그램의 시청률 중 적어도 하나를 이용하여 상기 방송 프로그램에 대한 최종 매칭 점수를 계산하는 단계;

상기 제2 대표문구 생성부의 방송 프로그램 결정부가 상기 인기 키워드를 포함하는 방송 프로그램 중에서 상기 최종 매칭 점수를 기초로 어느 하나의 방송 프로그램을 선택하는 단계; 및

상기 제2 대표문구 생성부의 생성부가 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계

를 포함하는 대표 문구 제공 방법.

청구항 16

제1항에 있어서,

상기 대표 문구 제공 시스템은 대표 문구 제공부를 더 포함하고,

상기 대표 문구 제공부가 상기 최종적인 대표 문구와 상기 인기 키워드를 결합하여 제공하는 단계

를 더 포함하는 대표 문구 제공 방법.

청구항 17

제1항 내지 제11항 및 제 13항 내지 제16항 중 어느 한 항의 방법을 실행하기 위한 프로그램이 기록된 컴퓨터에서 판독 가능한 기록 매체.

청구항 18

문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 제1 후보 대표 문구 생성부;

방송 데이터에서 상기 인기 키워드를 포함하는 방송 프로그램을 결정하고 상기 결정된 방송 프로그램의 방송 시간을 이용하여 상기 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 제2 후보 대표 문구 생성부; 및

상기 제1 후보 대표 문구 및 상기 제2 후보 대표 문구를 이용하여 상기 인기 키워드에 대해 최종적인 대표 문구를 결정하는 대표 문구 결정부

를 포함하는 대표 문구 제공 시스템.

청구항 19

제18항에 있어서,

상기 대표 문구 결정부는,

상기 제1 후보 대표 문구와 상기 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 20

제19항에 있어서,

상기 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 인기도 계산부

를 더 포함하는 대표 문구 제공 시스템.

청구항 21

제20항에 있어서,

상기 대표 문구 결정부는,

상기 문서의 인기도와 기설정된 기준 인기도를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 제1 후보 대표 문구를 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 22

제18항에 있어서,

상기 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 시간 점수 계산부

를 더 포함하는 대표 문구 제공 시스템.

청구항 23

제22항에 있어서,

상기 대표 문구 결정부는,

상기 제2 후보 대표 문구의 시간 점수와 기설정된 기준 시간 점수를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제2 후보 대표 문구를 최종적인 대표 문구로 결정하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 24

제18항에 있어서,

상기 제1 후보 대표 문구 생성부는,

상기 인기 키워드를 포함하는 문서들의 형태소를 분석하여 기준 단어를 결정하고, 결정된 기준 단어의 빈도수에 따라 대표 기준 단어를 결정하는 기준 단어 결정부; 및

상기 대표 기준 단어의 이전 또는 이후에 연속하는 단어와 상기 대표 기준 단어를 조합하여 상기 대표 기준 단어를 확장하고, 확장된 대표 기준 단어를 이용하여 상기 제1 후보 대표 문구를 생성하는 기준 단어 확장부

를 포함하는 대표 문구 제공 시스템.

청구항 25

제24항에 있어서,

상기 기준 단어 확장부는,

상기 인기 키워드를 포함하는 문서들에서 상기 대표 기준 단어의 이전 또는 이후에 연속하는 단어가 포함되는 조건부 확률을 계산하고, 상기 조건부 확률에 기초하여 상기 대표 기준 단어를 확장하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 26

제24항에 있어서,

상기 기준 단어 결정부는,

상기 인기 키워드를 포함하는 문서들의 형태소를 분석하여 상기 기준 단어의 빈도수를 카운트하고, 카운트된 빈도수에 기초하여 상기 대표 기준 단어를 결정하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 27

제18항에 있어서,

상기 인기 키워드를 포함하는 문서들을 수집하는 문서 수집부; 및

수집된 문서들에 대해 클러스터(Cluster)를 수행하는 클러스터 수행부

를 더 포함하고,

상기 제1 후보 대표 문구 생성부는,

상기 클러스터에 속하는 문서들을 기초로 대표 기준 단어를 확장하여 상기 제1 후보 대표 문구를 생성하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 28

삭제

청구항 29

제18항에 있어서,

상기 방송 프로그램 결정부는,

상기 방송 프로그램이 복수 개인 경우, 복수의 방송 프로그램 중에서 방송 프로그램의 방송 시작 시간 및 방송 종료 시간이 상기 인기 키워드가 노출된 시간과 가장 인접한 어느 하나의 방송 프로그램을 선택하고,

상기 제2 후보 대표 문구 생성부는,

선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 것을 특징으로 하는 대표 문구 제공 시스템.

청구항 30

제18항에 있어서,

상기 제2 후보 대표 문구 생성부는,

상기 인기 키워드가 여러 단어로 조합된 형태인 경우, 상기 인기 키워드를 형태소 분석하여 복수의 단어들로 분리하고, 상기 인기 키워드를 포함하는 방송 프로그램 중에서 분리된 복수의 단어들에 기초하여 어느 하나의 방송 프로그램을 선택하는 방송 프로그램 결정부; 및

선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 생성부

를 포함하는 대표 문구 제공 시스템.

청구항 31

제18항에 있어서,

상기 제2 후보 대표 문구 생성부는,

상기 인기 키워드를 포함하는 방송 프로그램에 매칭 점수, 시간 점수, 및 방송국 가중치를 부가하고, 상기 매칭 점수, 시간 점수, 방송국 가중치 및 방송 프로그램의 시청률 중 적어도 하나를 이용하여 상기 방송 프로그램에 대한 최종 매칭 점수를 계산하는 매칭 점수 계산부;

상기 인기 키워드를 포함하는 방송 프로그램 중에서 상기 최종 매칭 점수를 기초로 어느 하나의 방송 프로그램을 선택하는 방송 프로그램 결정부; 및

선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 생성부

를 포함하는 대표 문구 제공 시스템.

청구항 32

제18항에 있어서,

상기 최종적인 대표 문구와 상기 인기 키워드를 결합하여 제공하는 대표 문구 제공부

를 더 포함하는 대표 문구 제공 시스템.

명세서

기술분야

본 발명은 실시간 인기 키워드에 대한 대표 문구를 제공하는 방법 및 시스템에 관한 것으로, 보다 구체적으로 사이트에 노출된 인기 키워드가 실시간으로 급상승한 원인인 대표 문구를 문서와 방송 데이터에 기초하여 추출하는 방법 및 시스템에 관한 것이다.

[0001]

배경 기술

- [0002] 일반적으로, 키워드 검색을 제공하는 대부분의 웹 페이지는 현재 시점에서 실시간으로 검색 요청이 급증하는 검색어인 인기 키워드를 노출한다. 이때, 웹 페이지 상에는 선택된 인기 키워드와 인기 키워드의 순위가 노출된다.
- [0003] 특히, 인기 키워드는 웹 페이지 상에 단순히 연예인 이름이거나, 국제 또는 국내 단체명으로 노출된다. 이에 따라, 인기 키워드가 어떤 내용을 포함하고 있는지 알고자 하는 경우, 사용자는 마우스 클릭 등을 이용하여 특정 인기 키워드를 선택할 수 있다. 그러면, 웹 페이지에는 선택된 인기 키워드와 관련된 기사, 문서 등이 노출될 수 있다.
- [0004] 이처럼, 인기 키워드에 대한 구체적인 내용을 알고 싶은 경우뿐만 아니라, 인기 키워드가 된 원인 등과 같이 개략적인 내용을 알고자 하는 경우에도 사용자는 인기 키워드를 선택하고, 선택된 인기 키워드에 대한 다수의 기사들을 모두 읽어야 하는 번거로움이 존재한다.
- [0005] 또한, 대부분의 사용자들은 텔레비전이나 라디오를 청취하다가 출연진의 정보가 궁금하여 검색을 수행하거나, 특정 방송 프로그램이 언제 시작되는지, 혹은 오늘 방송하는 특정 방송 프로그램의 내용이 궁금해서 검색을 수행한다. 이에 따라, 실시간 인기 키워드로는 연예인 이름이나 방송 프로그램 제목이 대부분이다.
- [0006] 따라서, 실시간 인기 키워드에 해당하는 방송 프로그램을 제공하여 실시간 인기 키워드가 된 원인을 사용자에게 알려줄 수 있는 방안이 필요하다.
- [0007] 또한, 사용자가 인기 키워드 및 인기 키워드와 관련된 기사들을 단계적으로 선택하지 않더라도 실시간으로 인기 키워드가 된 원인을 제공할 수 있는 방안이 필요하다.

발명의 내용

해결하려는 과제

- [0008] 본 발명은 문서를 이용하여 실시간으로 인기 키워드가 노출된 원인과 관련된 제1 후보 대표 문구를 제공함으로써 인기 키워드로 제시된 이유에 대한 궁금증을 해결할 수 있는 방법 및 시스템을 제공한다.
- [0009] 본 발명은 방송 데이터를 이용하여 실시간으로 인기 키워드가 노출된 원인과 관련된 제2 후보 대표 문구를 제공함으로써 인기 키워드로 제시된 이유에 대한 궁금증을 해결할 수 있는 방법 및 시스템을 제공한다.
- [0010] 본 발명은 제1 후보 대표 문구 및 제2 후보 대표 문구에 대해 인기도와 방송 적합도를 고려하여 인기 키워드에 대한 최종적인 대표 문구를 결정함으로써 인기 키워드가 노출된 원인을 보다 정확하게 제공할 수 있는 방법 및 시스템을 제공한다.

과제의 해결 수단

- [0011] 본 발명의 일실시예에 따른 대표 문구 제공 방법은, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계, 및 상기 제1 후보 대표 문구 및 상기 제2 후보 대표 문구를 이용하여 상기 인기 키워드에 대한 최종적인 대표 문구를 결정하는 단계를 포함할 수 있다.
- [0012] 또한, 상기 최종적인 대표 문구를 결정하는 단계는, 상기 제1 후보 대표 문구와 상기 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정할 수 있다.
- [0013] 또한, 상기 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 단계를 더 포함할 수 있다.
- [0014] 또한, 상기 대표 문구를 결정하는 단계는, 상기 문서의 인기도와 기설정된 기준 인기도를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제1 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다.
- [0015] 또한, 상기 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 단계를 더 포함할 수 있다.
- [0016] 또한, 상기 대표 문구를 결정하는 단계는, 상기 제2 후보 대표 문구의 시간 점수와 기설정된 기준 시간 점수를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제2 후보 대표 문구를 최종적인 대표

문구로 결정할 수 있다.

- [0017] 또한, 상기 제1 후보 대표 문구를 생성하는 단계는, 상기 인기 키워드를 포함하는 문서들의 기준 단어를 결정하는 단계, 상기 기준 단어 중에서 대표 기준 단어를 결정하는 단계, 대표 기준 단어의 이전 또는 이후에 연속하는 단어와 상기 대표 기준 단어를 조합하여 상기 대표 기준 단어를 확장하는 단계, 및 확장된 대표 기준 단어를 이용하여 상기 제1 후보 대표 문구를 생성하는 단계를 포함할 수 있다.
- [0018] 또한, 상기 대표 기준 단어를 확장하는 단계는, 상기 인기 키워드를 포함하는 문서들에서 상기 대표 기준 단어의 이전 또는 이후에 연속하는 단어가 포함되는 조건부 확률을 계산하는 단계, 및 상기 조건부 확률에 기초하여 상기 대표 기준 단어를 확장하는 단계를 포함할 수 있다.
- [0019] 또한, 상기 대표 기준 단어를 결정하는 단계는, 상기 인기 키워드를 포함하는 문서들의 형태소를 분석하여 상기 기준 단어의 빈도수를 카운트하는 단계, 및 카운트된 빈도수에 기초하여 상기 대표 기준 단어를 결정하는 단계를 포함할 수 있다.
- [0020] 또한, 상기 인기 키워드를 포함하는 문서들을 수집하는 단계, 및 수집된 문서들에 대해 클러스터(Cluster)를 수행하는 단계를 더 포함할 수 있다. 그러면, 상기 제1 후보 대표 문구를 생성하는 단계는, 상기 클러스터에 속하는 문서들을 기초로 대표 기준 단어를 확장하여 상기 제1 후보 대표 문구를 생성할 수 있다.
- [0021] 또한, 상기 제1 후보 대표 문구를 생성하는 단계는, 상기 수집된 문서들에 가중치를 부가하는 단계, 및 상기 가중치를 이용하여 클러스터의 노출 우선 순위를 결정하는 단계를 더 포함할 수 있다.
- [0022] 또한, 상기 제2 후보 대표 문구를 생성하는 단계는, 상기 방송 데이터에서 상기 인기 키워드를 포함하는 방송 프로그램을 결정하는 단계, 및 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계를 포함할 수 있다.
- [0023] 또한, 상기 제2 후보 대표 문구를 생성하는 단계는, 상기 인기 키워드가 여러 단어로 조합된 형태인 경우, 상기 인기 키워드를 형태소 분석하여 복수의 단어들로 분리하는 단계, 상기 인기 키워드를 포함하는 방송 프로그램 중에서 분리된 복수의 단어들에 기초하여 어느 하나의 방송 프로그램을 선택하는 단계, 및 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계를 포함할 수 있다.
- [0024] 또한, 상기 제2 후보 대표 문구를 생성하는 단계는, 상기 인기 키워드를 포함하는 방송 프로그램에 매칭 점수, 시간 점수, 및 방송국 가중치를 부가하는 단계, 상기 매칭 점수, 시간 점수, 방송국 가중치 및 방송 프로그램의 시청률 중 적어도 하나를 이용하여 상기 방송 프로그램에 대한 최종 매칭 점수를 계산하는 단계, 상기 인기 키워드를 포함하는 방송 프로그램 중에서 상기 최종 매칭 점수를 기초로 어느 하나의 방송 프로그램을 선택하는 단계, 및 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 단계를 포함할 수 있다.
- [0025] 또한, 상기 최종적인 대표 문구와 상기 인기 키워드를 결합하여 제공하는 단계를 더 포함할 수 있다.
- [0026] 본 발명의 일실시예에 따른 대표 문구 제공 시스템은, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 제1 후보 대표 문구 생성부, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 제2 후보 대표 문구 생성부, 및 상기 제1 후보 대표 문구 및 상기 제2 후보 대표 문구를 이용하여 상기 인기 키워드에 대한 최종적인 대표 문구를 결정하는 대표 문구 결정부를 포함할 수 있다.
- [0027] 또한, 상기 대표 문구 결정부는, 상기 제1 후보 대표 문구와 상기 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정할 수 있다.
- [0028] 또한, 상기 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 인기도 계산부를 더 포함할 수 있다. 그러면, 상기 대표 문구 결정부는, 상기 문서의 인기도와 기설정된 기준 인기도를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 제1 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다.
- [0029] 또한, 상기 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 시간 점수 계산부를 더 포함할 수 있다. 그러면, 상기 대표 문구 결정부는, 상기 제2 후보 대표 문구의 시간 점수와 기설정된 기준 시간 점수를 이용하여 상기 인기 키워드에 대한 적합도가 더 높은 것으로 결정된 상기 제2 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다.

- [0030] 또한, 상기 제1 후보 대표 문구 생성부는, 상기 인기 키워드를 포함하는 문서들의 형태소를 분석하여 기준 단어를 결정하고, 결정된 기준 단어의 빈도수에 따라 대표 기준 단어를 결정하는 기준 단어 결정부, 및 상기 대표 기준 단어의 이전 또는 이후에 연속하는 단어와 상기 대표 기준 단어를 조합하여 상기 대표 기준 단어를 확장하고, 확장된 대표 기준 단어를 이용하여 상기 제2 후보 대표 문구를 생성하는 기준 단어 확장부를 포함할 수 있다.
- [0031] 또한, 상기 인기 키워드를 포함하는 문서들을 수집하는 문서 수집부, 및 수집된 문서들에 대해 클러스터(Cluster)를 수행하는 클러스터 수행부를 더 포함할 수 있다.
- [0032] 또한, 상기 제2 후보 대표 문구 생성부는, 상기 방송 데이터에서 상기 인기 키워드를 포함하는 방송 프로그램을 결정하는 방송 프로그램 결정부, 및 결정된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 생성부를 포함할 수 있다.
- [0033] 또한, 상기 제2 후보 대표 문구 생성부는, 상기 인기 키워드가 여러 단어로 조합된 형태인 경우, 상기 인기 키워드를 형태소 분석하여 복수의 단어들로 분리하고, 상기 인기 키워드를 포함하는 방송 프로그램 중에서 분리된 복수의 단어들에 기초하여 어느 하나의 방송 프로그램을 선택하는 방송 프로그램 결정부, 및 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 생성부를 포함할 수 있다.
- [0034] 또한, 상기 제2 후보 대표 문구 생성부는, 상기 인기 키워드를 포함하는 방송 프로그램에 매칭 점수, 시간 점수, 및 방송국 가중치를 부가하고, 상기 매칭 점수, 시간 점수, 방송국 가중치 및 방송 프로그램의 시청률 중 적어도 하나를 이용하여 상기 방송 프로그램에 대한 최종 매칭 점수를 계산하는 매칭 점수 계산부, 상기 인기 키워드를 포함하는 방송 프로그램 중에서 상기 최종 매칭 점수를 기초로 어느 하나의 방송 프로그램을 선택하는 방송 프로그램 결정부, 및 선택된 방송 프로그램의 방송 시간을 이용하여 상기 제2 후보 대표 문구를 생성하는 생성부를 포함할 수 있다.
- [0035] 또한, 상기 대표 문구와 상기 인기 키워드를 결합하여 제공하는 대표 문구 제공부를 포함할 수 있다.

발명의 효과

- [0036] 본 발명의 일실시예에 따르면, 문서를 이용하여 실시간으로 인기 키워드가 노출된 원인과 관련된 제1 후보 대표 문구를 제공함으로써 인기 키워드로 제시된 이유에 대한 궁금증이 해결될 수 있다.
- [0037] 본 발명의 일실시예에 따르면, 방송 데이터를 이용하여 실시간으로 인기 키워드가 노출된 원인과 관련된 제2 후보 대표 문구를 제공함으로써 인기 키워드로 제시된 이유에 대한 궁금증이 해결될 수 있다.
- [0038] 본 발명의 일실시예에 따르면, 제1 후보 대표 문구 및 제2 후보 대표 문구에 대해 인기도와 방송 적합도를 고려하여 인기 키워드에 대한 최종적인 대표 문구를 결정함으로써 인기 키워드가 노출된 원인이 보다 정확하게 제공될 수 있다.

도면의 간단한 설명

- [0039] 도 1은 본 발명의 일실시예에 따른 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 도면이다.
- 도 2는 본 발명의 일실시예에 따른 대표 문구 제공 시스템의 구성을 도시한 블록도이다.
- 도 3는 본 발명의 일실시예에 따른 도 2의 대표 문구 제공 시스템의 동작을 설명하기 위해 제공되는 흐름도이다.
- 도 4는 본 발명의 일실시예에 따른 제1 후보 대표 문구 생성부의 세부 구성을 도시한 블록도이다.
- 도 5는 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출하는 화면을 도시한 도면이다.
- 도 6은 본 발명의 일 실시예에 따른 제1 후보 대표 문구를 생성하는 방법을 설명하기 위해 제공되는 흐름도이다.
- 도 7은 본 발명의 일실시예에 따른 제2 후보 대표 문구 생성부의 세부 구성을 도시한 블록도이다.
- 도 8은 제2 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출하는 화면을 도시한 도면이다.
- 도 9는 본 발명의 일실시예에 따라 방송 시간을 이용하여 제2 후보 대표 문구를 제공하는 과정을 설명하기 위해 제공되는 흐름도이다.

도 10은 여러 단어가 조합된 형태를 갖는 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 흐름도이다.

도 11은 본 발명의 일실시예에 따라 매칭 점수를 이용하여 제2 후보 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0040] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.
- [0041] 도 1은 본 발명의 일실시예에 따른 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 도면이다.
- [0042] 도 1을 따르면, 대표 문구 제공 시스템(30)은 실시간 인기 키워드(40)와 관련된 문서들(10) 및 실시간 인기 키워드(40)와 관련된 방송 데이터들(20)을 수집할 수 있다. 여기서, 대표 문구 제공 시스템(30)은 인기 키워드를 포함하는 텍스트 문서(document), 인기 키워드와 관련된 기사(article), 및 상기 인기 키워드의 정지 영상과 같은 문서들(10)을 포털 사이트, 뉴스 사이트, 블로그, 카페, 마이크로 블로그와 같은 커뮤니티 등으로부터 수집할 수 있다.
- [0043] 이때, 대표 문구 제공 시스템(30)은 수집된 문서들의 형태소를 분석하여 인기 키워드에 대한 제1 후보 대표 문구를 생성할 수 있다. 여기서, 제1 후보 대표 문구는 웹 페이지에 노출하고자 하는 인기 키워드가 실시간 인기 키워드가 된 원인을 나타내는 문구로서, 문서를 기초로 생성된 문구를 의미한다.
- [0044] 그리고, 방송 데이터를 수집하는 경우, 대표 문구 제공 시스템(30)은 기설정된 시간마다 주기적으로 방송 데이터들을 방송국으로부터 수집할 수 있다. 여기서, 대표 문구 제공 시스템(30)은 공중파 TV 방송, 케이블 TV 방송, 라디오 방송, 인터넷 방송, 위성 TV 방송과 관련된 방송국 사이트, 뉴스 사이트 등으로부터 방송 편성표, 전자 프로그램 가이드와 같은 방송 데이터들(20)을 수집할 수 있다.
- [0045] 이때, 대표 문구 제공 시스템(30)은 수집된 방송 데이터에서 인기 키워드를 포함하는 방송 프로그램을 결정할 수 있다. 그러면, 대표 문구 제공 시스템(30)은 결정된 방송 프로그램을 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성할 수 있다. 여기서, 제2 후보 대표 문구는 웹 페이지에 노출하고자 하는 인기 키워드가 실시간 인기 키워드가 된 원인을 나타내는 문구로서, 방송 데이터를 기초로 생성된 문구를 의미한다.
- [0046] 일례로, 대표 문구 제공 시스템(30)은 결정된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다. 여기서, 방송 시간은 방송 프로그램의 방송 시작 시간 및 방송 종료 시간을 포함할 수 있다.
- [0047] 그리고, 대표 문구 제공 시스템(30)은 제1 후보 대표 문구와 제2 후보 대표 문구를 이용하여 인기 키워드에 대한 대표 문구를 결정할 수 있다. 이때, 대표 문구 제공 시스템(30)은 제1 후보 대표 문구와 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정할 수 있다. 여기서, 제1 후보 대표 문구와 제2 후보 대표 문구는 인기 키워드가 실시간으로 급상승하게 된 원인인 대표 문구가 될 가능성이 높은 후보군을 의미한다.
- [0048] 일례로, 대표 문구 제공 시스템(30)은 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도에 기초하여 제1 후보 대표 문구의 적합도를 결정할 수 있다. 그리고, 대표 문구 제공 시스템(30)은 제1 후보 대표 문구의 적합도에 기초하여 제1 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다.
- [0049] 이때, 제1 후보 대표 문구가 적합하지 않은 것으로 결정된 경우, 대표 문구 제공 시스템(30)은 제2 후보 대표 문구의 시간 점수에 기초하여 제2 후보 대표 문구의 적합도를 결정할 수 있다. 이어, 대표 문구 제공 시스템(30)은 제2 후보 대표 문구의 적합도에 기초하여 제2 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다.
- [0050] 그러면, 대표 문구 제공 시스템(30)은 최종적인 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0051] 도 2는 본 발명의 일실시예에 따른 대표 문구 제공 시스템의 구성을 도시한 블록도이다.
- [0052] 도 2에 따르면, 대표 문구 제공 시스템(200)은 제1 후보 대표 문구 생성부(210), 제2 후보 대표 문구 생성부(220), 인기도 계산부(230), 대표 문구 결정부(240), 시간 점수 계산부(250), 및 대표 문구 제공부(260)를 포함할 수 있다.
- [0053] 먼저, 제1 후보 대표 문구 생성부(210)는 문서(110)를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성할 수 있다. 여기서, 제1 후보 대표 문구는 인기 키워드가 제시된 근거를 나타내는 문구로서, 문서(110)에 기초하

여 추출될 수 있다.

- [0054] 일례로, 제1 후보 대표 문구 생성부(210)는 실시간 인기 키워드와 관련된 문서들의 형태소를 분석하여 제1 후보 대표 문구를 생성할 수 있다. 여기서, 제1 후보 대표 문구를 생성하는 자세한 과정은 도 4 내지 도 6을 참조하여 후술하기로 한다.
- [0055] 이어, 제2 후보 대표 문구 생성부(220)는 방송 데이터(120)를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성할 수 있다. 여기서, 제2 후보 대표 문구는 인기 키워드가 제시된 근거를 나타내는 문구로서, 방송 데이터(120)에 기초하여 추출될 수 있다.
- [0056] 일례로, 제2 후보 대표 문구 생성부(220)는 방송 데이터(120)에서 인기 키워드를 포함하는 방송 프로그램을 결정하고, 결정된 방송 프로그램의 방송 시간에 기초하여 제2 후보 대표 문구를 생성할 수 있다. 여기서, 제2 후보 대표 문구를 생성하는 자세한 과정은 도 7 내지 도 11을 참조하여 후술하기로 한다.
- [0057] 인기도 계산부(230)는 인기 키워드와 관련하여 수집된 문서들 중에서 제1 후보 대표 문구를 생성하기 위해 이용된 문서를 추출할 수 있다. 그리고, 인기도 계산부(230)는 추출된 문서의 인기도를 계산할 수 있다.
- [0058] 일례로, 인기도 계산부(230)는 추출된 문서의 클릭수 및 노출수를 조합한 검색 대비 클릭 비율(CTR)을 이용하여 문서의 인기도를 계산할 수 있다. 이때, 인기도 계산부(230)는 추출된 문서의 클릭 로그 자료를 이용하여 클릭수를 계산할 수 있다. 그리고, 인기도 계산부(230)는 추출된 문서의 검색 로그 자료를 이용하여 문서의 노출수를 계산할 수 있다. 여기서, 클릭 로그 자료 및 검색 로그 자료는 최근 30분 이내로 한정될 수 있다.
- [0059] 그러면, 대표 문구 결정부(240)는 제1 후보 대표 문구의 적합도 및 제2 후보 대표 문구의 적합도에 기초하여 제1 후보 대표 문구 및 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정할 수 있다.
- [0060] 일례로, 적합도 결정에 문서의 인기도와 기설정된 기준 인기도를 이용하고, 추출된 문서가 한 개인 경우, 인기도 계산부(230)는 하나의 문서 인기도를 최종 인기도로 계산할 수 있다. 이때, 추출된 문서가 복수 개인 경우, 인기도 계산부(230)는 복수의 문서 인기도를 조합하여 최종 인기도를 계산할 수 있다.
- [0061] 그러면, 대표 문구 결정부(240)는 최종 인기도와 기준 인기도를 이용하여 제1 후보 대표 문구의 적합도를 결정할 수 있다.
- [0062] 일례로, 기준 인기도가 α 로 기설정된 경우, 대표 문구 결정부(240)는 문서의 인기도와 기준 인기도 α 를 비교할 수 있다. 그리고, 최종 인기도가 기준 인기도 α 이상인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구가 제2 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은 것으로 결정할 수 있다. 그러면, 대표 문구 결정부(240)는 제1 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다. 즉, 최종 인기도가 기준 인기도 α 이상인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구를 생성하는 데 이용한 문서가 인기도를 획득한 것으로 결정할 수 있다. 여기서, 기준 인기도 α 는 변경될 수 있다.
- [0063] 다른 예로, 최종 인기도가 기준 인기도 α 미만인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구가 인기 키워드에 대해 최종적인 대표 문구로 적합하지 않은 것으로 결정할 수 있다.
- [0064] 그러면, 시간 점수 계산부(250)는 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 이때, 시간 점수 계산부(250)는 1)제2 후보 대표 문구를 생성하는 데 이용된 방송 프로그램의 방송 시간, 및 2)인기 키워드의 노출 시간에 기초하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 여기서, 방송 시간은, 방송 시작 시간 및 방송 종료 시간을 포함할 수 있다.
- [0065] 일례로, 노출 시간이 방송 시작 시간 이상이고, 방송 종료 시간 이하인 경우, 시간 점수 계산부(250)는 제2 후보 대표 문구의 시간 점수를 기설정된 기준 점수 β 로 계산할 수 있다. 즉, 방송 프로그램이 방송되는 시간에 인기 키워드가 노출된 경우, 시간 점수 계산부(250)는 제2 후보 대표 문구의 시간 점수를 기준 점수 β 로 계산할 수 있다.
- [0066] 다른 예로, 노출 시간이 방송 시작 시간 이전이거나, 또는 노출 시간이 방송 종료 시간 이후인 경우, 시간 점수 계산부(250)는 기준 점수 β 에서 오차 단위 γ 마다 오차 점수 δ 를 감산하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 여기서, 오차 단위 γ 및 오차 점수 δ 는 기설정되며, 변경될 수 있다.
- [0067] 이때, 노출 시간이 방송 시작 시간 이전인 경우, 시간 점수 계산부(250)는 노출 시간과 방송 시작 시간 간의 차이값을 계산할 수 있다. 그리고, 시간 점수 계산부(250)는 차이값을 기준으로 오차 단위마다 기준 점수에서 오

차 점수를 감산하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 예를 들어, 차이값이 20분이고, 오차 단위 γ 가 10분, 오차 점수 δ 가 5로 기설정된 경우, 시간 점수 계산부(250)는 차이값을 기준으로 계산된 오차 값 2와 오차 점수 5의 곱을 기준 점수 β 에서 감산($\beta-(2 \times 5)$)할 수 있다. 그리고, 시간 점수 계산부(250)는 감산 결과인 $\beta-10$ 을 제2 후보 대표 문구의 시간 점수로 계산할 수 있다.

[0068] 그러면, 대표 문구 결정부(240)는 제2 후보 대표 문구의 시간 점수와 기설정된 기준 시간 점수 τ 를 이용하여 제2 후보 대표 문구가 제1 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은지 여부를 결정할 수 있다.

[0069] 일례로, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하가 아닌 경우, 대표 문구 결정부(240)는 제2 후보 대표 문구가 인기 키워드에 대한 적합도가 더 높은 것으로 결정할 수 있다. 이에 따라, 대표 문구 결정부(240)는 제2 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다. 이때, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하인 경우, 대표 문구 결정부(240)는 제2 후보 대표 문구가 인기 키워드에 대해 최종적인 대표 문구로 적합하지 않다고 결정할 수 있다.

[0070] 대표 문구 제공부(260)는 인기 키워드에 대해 최종적으로 결정된 대표 문구와 인기 키워드를 결합하여 제공할 수 있다. 그러면, 사용자는 인기 키워드의 인기도가 실시간으로 급상승한 이유를 최종적으로 결정된 대표 문구를 통해 확인할 수 있다.

[0071] 일례로, 제1 후보 대표 문구가 "홍콩 여배우 BBB와 비밀 데이트"이고, 제1 후보 대표 문구를 인기 키워드에 대해 최종적인 대표 문구로 결정한 경우, 대표 문구 제공부(260)는 도 5와 같이, 대표 문구 "홍콩 여배우 BBB와 비밀 데이트"(521)와 인기 키워드 "AAA"(510)를 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부(260)는 인기 키워드 "AAA"에 대한 사진(522)을 인기 키워드(510) 및 대표 문구(521)와 결합하여 웹 페이지에 노출할 수도 있다.

[0072] 다른 예로, 제2 후보 대표 문구가 "22:00부터 방송 시작"이고, 제2 후보 대표 문구를 인기 키워드에 대해 최종적인 대표 문구로 결정한 경우, 대표 문구 제공부(260)는 도 8과 같이, 대표 문구 제공부(260)는 인기 키워드 "XXX"(810)와 대표 문구 "22:00부터 방송 시작"(821)을 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부(260)는 인기 키워드 "XXX"의 방송 프로그램에 대한 화보 사진(822)을 인기 키워드(810) 및 대표 문구(821)와 결합하여 웹 페이지에 노출할 수도 있다.

[0073] 이상의 도 2에서는, 대표 문구 제공 시스템(200)이 제1 후보 대표 문구 생성부(210), 제2 후보 대표 문구 생성부(220), 인기도 계산부(230), 대표 문구 결정부(240), 시간 점수 계산부(250), 및 대표 문구 제공부(260)를 포함하는 것으로 설명하였으나, 이는 실시예에 해당되며, 본 명세서에서 사용되는 대표 문구 제공 시스템의 구성은 사용자, 운용자의 의도 또는 본 발명이 속하는 분야의 관례 등에 따라 달라질 수 있다. 따라서, 본 대표 문구 제공 시스템의 구성이 도 2에 의해 제한되거나 한정되는 것은 아니다.

[0074] 일례로, 본 발명의 일실시예에 따른 대표 문구 제공 시스템은, 제1 후보 대표 문구 생성부, 제2 후보 대표 문구 생성부, 및 대표 문구 결정부로 구성될 수 있다. 이때, 대표 문구 제공 시스템은 대표 문구 제공부를 더 포함할 수도 있다.

[0075] 다른 예로, 본 발명의 일실시예에 따른 대표 문구 제공 시스템은, 제1 후보 대표 문구 생성부, 제2 후보 대표 문구 생성부, 인기도 계산부, 및 대표 문구 결정부로 구성될 수 있다. 이때, 대표 문구 제공 시스템은 대표 문구 제공부를 더 포함할 수도 있다.

[0076] 또 다른 예로, 본 발명의 일실시예에 따른 대표 문구 제공 시스템은, 대표 문구 제공 시스템은, 제1 후보 대표 문구 생성부, 제2 후보 대표 문구 생성부, 시간 점수 계산부, 및 대표 문구 결정부로 구성될 수도 있다. 이때, 대표 문구 제공 시스템은 대표 문구 제공부를 더 포함할 수도 있다.

[0077] 도 3는 본 발명의 일실시예에 따른 도 2의 대표 문구 제공 시스템의 동작을 설명하기 위해 제공되는 흐름도이다.

[0078] 먼저, 310 단계에서, 제1 후보 대표 문구 생성부(210)는 문서를 기초로 제1 후보 대표 문구를 생성할 수 있다. 이때, 제1 후보 대표 문구 생성부(210)는 실시간 인기 키워드와 관련된 문서들의 형태소를 분석하여 제1 후보 대표 문구를 생성할 수 있다.

[0079] 이어, 320 단계에서, 제2 후보 대표 문구 생성부(220)는 방송 데이터를 기초로 제2 후보 대표 문구를 생성할 수 있다. 일례로, 제2 후보 대표 문구 생성부(220)는 방송 데이터에 포함된 복수의 방송 프로그램들 중에서 인기

키워드를 포함하는 방송 프로그램을 결정할 수 있다. 그리고, 제2 후보 대표 문구 생성부(220)는 결정된 방송 프로그램의 방송 시간에 기초하여 제2 후보 대표 문구를 생성할 수 있다.

- [0080] 그리고, 330 단계에서, 인기도 계산부(230)는 제1 후보 대표 문구를 생성하는데 이용된 하나 이상의 문서의 인기도를 계산할 수 있다.
- [0081] 이때, 인기도 계산부(230)는 인기 키워드와 관련하여 수집된 문서들 중에서 제1 후보 대표 문구를 생성하기 위해 이용된 문서를 추출할 수 있다. 그리고, 인기도 계산부(230)는 추출된 문서의 클릭수 및 노출수를 조합한 검색 대비 클릭 비율(CTR)을 이용하여 문서의 인기도를 계산할 수 있다. 이때, 인기도 계산부(230)는 추출된 문서의 클릭 로그 자료를 이용하여 클릭수를 계산할 수 있다. 그리고, 인기도 계산부(230)는 추출된 문서의 검색 로그 자료를 이용하여 문서의 노출수를 계산할 수 있다. 여기서, 클릭 로그 자료 및 검색 로그 자료는 최근 30분 이내에 생성된 자료일 수 있으나, 본 발명은 상기 언급된 시간으로 한정되는 것은 아니다.
- [0082] 이처럼, 제1 후보 대표 문구를 생성하는 데 이용된 문서가 한 개인 경우, 인기도 계산부(230)는 한 개의 문서 인기도를 계산할 수 있다. 그리고, 제1 후보 대표 문구를 생성하는 데 이용된 문서가 복수 개인 경우, 인기도 계산부(230)는 복수의 문서에 대응하는 문서 인기도를 각각 계산할 수 있다.
- [0083] 이에 따라, 335 단계에서, 인기도 계산부(230)는 복수의 문서 인기도를 조합하여 최종 인기도를 계산할 수 있다. 이때, 계산된 문서 인기도가 한 개인 경우, 인기도 계산부(230)는 하나의 문서 인기도를 최종 인기도로 계산할 수 있다.
- [0084] 이어, 대표 문구 결정부(240)는 제1 후보 대표 문구의 적합도 및 제2 후보 대표 문구의 적합도에 기초하여 제1 후보 대표 문구 및 제2 후보 대표 문구 중 인기 키워드에 대한 적합도가 더 높은 것을 최종적인 대표 문구로 결정할 수 있다.
- [0085] 그리고, 340 단계에서, 대표 문구 결정부(240)는 최종 인기도와 기설정된 기준 인기도를 이용하여 제1 후보 대표 문구가 인기 키워드에 대한 적합도가 더 높은지 여부를 결정할 수 있다.
- [0086] 이때, 최종 인기도가 기준 인기도 이상인 경우, 350 단계에서, 대표 문구 결정부(240)는 제1 후보 대표 문구가 제2 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은 것으로 결정할 수 있다. 그리고, 대표 문구 결정부(240)는 제1 후보 대표 문구를 최종적인 대표 문구로 결정할 수 있다. 즉, 최종 인기도가 기준 인기도 α 이상인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구를 생성하는 데 이용한 문서가 인기도를 획득한 것으로 결정할 수 있다.
- [0087] 그리고, 최종 인기도가 기준 인기도 미만인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구가 인기 키워드에 대한 적합도가 더 높지 않은 것으로 결정할 수 있다. 즉, 대표 문구 결정부(240)는 제1 후보 대표 문구가 최종적인 대표 문구로 적합하지 않은 것으로 결정할 수 있다.
- [0088] 그러면, 360 단계에서, 시간 점수 계산부(250)는 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 이때, 시간 점수 계산부(250)는 제2 후보 대표 문구를 생성하는 데 이용된 방송 프로그램의 방송 시간, 및 인기 키워드의 노출 시간에 기초하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 여기서, 방송 시간은, 방송 시작 시간 및 방송 종료 시간을 포함할 수 있다.
- [0089] 일례로, 노출 시간이 방송 시작 시간 이상이고, 방송 종료 시간 이하인 경우, 시간 점수 계산부(250)는 제2 후보 대표 문구의 시간 점수를 기설정된 기준 점수 β 로 계산할 수 있다.
- [0090] 다른 예로, 노출 시간이 방송 시작 시간 이전이거나, 또는 노출 시간이 방송 종료 시간 이후인 경우, 시간 점수 계산부(250)는 기준 점수 β 에서 오차 단위 γ 마다 오차 점수 δ 를 감산하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다.
- [0091] 이때, 노출 시간이 방송 시작 시간 이전인 경우, 시간 점수 계산부(250)는 노출 시간과 방송 시작 시간 간의 차이값을 계산할 수 있다. 그리고, 시간 점수 계산부(250)는 차이값을 기준으로 오차 단위마다 기준 점수에서 오차 점수를 감산하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다. 예를 들어, 차이값이 20분이고, 오차 단위 γ 가 10분, 오차 점수 δ 가 5로 기설정된 경우, 시간 점수 계산부(250)는 차이값을 기준으로 계산된 오차 값 2와 오차 점수 5의 곱을 기준 점수 β 에서 감산($\beta - (2 \times 5)$)할 수 있다. 그리고, 시간 점수 계산부(250)는 감산 결과인 $\beta - 10$ 을 제2 후보 대표 문구의 시간 점수로 계산할 수 있다.
- [0092] 마찬가지로, 노출 시간이 방송 종료 시간 이후인 경우, 시간 점수 계산부(250)는 노출 시간과 방송 종료 시간

간의 차이값을 계산할 수 있다. 그 리고, 시간 점수 계산부(250)는 차이값을 기준으로 오차 단위마다 기준 점수에서 오차 점수를 감산하여 제2 후보 대표 문구의 시간 점수를 계산할 수 있다.

- [0093] 이어, 370 단계에서, 대표 문구 결정부(240)는 시간 점수와 기설정된 기준 시간 점수 τ 를 이용하여 제2 후보 대표 문구가 제1 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은지 여부를 결정할 수 있다.
- [0094] 이때, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하인 경우, 대표 문구 결정부(240)는 제2 후보 대표 문구가 인기 키워드에 대한 적합도가 더 높지 않은 것으로 결정할 수 있다. 즉, 대표 문구 결정부(240)는 제2 후보 대표 문구가 최종적인 대표 문구로 적합하지 않다고 결정할 수 있다.
- [0095] 그리고, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하가 아닌 경우, 380 단계에서, 대표 문구 결정부(240)는 제2 후보 대표 문구가 제1 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은 것으로 결정할 수 있다. 즉, 최종 인기도가 기준 인기도 이상이고, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하가 아닌 경우, 제2 후보 대표 문구가 제1 후보 대표 문구보다 인기 키워드에 대한 적합도가 더 높은 것으로 결정할 수 있다. 이에 따라, 대표 문구 결정부(240)는 제2 후보 대표 문구를 인기 키워드에 대해 최종적인 대표 문구로 결정할 수 있다.
- [0096] 그러면, 390 단계에서, 대표 문구 제공부(260)는 대표 문구와 인기 키워드를 결합하여 제공할 수 있다.
- [0097] 일례로, 제1 후보 대표 문구가 인기 키워드에 대해 최종적인 대표 문구로 결정된 경우, 도 5와 같이, 대표 문구 제공부(260)는 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0098] 다른 예로, 제2 후보 대표 문구가 인기 키워드에 대해 최종적인 대표 문구로 결정된 경우, 도 8과 같이, 대표 문구 제공부(260)는 제2 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0099] 한편, 도 3의 370단계에서, 제2 후보 대표 문구의 시간 점수가 기준 시간 점수 이하인 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구 및 제2 후보 대표 문구 전부 인기 키워드에 대해 최종적인 대표 문구로 결정하지 않을 수 있다. 이처럼, 제1 후보 대표 문구 및 제2 후보 대표 문구 전부 인기 키워드에 대해 최종적인 대표 문구로 결정되지 않으면, 대표 문구 결정부(240)는 제1 후보 대표 문구 및 제2 후보 대표 문구 중 어느 하나를 랜덤하게 선택하여 인기 키워드에 대해 최종적인 대표 문구로 결정할 수 있다. 그러면, 대표 문구 제공부(260)는 랜덤하게 선택된 대표 문구와 인기 키워드를 결합하여 노출할 수 있다.
- [0100] 또한, 제1 후보 대표 문구 및 제2 후보 대표 문구 모두 최종적인 대표 문구로 결정되지 않은 경우, 대표 문구 결정부(240)는 제1 후보 대표 문구 및 제2 후보 대표 문구를 모두 인기 키워드에 대해 최종적인 대표 문구로 결정할 수도 있다. 그러면, 대표 문구 제공부는, 제1 후보 대표 문구 및 제2 후보 대표 문구, 그리고 인기 키워드를 결합하여 노출할 수 있다.
- [0101] 또한, 제1 후보 대표 문구 및 제2 후보 대표 문구 모두 최종적인 대표 문구로 결정되지 않은 경우, 대표 제공부(260)는 대표 문구 없이 인기 키워드만을 노출할 수도 있다.
- [0102] 이상의 도 3에서는, 인기도 계산, 시간 점수 계산을 이용하여 제1 및 제2 후보 대표 문구의 적합도를 결정하고, 결정된 적합도에 기초하여 최종적인 대표 문구를 결정 및 제공하는 구성에 대해 설명하였으나, 이는 실시예에 해당되며, 본 명세서에서 사용되는 대표 문구 제공 방법의 구성은 사용자, 운용자의 의도 또는 본 발명이 속하는 분야의 관례 등에 따라 달라질 수 있다. 따라서, 본 대표 문구 제공 방법의 구성이 도 3에 의해 제한되거나 한정되는 것은 아니다.
- [0103] 일례로, 본 발명의 일실시예에 따른 대표 문구 제공 방법은, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계, 및 제1 후보 대표 문구 및 제2 후보 대표 문구를 이용하여 인기 키워드에 대해 최종적인 대표 문구를 결정하는 단계를 포함할 수 있다. 이 때, 대표 문구 제공 방법은 최종적인 대표 문구와 인기 키워드를 결합하여 제공하는 단계를 더 포함할 수 있다.
- [0104] 다른 예로, 대표 문구 제공 방법은, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계, 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 단계, 제1 후보 대표 문구 및 제2 후보 대표 문구를 이용하여 인기 키워드에 대해 최종적인 대표 문구를 결정하는 단계를 포함할 수 있다. 이 때, 대표 문구 제공 방법은 최종적인 대표 문구와 인기 키워드를 결합하여 제공하는 단계를 더 포함할 수 있다.

[0105] 또 다른 예로, 본 발명의 일실시예에 따른 대표 문구 제공 방법은, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계, 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 단계, 그리고 제1 후보 대표 문구 및 제2 후보 대표 문구를 이용하여 인기 키워드에 대해 최종적인 대표 문구를 결정하는 단계를 포함할 수 있다. 이때, 대표 문구 제공 방법은 최종적인 대표 문구와 인기 키워드를 결합하여 제공하는 단계를 더 포함할 수 있다.

[0106] 또 다른 예로, 문서를 기초로 인기 키워드에 대한 제1 후보 대표 문구를 생성하는 단계, 방송 데이터를 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 단계, 제1 후보 대표 문구를 생성하기 위해 이용된 문서의 인기도를 계산하는 단계, 제2 후보 대표 문구를 생성하기 위해 이용된 방송 데이터의 방송 시간과 인기 키워드의 노출 시간을 이용하여 제2 후보 대표 문구의 시간 점수를 계산하는 단계, 및 제1 후보 대표 문구 및 제2 후보 대표 문구를 이용하여 인기 키워드에 대해 최종적인 대표 문구를 결정하는 단계를 포함할 수 있다. 이때, 대표 문구 제공 방법은 최종적인 대표 문구와 인기 키워드를 결합하여 제공하는 단계를 더 포함할 수 있다.

[0107] 이하에서는, 도 4 내지 도 6을 참조하여 제1 후보 대표 문구를 생성하는 구성, 및 제1 후보 대표 문구가 대표 문구로 결정된 경우, 제1 후보 대표 문구와 인기 키워드를 결합하여 제공하는 구성에 대해 상세히 설명하기로 한다.

[0108] 도 4는 본 발명의 일실시예에 따른 제1 후보 대표 문구 생성부의 세부 구성을 도시한 블록도이다.

[0109] 도 4에 따르면, 도 2의 대표 문구 제공 시스템은, 문서 수집부, 및 클러스터 수행부를 더 포함할 수 있다.

[0110] 먼저, 문서 수집부(410)는 웹 페이지 상에 실시간 급상승 검색어로 노출하고자 하는 인기 키워드와 관련된 문서들을 각종 포털 사이트, 뉴스 사이트, 블로그, 마이크로 블로그 등을 통해 수집할 수 있다.

[0111] 일례로, 문서 수집부(410)는 인기 키워드 AAA에 대한 문서들을 아래의 표 1과 같이 수집할 수 있다.

표 1

문서	
NNN_news#0	AAA, 홍콩, 최고 여배우, 'BBB'와 中서 비밀 데이트?
NNN_news#1	AAA, 홍콩 여배우 BBB와 몰래 데이트?
MMM_news#3	AAA-BBB, 연인? 콘서트 항상 인기 폭발
MMM_news#1	AAA, 홍콩 여배우 BBB와 비밀 데이트?
LLL#0	AAA, BBB 콘서트 영상 인기

[0113] 클러스터 수행부(420)는 대해 N-그램 형태소 분석을 이용하여 수집된 문서들에 대해 클러스터를 수행할 수 있다. 여기서, N-그램은 단어를 N개까지 연이어 붙여서 사용하는 것으로, 1-그램은 단어 하나를 의미하고, 2-그램은 두 개의 단어가 연이어 결합된 형태를 의미한다. 이처럼, N-그램 형태소 분석을 통해 클러스터를 수행함에 따라, 클러스터의 품질이 향상될 수 있다.

[0114] 이때, 1-그램 형태소 분석의 경우, 잘못된 클러스터가 생성될 수 있으나, 5-그램까지 이용하여 형태소 분석을 하는 경우, AAA 키워드와 연관성이 높은 문서들끼리 묶임 확률이 증가한다. 일례로, 클러스터 수행부(420)는 표 1의 NNN_news#0에 대해 5-그램 형태소 분석을 이용하여 클러스터 인풋 데이터(Cluster Input Data)를 생성할 수 있다. 이때, 생성된 클러스터 인풋 데이터는 아래의 표 2와 같을 수 있다.

표 2

문서	클러스터 인풋 데이터
NNN_news#0	AAA AAA홍콩 AAA홍콩최고 AAA홍콩최고여배우 AAA홍콩최고여배우BBB AAA홍콩최고여배우BBB와 AAA홍콩최고여배우BBB와中 홍콩 홍콩최고 홍콩최고여배우 홍콩최고여배우BBB와中 홍콩최고여배우BBB와 中서비밀 최고 여배우 홍콩최고여배우BBB와中서비밀 최고여배우BBB와 中서비밀 여배우 여배우BBB 여배우BBB와 여배우BBB와中 여배우BBB와 中서 여배우BBB와中서비밀 여배우BBB와中서비밀데이트 BBB BBB와 BBB와中 BBB와中서 BBB와中서비밀 BBB와中서비밀데이트 中 中서 中서비밀 中서비밀데이트 서 서비밀 서비밀데이트 비밀 비밀데이트 데이터

[0116] 표 2에서는 문서 NNN_news#0에 대해 생성된 클러스터 인풋 데이터를 설명하고 있으나, 클러스터 수행부(420)는 수집된 모든 문서들에 대해 N-그램 형태소 분석을 이용하여 클러스터 인풋 데이터를 생성할 수 있다.

[0117] 그러면, 클러스터 수행부(420)는 생성된 클러스터 인풋 데이터와 카테고리컬(Categorical) 클러스터링 기법을 이용하여 수집된 모든 문서들에 대해 클러스터를 수행할 수 있다. 이때, 클러스터 수행부(420)는 자카드(Jaccard) 계수를 기반으로 하는 계층적 클러스터링 기법을 이용하여 클러스터를 수행할 수 있다. 일례로, 클러스터 수행부(420)는 표 1에서 수집된 문서들에 대해 상기 계층적 클러스터링 기법을 이용하여 클러스터를 수행함으로써 아래의 표 3 및 표 4와 같은 2개의 클러스터를 생성할 수 있다.

표 3

[0118]

# 클러스터 A	
MMM_news#1	AAA, 홍콩 여배우 BBB와 비밀 데이트?
NNN_news#0	AAA, 홍콩, 최고 여배우, 'BBB'와 中서 비밀 데이트?
NNN_news#1	AAA, 홍콩 여배우 BBB와 몰래 데이트?

표 4

[0119]

# 클러스터 B	
LLL#0	AAA, BBB 콘서트 영상 인기
MMM_news#3	AAA-BBB, 연인? 콘서트 항상 인기 폭발

[0120] 표 3에 따르면, 클러스터 A는 인기 키워드 AAA와 BBB 간의 데이트와 관련된 문서들을 포함하는 클러스터이다. 그리고, 표 4에 따르면, 클러스터 B는 인기 키워드 AAA와 BBB간의 콘서트와 관련된 문서들을 포함하는 클러스터이다.

[0121] 제1 후보 대표 문구 생성부(430)는 생성된 복수의 클러스터에서 각 클러스터를 대표하는 문장 하나를 조합하여 인기 키워드에 대한 제1 후보 대표 문구를 생성할 수 있다. 여기서, 제1 후보 대표 문구 생성부(430)는 기준 단어 결정부(431) 및 기준 단어 확장부(432)를 포함할 수 있다.

[0122] 기준 단어 결정부(431)는 하나의 클러스터에 포함된 모든 문서들의 형태소를 분석하여 하나 이상의 기준 단어를 결정할 수 있다. 일례로, 클러스터 A에 포함된 문서들의 형태소를 분석하는 경우, 기준 단어 결정부(431)는 문서 MMM_news#1, NNN_news#0, 및 NNN_news#1의 형태소를 분석하여 "홍콩", "데이트", "BBB", "여배우", "비밀", "몰래"를 기준 단어로 결정할 수 있다.

[0123] 이때, 기준 단어 결정부(431)는 결정된 기준 단어가 상기 문서들 MMM_news#1, NNN_news#0, 및 NNN_news#1에 포함되는 빈도수를 카운트할 수 있다. 일례로, 클러스터 A에 대해 결정된 기준 단어와 카운트된 기준 단어의 빈도수는 아래의 표 5와 같을 수 있다.

표 5

[0124]

기준 단어	빈도수
홍콩	232
데이트	224
BBB	215
여배우	152
비밀	142
몰래	70

[0125] 기준 단어 결정부(431)는 카운트된 기준 단어의 빈도수가 가장 높은 순으로 2개의 기준 단어를 대표 기준으로 결정할 수 있다. 일례로, 표 5에 따르면, 기준 단어 결정부(431)는 빈도수가 높은 상위 2개의 "홍콩"과 "데이트"를 대표 기준으로 결정할 수 있다.

[0126] 기준 단어 확장부(432)는 결정된 대표 기준 단어와 연속하는 단어가 클러스터에 속하는 문서들에 포함되는 조건

부 확률을 계산할 수 있다. 이때, 기준 단어 확장부(432)는 바이그램(bi-gram) 조건부 확률을 이용하여 대표 기준 단어와 이전 또는 이후에 연속하는 단어가 포함될 조건부 확률을 계산할 수 있다.

[0127] 일례로, 기준 단어 확장부(232)는 클러스터 A에 속하는 문서들에서 대표 기준 단어 "홍콩"과 연속하는 단어 "여배우" 및 "최고"가 포함되는 조건부 확률을 계산할 수 있다. 이때, 기준 단어 확장부(432)는 아래의 수학적 식 1을 이용하여 "홍콩" 이후에 연속하는 단어 "여배우"가 클러스터 A에 속하는 문서들에 포함되는 조건부 확률을 2/3로 계산할 수 있다. 보다 상세하게는, 클러스터 A에 속하는 문서 MMM_news#1, NNN_news#0, 및 NNN_news#1 중에서 "홍콩 여배우"가 포함된 문서는 MMM_news#1 및 NNN_news#1이다. 이에 따라, 기준 단어 확장부(232)는 "홍콩 여배우"의 조건부 확률을 2/3로 계산할 수 있다.

[0128] 동일한 방법으로, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서들에 "홍콩" 이후에 연속하는 단어 "최고"가 포함되는 조건부 확률을 1/3로 계산할 수 있다. 보다 상세하게는, 클러스터 A에 속하는 문서들 중에 "홍콩 최고"를 포함하는 문서는 NNN_news#0 하나이므로, 기준 단어 확장부(432)는 아래의 수학적 식 1을 이용하여 "홍콩 최고"의 조건부 확률을 1/3로 계산할 수 있다.

수학적 식 1

$$P(\text{여배우}|\text{홍콩}) = \frac{P(\text{홍콩여배우})}{P(\text{홍콩})} = \frac{2}{3}$$

[0129]

$$P(\text{최고}|\text{홍콩}) = \frac{P(\text{홍콩최고})}{P(\text{홍콩})} = \frac{1}{3}$$

[0130]

[0131] 그리고, 기준 단어 확장부(432)는 계산된 조건부 확률 중 확률 값이 큰 확률이 기설정된 기준 확률 이상이면, 확률 값이 큰 확률에 해당하는 단어로 대표 기준 단어를 확장할 수 있다.

[0132] 일례로, 기준 확률이 1/2로 기설정된 경우, 기준 단어 확장부(432)는 계산된 조건부 확률을 비교하여 확률 값이 조건부 확률을 선택할 수 있다. 그러면, 기준 단어 확장부(432)는 조건부 확률 2/3와 1/3을 비교하여 2/3를 선택할 수 있다. 그리고, 기준 단어 확장부(432)는 선택한 조건부 확률 2/3이 기준 확률 1/2 이상인지를 판단할 수 있다. 이때, 조건부 확률 2/3이 기준 확률 1/2 이상이므로, 기준 단어 확장부(432)는 조건부 확률 2/3에 해당하는 단어 "홍콩 여배우"로 대표 기준 단어를 확장할 수 있다. 즉, 대표 기준 단어가 "홍콩"에서 "홍콩 여배우"로 확장될 수 있다.

[0133] 이때, 기준 단어 확장부(432)는 확장하고자 하는 단어의 조건부 확률이 기준 확률 미만이 될 때까지 대표 기준 단어를 계속하여 확장할 수 있다.

[0134] 일례로, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서들에 확장된 단어 "여배우" 이후에 연속하는 단어가 포함될 조건부 확률을 계산할 수 있다. 이때, 기준 단어 확장부(432)는 위의 수학적 식 1과 동일한 방법으로 P(BBB와|여배우)의 조건부 확률을 1로 계산할 수 있다. 보다 상세하게는, 클러스터 A에 속하는 모든 문서에 여배우 이후에 연속하여 단어 "BBB와"가 포함되므로, 기준 단어 확장부(432)는 P(BBB와|여배우)의 조건부 확률을 3/3=1로 계산할 수 있다. 그리고, 기준 단어 확장부(432)는 계산된 P(BBB와|여배우)의 조건부 확률 1이 기설정된 기준 확률 1/2 이상이므로, 대표 기준 단어를 "홍콩 여배우"에서 "홍콩 여배우 BBB와"로 확장할 수 있다.

[0135] 이때, 기준 단어 확장부(432)는 확장된 단어의 조건부 확률이 기준 확률 미만이면, 대표 기준 단어의 확장을 종료할 수 있다. 일례로, 바이그램 조건부 확률을 이용하는 경우, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서들에서 확장된 단어 "BBB와" 이후에 연속하는 단어를 포함하는 조건부 확률이 1/3, 1/3, 1/3로 기준 확률 1/2 미만이므로 대표 기준 단어의 확장을 종료할 수 있다.

[0136] 보다 상세하게는, 문서 MMM_news#1은 "BBB와" 연속하는 단어로 "비밀"을 포함하고, NNN_news#0는 "BBB와" 연속하는 단어로 "중서"를 포함하고, 및 NNN_news#1는 "BBB와" 연속하는 단어로 "몰래"를 포함한다. 이에 따라, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서들에서 "BBB와" 이후에 연속하는 단어를 포함하는 조건부 확률을 1/3, 1/3, 1/3로 계산할 수 있다. 이때, 계산된 조건부 확률이 모두 동일하므로, 기준 확장부(432)는 이중 하나인 1/3을 선택하고, 선택된 조건부 확률 1/3과 기준 확률 1/2을 비교할 수 있다. 그리고, 기준 확장부

(432)는 선택된 조건부 확률 1/2이 기준 확률 1/2 미만이므로 대표 기준 단어의 확장을 종료할 수 있다. 이에 따라, 기준 단어 확장부(432)는 바이그램 조건부 확률을 이용하여 대표 기준 단어 "홍콩"을 "홍콩 여배우 BBB와"까지 확장할 수 있다.

- [0137] 이어, 기준 단어 확장부(432)는 첫 번째로 빈도수가 높은 대표 기준 단어 "홍콩"의 확장이 종료되면, 두 번째로 빈도수가 높은 대표 기준 단어 "데이트"를 확장할 수 있다. 이때, 기준 단어 확장부(432)는 "홍콩"과 마찬가지로 바이그램 조건부 확률을 이용하여 "데이트"를 확장할 수 있다.
- [0138] 일례로, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서들에서 "데이트" 이전에 연속하는 단어 "비밀 데이트" 및 "몰래 데이트"의 조건부 확률을 2/3, 1/3로 각각 계산할 수 있다. 그리고, 기준 단어 확장부(432)는 계산된 조건부 확률 중 확률 값이 큰 2/3를 선택할 수 있다. 이어, 기준 단어 확장부(432)는 선택된 조건부 확률 2/3이 기준 확률 1/2 이상이므로, 2/3에 해당하는 단어 "비밀 데이트"로 대표 기준 단어를 확장할 수 있다.
- [0139] 이때, 기준 단어 확장부(432)는 "홍콩"과 마찬가지로, 확장된 단어의 바이그램 조건부 확률이 기준 확률 미만일 때까지 대표 기준 단어를 계속 확장할 수 있다. 그러면, 기준 단어 확장부(432)는 대표 기준 단어 "데이트"를 최종적으로 "비밀 데이트"까지 확장할 수 있다.
- [0140] 그리고, 기준 단어 확장부(432)는 확장된 대표 기준 단어들을 조합하여 제1 후보 대표 문구를 생성할 수 있다. 일례로, 기준 단어 확장부(432)는 "홍콩 여배우 BBB와"와 "비밀 데이트"를 조합하여 "홍콩 여배우 BBB와 비밀 데이트"를 제1 후보 대표 문구로 생성할 수 있다. 이때, 제1 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는, 제1 후보 대표 문구 "홍콩 여배우 BBB와 비밀 데이트"와 인기 키워드 "AAA"를 결합하여 웹 페이지에 노출할 수 있다.
- [0141] 동일한 방법으로, 제1 후보 대표 문구 생성부(430)는 클러스터 B에 포함된 모든 문서들의 형태소를 분석하여 하나 이상의 기준 단어를 결정할 수도 있다. 일례로, 기준 단어 결정부(431)는 문서 LLL#0 및 MMM_news#3의 형태소를 분석하여 "BBB", "콘서트", "영상", "인기", "폭발", "연인"을 기준 단어로 결정할 수 있다. 이때, 기준 단어 결정부(431)는 결정된 기준 단어가 상기 문서들 LLL#0, MMM_news#3에 포함되는 빈도수를 카운트하고, 카운트된 기준 단어의 빈도수가 가장 높은 순으로 2개 이상의 기준 단어를 대표 기준 단어로 결정할 수 있다. 그러면, 기준 단어 확장부(432)는 결정된 대표 기준 단어와 연속하는 단어가 클러스터 B에 속하는 문서들에 포함되는 조건부 확률을 계산할 수 있다.
- [0142] 그리고, 기준 단어 확장부(432)는 계산된 조건부 확률 중 확률 값이 큰 확률이 기설정된 기준 확률 이상이면, 확률 값이 큰 확률에 해당하는 단어로 대표 기준 단어를 확장할 수 있다. 이때, 기준 단어 확장부(432)는 확장하고자 하는 단어의 조건부 확률이 기준 확률 미만일 때까지 대표 기준 단어를 계속하여 확장할 수 있다. 마지막으로, 기준 단어 확장부(432)는 확장된 대표 기준 단어들을 조합하여 클러스터 B에 포함된 문서들에 대한 제1 후보 대표 문구를 생성할 수 있다.
- [0143] 이때, 제1 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 생성된 클러스터 A 및 클러스터 B에 대한 제1 후보 대표 문구를 인기 키워드 "AAA"와 각각 결합하여 웹 페이지에 노출할 수 있다. 일례로, 대표 문구 제공부는 클러스터 A 및 클러스터 B에 노출 우선 순위를 부여하고, 노출 우선 순위가 높은 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0144] 보다 상세하게는, 클러스터 A의 노출 우선 순위가 80이고, 클러스터 B의 노출 우선 순위가 60인 경우, 인기 키워드 "AAA" 검색이 요청되면, 대표 문구 제공부는 클러스터 A의 대표 문구와 인기 키워드 "AAA"를 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부는 복수의 클러스터들에 포함된 문서들의 가중치를 이용하여 노출 우선 순위를 결정할 수 있다. 일례로, 대표 문구 제공부는 클러스터에 포함된 문서들의 가중치의 합이 높은 순으로 노출 우선 순위를 높게 부여할 수 있다. 그러면, 대표 문구 제공부는 복수의 클러스터들 중 노출 우선 순위가 가장 높은 클러스터를 선택하고, 선택한 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0145] 다른 예로, 대표 문구 제공부는 클러스터의 크기(size)를 이용하여 복수의 클러스터 중 어느 하나의 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다. 여기서, 클러스터의 크기는 클러스터에 속하는 문서의 개수이다. 즉, 대표 문구 제공부는 클러스터 A 및 클러스터 B 중 문서를 많이 포함하고 있는 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0146] 이때, 대표 문구 제공부는 클러스터에 속하는 문서의 작성 시간에 따라 가중치를 부가하여 문서의 크기를 계산할 수도 있다. 그러면, 대표 문구 제공부는 최근에 작성된 문서일수록 높은 가중치를 부가하여, 클러스터 A

에 포함된 문서의 개수가 클러스터 B에 포함된 문서의 개수보다 작더라도 최근에 작성된 문서의 개수가 클러스터 B보다 많은 경우, 클러스터 A의 크기가 클러스터 B의 크기보다 클 수도 있다. 이 경우, 대표 문구 제공부는 클러스터 A의 제1 후보 대표 문구와 인기 키워드를 결합하여 노출할 수 있다.

[0147] 마찬가지로, 클러스터 B에 포함된 문서의 개수가 클러스터 A에 포함된 문서의 개수보다 작더라도, 클러스터 B에 클러스터 A에 포함된 문서들보다 최근에 작성된 문서가 많은 경우, 클러스터 B의 크기가 클러스터 A의 크기보다 클 수도 있다. 그러면, 대표 문구 제공부는 클러스터 B의 제1 대표 문구와 인기 키워드를 결합하여 노출할 수도 있다.

[0148] 이때, 대표 문구 제공부는 인기 키워드들에 해당하는 제1 후보 대표 문구를 모두 웹 페이지 상에 노출할 수 있을 뿐만 아니라, 웹 페이지에 노출된 인기 키워드들 중 마우스, 터치, 포인터 등에 의해 활성화된 인기 키워드의 제1 대표 문구를 웹 페이지에 노출할 수도 있다.

[0149] 일례로, 도 5와 같이, 실시간 인기 키워드들을 포함하는 목록(500)에서 마우스가 AAA 키워드(510)를 가리키는 경우, 대표 문구 제공부는 인기 키워드 AAA(510)와 AAA에 대한 제1 후보 대표 문구(521)를 함께 웹 페이지에 노출할 수 있다.

[0150] 이때, 대표 문구 제공부는 인기 키워드에 대한 대표 문구(521)와 함께 인기 키워드에 대한 정지 영상 등의 사진(522)을 함께 웹 페이지(520)에 노출할 수도 있다. 이를 위해, 대표 문구 제공부는 수집된 문서들의 작성 시간에 따라 문서들에 가중치를 부가할 수 있다. 일례로, 대표 문구 제공부는 작성 시간이 최근에 해당할수록 가중치를 높게 부여할 수 있다. 그러면, 대표 문구 제공부는 클러스터에 포함된 문서들 중에서 인기 키워드에 대한 사진을 포함하고 있는 문서를 추출할 수 있다. 그리고, 대표 문구 제공부는 추출된 문서들 중 가중치가 가장 높은 문서에 포함된 사진을 인기 키워드에 대한 제1 후보 대표 문구 및 인기 키워드와 결합하여 제공할 수 있다.

[0151] 이상에서는, 기설정된 기준 확률을 이용하여 대표 기준 단어를 확장함에 따라 제1 후보 대표 문구를 생성하는 과정에 대해 설명하였으나, 기준 단어 확장부(432)는 아래의 수학적 2를 이용하여 기준 확률을 계산할 수도 있다. 그러면, 기준 단어 확장부(432)는 계산된 기준 확률과 조건부 확률을 비교하여 대표 기준 단어를 확장할 수도 있다.

수학적 2

$$\text{기준 확률}_{K_Cluster} = \frac{1}{N(K_Cluster)^{\alpha}}$$

[0152]

[0153] 수학적 2에서, N(K_Cluster)는 K 클러스터에 포함된 문서의 개수이고, α는 임의의 변수이다. 일례로, α로는 0.3, 0.4, 0.5, 0.7 등의 실수가 이용될 수 있다.

[0154] 수학적 2에 따르면, 기준 단어 확장부(432)는 클러스터에 속하는 문서의 개수에 기초하여 기준 확률을 계산할 수 있다. 일례로, α=0.5가 이용되는 경우, 기준 단어 확장부(432)는 클러스터 A에 속하는 문서의 개수가 3개

$$\frac{1}{3^{0.5}} = 0.577$$

이므로, 클러스터 A에 대한 기준 확률을 0.577으로 계산할 수 있다.

[0155] 도 6은 본 발명의 일 실시예에 따른 제1 후보 대표 문구를 생성하는 방법을 설명하기 위해 제공되는 흐름도이다.

[0156] 먼저, 610 단계에서, 문서 수집부(410)는 웹 페이지에 노출하고자 하는 실시간 인기 키워드와 관련된 문서들을 포털 등을 통해 수집할 수 있다.

[0157] 이어, 620 단계에서, 클러스터 수행부(420)는 수집된 문서들의 형태소를 분석하여 클러스터를 수행할 수 있다.

[0158] 그리고, 630 단계에서, 기준 단어 결정부(431)는 클러스터에 포함된 문서들의 형태소를 분석하여, 문서 별로 기준 단어를 결정할 수 있다. 이때, 기준 단어 결정부(431)는 형태소 분석을 통해 결정된 기준 단어가 문서들에 포함되는 빈도수를 카운트할 수 있다. 일례로, 클러스터 A에 포함된 문서들을 대상으로 결정된 기준 단어 및

기준 단어의 빈도수는 위의 표 5와 같을 수 있다.

- [0159] 이어, 640 단계에서, 기준 단어 결정부(431)는 기준 단어의 빈도수에 기초하여 대표 기준 단어를 결정할 수 있다. 일례로, 기준 단어 결정부(431)는 위의 표 5에 따라 결정된 기준 단어들 중 빈도수가 높은 2개의 기준 단어 "홍콩" 및 "데이트"를 대표 기준 단어로 결정할 수 있다.
- [0160] 그러면, 650 단계에서, 기준 단어 확장부(432)는 바이그램(bi-gram) 조건부 확률을 이용하여 결정된 대표 기준 단어의 조건부 확률을 계산할 수 있다. 여기서, 조건부 확률은 클러스터에 포함되는 문서들에서 대표 기준 단어와 이전 또는 이후에 연속하는 단어가 포함되는 확률을 의미한다.
- [0161] 일례로, 기준 단어 확장부(432)는 클러스터 A 포함된 문서들에서 대표 기준 단어 "홍콩" 이후에 단어 "여배우"가 연속하는 조건부 확률을 2/3 및 1/3로 계산할 수 있다. 이때, 기준 단어 확장부(432)는 계산된 조건부 확률 중 큰 조건부 확률 2/3을 선택할 수 있다.
- [0162] 이어, 660 단계에서, 기준 단어 확장부(432)는 선택된 조건부 확률이 기설정된 기준 확률 이상인지 여부를 판단할 수 있다. 일례로, 기준 확률이 1/2로 기설정된 경우, 기준 단어 확장부(432)는 계산된 "홍콩 여배우"의 조건부 확률 2/3이 기준 확률 1/2 이상인지를 판단할 수 있다.
- [0163] 이때, 계산된 조건부 확률이 기준 확률 이상으로 판단된 경우, 기준 단어 확장부(432)는 대표 기준 단어를 선택된 조건부 확률에 해당하는 단어로 확장할 수 있다. 그리고, 680 단계에서, 기준 단어 확장부(432)는 확장된 대표 기준 단어를 650 내지 660과 동일한 방법을 이용하여 계속 확장할 수 있다. 일례로, 기준 단어 확장부(432)는 "여배우" 이후에 "BBB와"가 연속하는 조건부 확률을 계산하고, 계산된 조건부 확률과 기준 확률을 비교함으로써 대표 기준 단어를 "BBB와"로 확장할 수 있다.
- [0164] 이때, 계산된 조건부 확률이 기준 확률 미만으로 판단된 경우, 670 단계에서, 기준 단어 확장부(432)는 대표 기준 단어의 확장을 종료할 수 있다.
- [0165] 일례로, 기준 단어 확장부(432)는 클러스터 A에 포함된 문서들에서 대표 기준 단어 "BBB와" 이후에 연속하는 단어가 포함될 조건부 확률을 계산할 수 있다. 이때, 계산된 조건부 확률이 1/3로 모두 기준 확률 1/2 미만이므로, 기준 단어 확장부(432)는 대표 기준 단어의 확장을 종료할 수 있다. 그러면, 확장된 대표 기준 단어는 "홍콩 여배우 BBB와"가 될 수 있다.
- [0166] 마찬가지로, 기준 단어 확장부(432)는 "홍콩"과 동일한 방법으로 640 단계에서 결정된 대표 기준 단어 "데이트"에 대해 확장을 수행할 수 있다. 그러면, 기준 단어 확장부(432)는 대표 기준 단어를 "비밀 데이트"로 확장할 수 있다.
- [0167] 이어, 690 단계에서, 기준 단어 확장부(432)는 확장된 대표 기준 단어들을 조합하여 제1 후보 대표 문구를 생성할 수 있다. 일례로, 기준 단어 확장부(432)는 대표 기준 단어 "홍콩"을 기초로 확장된 "홍콩 여배우 BBB와"와 대표 기준 단어 "데이트"를 기초로 확장된 "비밀 데이트"를 조합하여 "홍콩 여배우 BBB와 비밀 데이트"를 제1 후보 대표 문구로 생성할 수 있다.
- [0168] 이때, 제1 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 생성된 제1 후보 대표 문구와 인기 키워드를 결합하여 제공할 수 있다.
- [0169] 일례로, 도 5와 같이, 대표 문구 제공부는 인기 키워드 목록(500)에 포함된 모든 인기 키워드들에 대한 대표 문구들을 웹 페이지에 노출할 수 있다. 이외에, 대표 문구 제공부는 인기 키워드 목록(500)에 포함된 인기 키워드들 중 마우스 등에 의해 활성화된 인기 키워드에 대한 대표 문구만을 웹 페이지에 노출할 수도 있다.
- [0170] 다른 예로, 대표 문구 제공부는 제1 후보 대표 문구(521)와 함께 인기 키워드에 대한 정지 영상 등의 사진(522)을 함께 노출할 수도 있다. 이때, 대표 문구 제공부는 수집된 문서들의 작성 시간에 따라 문서들에 가중치를 부가하고, 부가된 가중치가 가장 높은 문서에 포함된 사진을 인기 키워드에 대한 제1 후보 대표 문구와 함께 웹 페이지에 노출할 수 있다. 그리고, 가중치가 가장 높은 문서에 사진이 포함되지 않은 경우, 대표 문구 제공부는, 가중치가 차순위인 문서에 포함된 사진을 이용할 수도 있다. 여기서, 대표 문구 제공부는 문서의 작성 시간이 최근에 해당할수록 가중치를 높게 부여할 수 있다.
- [0171] 이상의 도 6에서는 하나의 클러스터 A에 대한 제1 후보 대표 문구를 생성하고, 생성한 클러스터 A의 제1 후보 대표 문구와 인기 키워드를 결합하여 제공하는 것에 대해 설명하였으나, 본 대표 문구 제공 방법은, 복수의 클러스터에 대한 제1 후보 대표 문구를 생성할 수 있다. 그리고, 제1 후보 대표 문구가 대표 문구로 결정된

경우, 대표 문구 제공부는, 생성한 복수의 제1 후보 대표 문구 중 적어도 하나와 인기 키워드를 결합하여 웹 페이지에 노출할 수도 있다.

- [0172] 일례로, 대표 문구 제공부는 생성된 클러스터 A 및 클러스터 B에 대한 제1 후보 대표 문구를 인기 키워드 "AA A"와 각각 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부는 클러스터 A 및 클러스터 B에 노출 우선 순위를 부여하고, 노출 우선 순위가 높은 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0173] 보다 상세하게는, 클러스터 A의 노출 우선 순위가 80이고, 클러스터 B의 노출 우선 순위가 60인 경우, 인기 키워드 "AAA" 검색이 요청되면, 대표 문구 제공부는 클러스터 A의 제1 후보 대표 문구와 인기 키워드 "AAA"를 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부는 복수의 클러스터들에 포함된 문서들의 가중치를 이용하여 노출 우선 순위를 결정할 수 있다. 일례로, 대표 문구 제공부는 클러스터에 포함된 문서들의 가중치의 합이 높은 순으로 노출 우선 순위를 높게 부여할 수 있다. 그러면, 대표 문구 제공부는 복수의 클러스터들 중 노출 우선 순위가 가장 높은 클러스터를 선택하고, 선택한 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0174] 다른 예로, 대표 문구 제공부는 클러스터의 크기(size)를 이용하여 복수의 클러스터 중 어느 하나의 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다. 여기서, 클러스터의 크기는 클러스터에 속하는 문서의 개수이다. 즉, 대표 문구 제공부는 클러스터 A 및 클러스터 B 중 문서를 많이 포함하고 있는 클러스터의 제1 후보 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0175] 이때, 대표 문구 제공부는 클러스터에 속하는 문서의 작성 시간에 따라 가중치를 부가하여 문서의 크기를 계산할 수도 있다. 그러면, 대표 문구 제공부는 최근에 작성된 문서일수록 높은 가중치를 부가하여, 클러스터 A에 포함된 문서의 개수가 클러스터 B에 포함된 문서의 개수보다 작더라도 최근에 작성된 문서의 개수가 클러스터 B보다 많은 경우, 클러스터 A의 크기가 클러스터 B의 크기보다 클 수도 있다. 이 경우, 대표 문구 제공부는 클러스터 A의 제1 후보 대표 문구와 인기 키워드를 결합하여 노출할 수 있다.
- [0176] 마찬가지로, 클러스터 B에 포함된 문서의 개수가 클러스터 A에 포함된 문서의 개수보다 작더라도, 클러스터 B에 클러스터 A에 포함된 문서들보다 최근에 작성된 문서가 많은 경우, 클러스터 B의 크기가 클러스터 A의 크기보다 클 수도 있다. 그러면, 대표 문구 제공부는 클러스터 B의 제1 후보 대표 문구와 인기 키워드를 결합하여 노출할 수도 있다.
- [0177] 지금까지, 도 4 내지 도 6을 참조하여 제1 후보 대표 문구를 생성하고, 제1 후보 대표 문구가 대표 문구로 결정됨에 따라 제1 후보 대표 문구와 인기 키워드를 결합하여 제공하는 구성에 대해 설명하였다. 이때, 도 4 내지 도 6에서는, 복수의 기준 단어들 중 빈도수가 높은 순으로 2개의 대표 기준 단어를 결정하고, 조건부 확률에 기초하여 대표 기준 단어를 확장 및 대표 문구를 생성하는 것으로 설명하였으나, 이는 실시예에 해당되며, 기준 단어 결정부(431)는 복수의 기준 단어들 중 빈도수가 가장 높은 하나의 대표 기준 단어를 결정하고, 기준 단어 확장부(432)는 결정된 대표 기준 단어를 확장하여 대표 문구를 생성할 수도 있다.
- [0178] 이하에서는, 도 7 내지 도 11을 참조하여, 제2 후보 대표 문구를 생성하는 구성, 및 제2 후보 대표 문구가 대표 문구로 결정된 경우, 제2 후보 대표 문구와 인기 키워드를 결합하여 제공하는 구성에 대해 설명하기로 한다.
- [0179] 도 7은 본 발명의 실시예에 따른 제2 후보 대표 문구 생성부의 세부 구성을 도시한 블럭도이다.
- [0180] 도 7에 따르면, 도 2의 대표 문구 제공 시스템은, 방송 데이터 수집부, 및 데이터 베이스를 더 포함할 수 있다.
- [0181] 방송 데이터 수집부(710)는 방송국으로부터 방송 데이터들을 수집할 수 있다. 여기서, 방송국은 공중파 TV 방송, 케이블 TV 방송, 라디오 방송, 인터넷 방송, 위성 TV 방송을 포함할 수 있다. 이때, 방송 데이터 수집부(710)는 매일 오전, 매일 점심, 매일 오후, 또는 매주 목요일 등과 같이, 기설정된 시간마다 주기적으로 방송 데이터들을 수집할 수 있다.
- [0182] 그리고, 방송 데이터 수집부(710)는 수집된 방송 데이터를 아래의 표 6의 형식으로 데이터 베이스(720)에 저장할 수 있다.

표 6

방송국	프로그램명	시작시간	종료시간	출연자이름	극중이름	방송내용
AAA	AAA 뉴스	20100312 20:00	20100312 21:00	KKK		선거 투표
BBB2	XXX	20100312 22:00	20100312 23:20	000, PPP	QQQ, RRR	노비, 도망, 장군
BBB1	YYY	20100312 22:00	20100312 23:00	JJJ		안경 비밀
BBB2	ZZZ	20100312 21:00	20100312 22:00	VVV, PPP		모내기 귀농

- [0184] 제2 후보 대표 문구 생성부(730)는 방송 데이터에서 인기 키워드를 포함하는 방송 프로그램을 결정할 수 있다. 이때, 제2 후보 대표 문구 생성부(730)는 인기 키워드의 노출 시간과 결정된 방송 프로그램의 방송 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 여기서, 방송 시간은 방송 시작 시간 및 방송 종료 시간을 포함할 수 있다.
- [0185] 일례로, 제2 후보 대표 문구 생성부(730)는 인기 키워드의 노출 시간이 방송 시작 시간 이전인 경우, 노출 시간과 방송 시작 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다.
- [0186] 다른 예로, 인기 키워드의 노출 시간이 방송 종료 시간 이후인 경우, 제2 후보 대표 문구 생성부(730)는 노출 시간과 방송 종료 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다.
- [0187] 또 다른 예로, 인기 키워드의 노출 시간이 방송 시작 시간과 방송 종료 시간 사이에 포함되는 경우, 제2 후보 대표 문구 생성부(730)는 방송 시작 시간 및 방송 종료 시간 중 어느 하나와 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 이때, 방송 시작 시간 2시간 전과 방송 종료 시간 2시간 이후에 방송 프로그램에 대한 검색이 많이 수행되므로, 기준값은 2시간으로 기설정될 수 있다.
- [0188] 그러면, 제2 후보 대표 문구 생성부(730)는 계산된 차이값과 기설정된 기준값을 비교하여 인기 키워드에 대한 대표 문구를 생성할 수 있다. 일례로, 제2 후보 대표 문구 생성부(730)는 계산된 차이값이 기설정된 기준값 이하이면, 결정된 방송 프로그램의 방송 시간에 기초하여 제2 후보 대표 문구를 생성할 수 있다.
- [0189] 보다 상세하게는, 제2 후보 대표 문구 생성부(730)는 방송 프로그램 결정부(731), 매칭 점수 계산부(732), 및 생성부(733)를 포함할 수 있다.
- [0190] 방송 프로그램 결정부(731)는 방송 데이터에 인기 키워드를 매칭하여, 방송 데이터에서 인기 키워드를 포함하는 방송 프로그램을 결정할 수 있다.
- [0191] 일례로, 인기 키워드로 "XXX"가 웹 페이지에 노출된 경우, 방송 프로그램 결정부(731)는 표 6의 방송 데이터에서 인기 키워드 "XXX"를 포함하는 방송 프로그램을 결정할 수 있다.
- [0192] 이때, 노출 시간이 방송 시작 시간 이전인 경우, 매칭 점수 계산부(732)는 결정된 방송 프로그램의 방송 시작 시간과 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 그러면, 생성부(733)는 계산된 차이값이 기설정된 기준값 이하이면, 인기 키워드 "XXX"가 방송 프로그램 "XXX"로 인해 실시간으로 인기 키워드가 된 원인으로 결정할 수 있다.
- [0193] 이에 따라, 생성부(733)는 인기 키워드의 원인으로 결정된 방송 프로그램을 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성할 수 있다.
- [0194] 일례로, 생성부(733)는 방송 프로그램 "XXX"의 방송 시작 시간을 기초로 "22:00부터 방송"을 제2 후보 대표 문구로 생성할 수 있다.
- [0195] 다른 예로, 인기 키워드로 "QQQ"가 웹 페이지에 노출된 경우, 방송 프로그램 결정부(731)는 표 6의 방송 데이터에서 인기 키워드 "QQQ"를 포함하는 방송 프로그램 "XXX"를 결정할 수 있다. 이때, 노출 시간이 방송 종료 시간 이후인 경우, 매칭 점수 계산부(732)는 결정된 방송 프로그램의 방송 종료 시간과 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 그러면, 생성부(733)는 계산된 차이값이 기설정된 기준값 이하이면, 인기 키워드 "QQQ"가 방송 프로그램 "XXX"로 인해 실시간으로 인기 키워드가 된 원인으로 결정할 수 있다.
- [0196] 이에 따라, 생성부(733)는 인기 키워드의 원인으로 결정된 방송 프로그램을 기초로 제2 후보 대표 문구를 생성

할 수 있다. 일례로, 생성부(733)는 방송 프로그램 "XXX"의 방송 시간을 기초로 "22:00부터 23:20까지 방송"을 제2 후보 대표 문구로 생성할 수 있다.

- [0197] 한편, 노출 시간이 방송 시작 시간과 방송 종료 시간 사이에 포함된 경우, 즉, 노출 시간이 방송 시작 시간 이후이고, 방송 종료 시간 이전인 경우, 매칭 점수 계산부(732)는 결정된 방송 프로그램의 방송 시작 시간과 방송 종료 시간 중 어느 하나와 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 그리고, 생성부(733)는 차이값에 따라 결정된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다.
- [0198] 이하에서는, 도 7을 계속 참조하여, 인기 키워드가 여러 단어가 조합된 구조인 경우, 제2 후보 대표 문구를 생성하는 과정에 대해 설명하기로 한다.
- [0199] 먼저, 여러 단어가 조합된 형태의 인기 키워드가 노출된 경우, 방송 프로그램 결정부(731)는 인기 키워드를 형태소 분석하여 복수의 단어들로 분리할 수 있다. 그리고, 방송 프로그램 결정부(731)는 방송 데이터에 분리된 단어를 매칭하여, 방송 데이터에서 분리된 단어를 포함하는 방송 프로그램을 결정할 수 있다. 이때, 방송 프로그램 결정부(731)는 방송 데이터에서 분리된 단어를 포함하는 방송 프로그램이 존재할 때까지 분리된 단어들을 계속하여 분리할 수 있다.
- [0200] 일례로, 인기 키워드가 여러 단어가 조합된 "XXX PPP QQQ" 구조인 경우, 방송 프로그램 결정부(731)는 인기 키워드를 형태소 분석하여 "XXX", "PPP", "QQQ"로 분리할 수 있다. 그리고, 방송 프로그램 결정부(731)는 방송 데이터와 매칭을 통해 분리된 단어들을 포함하는 방송 프로그램을 결정할 수 있다.
- [0201] 이때, PPP를 이용하는 경우를 예로 들면, 방송 프로그램 결정부(731)는 방송 데이터에서 분리된 단어 "PPP"를 포함하는 방송 프로그램으로 "XXX"와 "ZZZ"를 결정할 수 있다. 그러면, 방송 프로그램 결정부(731)는 방송 데이터에 분리된 단어들을 포함하는 방송 프로그램이 존재하므로, 분리된 단어 "PPP"를 계속 분리하는 것을 종료할 수 있다. 여기서, 방송 데이터에 분리된 단어들을 포함하는 방송 프로그램이 존재하지 않는 경우, 방송 프로그램 결정부(731)는 분리된 단어 "PPP"를 계속하여 분리할 수 있다.
- [0202] 이때, 분리된 단어를 포함하는 방송 프로그램이 복수 개인 경우, 방송 프로그램 결정부(731)는 결정된 방송 프로그램 중 분리된 단어들을 가장 많이 포함하는 방송 프로그램을 선택할 수 있다.
- [0203] 일례로, 방송 데이터에서 분리된 단어 "PPP"를 포함하는 방송 프로그램으로 "XXX"와 "ZZZ"가 결정된 경우, 방송 프로그램 결정부(731)는 방송 프로그램 "XXX" 및 "ZZZ"에 포함되는 분리된 단어의 개수를 카운트할 수 있다. 그러면, 방송 프로그램 결정부(731)는 방송 프로그램 "XXX"는 3개의 분리된 단어 "XXX", "PPP", 및 "QQQ"를 포함하고, "ZZZ"는 1개의 분리된 단어 "PPP"를 포함하는 것으로 카운트할 수 있다. 이에 따라, 방송 프로그램 결정부(731)는 가장 많은 분리된 단어를 포함하는 방송 프로그램 "XXX"를 선택할 수 있다.
- [0204] 다른 예로, 방송 프로그램 결정부(731)는 분리된 단어를 포함하는 방송 프로그램이 복수 개이고, 복수의 방송 프로그램이 동일한 개수의 분리된 단어들을 포함하는 경우, 방송 프로그램 결정부(731)는 결정된 복수의 방송 프로그램의 방송 시간에 기초하여 복수의 방송 프로그램 중 어느 하나의 방송 프로그램을 선택할 수 있다. 이때, 방송 프로그램 결정부(731)는 복수의 방송 프로그램의 방송 시작 시간과 노출 시간, 방송 종료 시간과 노출 시간을 비교하고, 비교 결과 노출 시간에 가장 가까운 방송 시간에 해당하는 방송 프로그램을 선택할 수 있다.
- [0205] 마찬가지로, 인기 키워드가 하나의 단어로 구성되고, 인기 키워드를 포함하는 방송 프로그램이 복수개인 경우에도, 방송 프로그램 결정부(731)는 복수의 방송 프로그램의 방송 시간 중 노출 시간에 가장 가까운 방송 프로그램을 선택할 수 있다.
- [0206] 그러면, 매칭 점수 계산부(732)는 선택한 방송 프로그램의 방송 시간과 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다. 그리고, 생성부(233)는 차이값을 기초로 결정된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다. 여기서, 방송 시간은 방송 시작 시간 및 방송 종료 시간을 포함할 수 있다. 이때, 차이값을 계산하는 과정은 인기 키워드가 하나의 단어로 구성된 경우, 앞에서 설명한 매칭 점수 계산부(732)에서 차이값을 계산하는 과정과 동일하므로 중복되는 설명은 생략하기로 한다.
- [0207] 한편, 인기 키워드를 포함하는 방송 프로그램이 복수 개로 결정된 경우, 방송 프로그램 결정부(731)는 복수의 방송 프로그램의 매칭 점수를 이용하여 복수의 방송 프로그램 중 어느 하나를 선택할 수도 있다. 여기서, 인기 키워드는 하나의 단어로 구성될 수도 있고, 여러 단어가 조합된 형태일 수도 있다.
- [0208] 이때, 매칭 점수 계산부(732)는 방송 데이터를 구성하는 항목들 중 인기 키워드를 포함하는 항목에 따라 매칭

점수를 부가하고, 매칭 점수와 가중치를 이용하여 최종 매칭 점수를 계산할 수 있다.

- [0209] 먼저, 매칭 점수 계산부(732)는 방송 데이터를 구성하는 항목들 중 인기 키워드가 프로그램 명, 출연진, 극중 이름, 및 방송 내용에 포함되는지 여부에 따라 인기 키워드를 포함하는 방송 프로그램 별로 매칭 점수를 부가할 수 있다. 일례로, 매칭 점수 계산부(732)는 방송 데이터를 구성하는 항목들 중 인기 키워드가 프로그램 명에 포함되는 경우 100, 출연진에 포함되는 경우 80, 극중 이름에 포함되는 경우 80, 및 방송 내용에 포함되는 경우 50을 매칭 점수로 부가할 수 있다.
- [0210] 그리고, 매칭 점수 계산부(732)는 인기 키워드가 노출된 시간과 인기 키워드를 포함하는 방송 프로그램의 방송 시간을 이용하여 방송 프로그램 별 시간 점수를 부가할 수 있다. 일례로, 인기 키워드가 노출된 시간이 기설정된 오차 범위 내에서 방송 시간과 일치하는 경우, 매칭 점수 계산부(732)는 시간 점수를 50으로 부가할 수 있다. 이때, 매칭 점수 계산부(732)는 인기 키워드가 노출된 시간이 인기 키워드를 포함하는 방송 프로그램의 방송 시간과 10분 오차 시 마다 50에서 -5점을 감산하여 시간 점수를 부가할 수 있다.
- [0211] 또한, 매칭 점수 계산부(732)는 인기 키워드를 포함하는 방송 프로그램이 속하는 방송의 형태에 따라 방송 프로그램 별로 방송국 가중치를 부가할 수 있다. 이때, 매칭 점수 계산부(732)는 공중파 방송, 케이블 방송, 라디오 방송 순으로 높은 가중치를 부가할 수 있다. 일례로, 매칭 점수 계산부(732)는 공중파 방송에 가중치 1, 케이블 방송에 가중치 0.1, 라디오 방송에 가중치 0.05를 부가할 수 있다.
- [0212] 마지막으로, 매칭 점수 계산부(732)는 인기 키워드를 포함하는 방송 프로그램 별로 부가된 매칭 점수, 시간 점수, 및 방송국 가중치와 방송 프로그램의 시청률을 이용하여 최종 매칭 점수를 방송 프로그램 별로 계산할 수 있다. 일례로, 매칭 점수 계산부(732)는 아래의 수학적 식 3을 이용하여 인기 키워드를 포함하는 방송 프로그램 별 최종 매칭 점수를 계산할 수 있다.

수학적 식 3

$$\text{최종 매칭 점수} = (P+T) \times S \times A$$

- [0213]
- [0214] 수학적 식 3에서, P는 프로그램명, 출연진, 극중 이름, 및 방송 내용에 따라 부가된 매칭 점수, T는 시간 점수, S는 방송국 가중치, A는 시청률이다.
- [0215] 수학적 식 3에 따르면, 매칭 점수 계산부(732)는 매칭 점수와 시간 점수의 합에 방송국 가중치를 곱하고, 계산된 곱에 시청률을 곱하여 방송 프로그램 별로 최종 매칭 점수를 계산할 수 있다. 일례로, 인기 키워드 PPP를 포함하는 방송 프로그램 "XXX" 및 "ZZZ"의 시청률이 30%, 10%인 경우, 매칭 점수 계산부(732)는 $(P+T) \times S$ 에 0.3을 곱하여 XXX의 최종 매칭 점수를 계산하고, $(P+T) \times S$ 에 0.1을 곱하여 ZZZ의 최종 매칭 점수를 계산할 수 있다.
- [0216] 그러면, 방송 프로그램 결정부(731)는 인기 키워드를 포함하는 방송 프로그램들 중 계산된 최종 매칭 점수가 가장 높은 방송 프로그램을 선택할 수 있다.
- [0217] 이어, 생성부(733)는 선택한 방송 프로그램의 방송 시간을 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성할 수 있다. 일례로, 생성부(733)는 도 8과 같이, 선택한 방송 프로그램 "XXX"의 방송 시간을 기초로 "22:00부터 방송"을 제2 후보 대표 문구(821)로 생성할 수 있다.
- [0218] 이때, 제2 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 제2 후보 대표 문구(821)와 인기 키워드(810)를 결합하여 제공할 수 있다. 일례로, 대표 문구 제공부는 제2 후보 대표 문구(821)와 함께 인기 키워드에 대한 정지 영상 등의 사진(822)을 함께 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부는 실시간 인기 키워드들을 포함하는 목록(800)에서 마우스, 포인터 등에 의해 활성화된 인기 키워드에 대한 대표 문구와 사진을 웹 페이지에 노출할 수 있다.
- [0219] 도 9는 본 발명의 실시시에 따라 방송 시간을 이용하여 제2 후보 대표 문구를 제공하는 과정을 설명하기 위해 제공되는 흐름도이다.
- [0220] 도 9에 따르면, 910 단계에서, 방송 데이터 수집부(720)는 방송국으로부터 방송 데이터를 주기적으로 수집하여 데이터 베이스(710)에 저장할 수 있다. 여기서, 방송국은 공중파 TV 방송, 케이블 TV 방송, 라디오 방송, 인터넷 방송, 위성 TV 방송을 포함할 수 있다.

- [0221] 이어, 920 단계에서, 방송 프로그램 결정부(731)는 방송 데이터에서 인기 키워드를 포함하는 방송 프로그램을 결정할 수 있다. 일례로, 방송 프로그램 결정부(731)는 표 6과 같은 구성을 갖는 방송 데이터에서 웹 페이지에 노출된 인기 키워드 "QQQ"를 포함하는 방송 프로그램 "XXX"를 결정할 수 있다.
- [0222] 이때, 930 단계에서, 인기 키워드가 노출된 시간이 결정된 방송 프로그램의 방송 시작 시간과 방송 종료 시간 사이에 포함되지 않는 경우, 940 단계에서, 매칭 점수 계산부(732)는 결정된 방송 프로그램의 방송 시간과 노출 시간 간의 차의 절대값으로서 차이값을 계산할 수 있다.
- [0223] 일례로, 노출 시간이 방송 시작 시간 이전이거나, 방송 종료 시간 이후인 경우, 매칭 점수 계산부(732)는 결정된 방송 프로그램의 방송 시작 시간과 노출 시간 간의 차이값, 또는 방송 종료 시간과 노출 시간 간의 차이값을 계산할 수 있다.
- [0224] 그리고, 950 단계에서, 매칭 점수 계산부(732)는 계산된 차이값이 기설정된 기준값 이하인지 비교할 수 있다. 일례로, 방송 시작 시간 이전 2시간과 방송 종료 시간 이후 2시간 동안 방송 프로그램과 관련된 키워드가 인기 키워드로 검색될 확률이 높으므로, 기준 값은 2시간으로 기설정될 수 있다.
- [0225] 이때, 차이값이 기준값 이하인 경우, 960 단계에서, 생성부(733)는 결정된 방송 프로그램의 방송 시간을 기초로 인기 키워드에 대한 제2 후보 대표 문구를 생성할 수 있다.
- [0226] 일례로, 노출 시간이 결정된 방송 프로그램 "XXX"의 방송 시작 시간 이전에 해당하는 경우, 생성부(733)는 결정된 방송 프로그램 "XXX"의 방송 시작 시간을 이용하여 "22:00부터 방송 시작"을 제2 후보 대표 문구로 생성할 수 있다.
- [0227] 다른 예로, 노출 시간이 결정된 방송 프로그램 "XXX"의 방송 종료 시간 이후에 해당하는 경우, 생성부(733)는 결정된 방송 프로그램 "XXX"의 방송 시작 시간 및 방송 종료 시간을 이용하여 "22:00부터 23:20까지 방송"을 제2 후보 대표 문구로 생성할 수 있다.
- [0228] 이때, 대표 문구 결정부에서 제2 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 제2 후보 대표 문구와 인기 키워드를 결합하여 제공할 수 있다. 일례로, 대표 문구 제공부는 인기 키워드 "QQQ"와 제2 후보 대표 문구 "22:00부터 방송 시작"을 결합하여 웹 페이지에 노출할 수 있다. 이때, 대표 문구 제공부는 인기 키워드 "QQQ"의 방송 프로그램 "XXX"에 대한 화보 사진을 인기 키워드 및 제2 후보 대표 문구와 결합하여 웹 페이지에 노출할 수도 있다.
- [0229] 한편, 960 단계에서, 결정된 방송 프로그램이 복수개인 경우, 방송 프로그램 결정부(931)는 결정된 방송 프로그램의 방송 시간을 기초로 어느 하나의 방송 프로그램을 선택할 수 있다. 일례로, 결정된 방송 프로그램들의 방송 시작 시간과 노출 시간을 비교하고, 결정된 방송 프로그램들의 방송 종료 시간과 노출 시간을 비교하여, 노출 시간에 가장 가까운 방송 시간을 갖는 방송 프로그램을 선택할 수 있다. 그러면, 생성부(733)는 선택된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다.
- [0230] 도 10은 여러 단어가 조합된 형태를 갖는 인기 키워드에 대한 제2 후보 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 흐름도이다.
- [0231] 도 10에 따르면, 1010 단계에서, 방송 프로그램 결정부(731)는 여러 단어가 조합된 인기 키워드를 형태소 분석하여 조합된 복수의 단어들을 분리할 수 있다.
- [0232] 일례로, 인기 키워드가 여러 단어가 조합된 "XXX PPP QQQ" 구조인 경우, 방송 프로그램 결정부(731)는 인기 키워드를 형태소 분석하여 "XXX", "PPP", "QQQ"로 분리할 수 있다.
- [0233] 이어, 1020 단계에서, 방송 프로그램 결정부(731)는 방송 데이터에서 분리된 단어들을 포함하는 방송 프로그램을 결정할 수 있다. 일례로, 방송 프로그램 결정부(731)는 방송 데이터에서 분리된 단어 "PPP"를 포함하는 방송 프로그램으로 "XXX"와 "ZZZ"를 결정할 수 있다. 그러면, 방송 프로그램 결정부(731)는 방송 프로그램 "XXX" 및 "ZZZ"에 포함되는 분리된 단어의 개수를 카운트할 수 있다. 결정부(231)는 방송 프로그램 "XXX"는 3개의 분리된 단어 "XXX", "PPP", 및 "QQQ"를 포함하고, "ZZZ"는 1개의 분리된 단어 "PPP"를 포함하는 것으로 카운트할 수 있다. 이에 따라, 방송 프로그램 결정부(731)는 가장 많은 분리된 단어를 포함하는 방송 프로그램 "XXX"를 선택할 수 있다.
- [0234] 이때, 1030 단계에서, 결정된 방송 프로그램이 복수 개이고, 1040 단계에서, 분리된 단어를 최대로 포함하는 방송 프로그램이 복수 개인 경우, 1050 단계에서, 방송 프로그램 결정부(731)는 결정된 방송 프로그램들 중 방송

시간이 인기 키워드의 노출 시간에 가장 가까운 방송 프로그램을 선택할 수 있다. 이때, 방송 프로그램 결정부(731)는 결정된 복수의 방송 프로그램의 방송 시작 시간 및 방송 종료 시간을 모두 노출 시간과 비교하여 노출 시간에 가장 가까운 어느 하나의 방송 프로그램을 선택할 수 있다.

- [0235] 이때, 1040 단계에서, 분리된 단어를 최대 포함하는 방송 프로그램이 복수개가 아닌 경우, 1060 단계에서, 방송 프로그램 결정부(731)는 분리된 단어를 최대 포함하는 방송 프로그램을 선택할 수 있다. 일례로, 방송 프로그램 "XXX"는 분리된 단어 3개를 포함하고, 방송 프로그램 "ZZZ"는 분리된 단어 1개를 포함하므로, 방송 프로그램 결정부(731)는 "XXX"를 선택할 수 있다.
- [0236] 이어, 1070 단계에서, 매칭 점수 계산부(732)는 선택된 방송 프로그램의 방송 시간과 노출 시간 간의 차이값을 계산할 수 있다. 이때, 차이값이 기설정된 기준값 이하인 경우, 1090 단계에서, 생성부(733)는 선택된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다. 그리고, 대표 문구 결정부에서 제2 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 제2 후보 대표 문구와 인기 키워드를 결합하여 제공할 수 있다. 여기서, 1070 내지 1090 단계의 동작은 앞에서 설명한 940 내지 960 단계와 중복되므로 자세한 설명은 생략하기로 한다.
- [0237] 도 11은 본 발명의 일 실시예에 따라 매칭 점수를 이용하여 제2 후보 대표 문구를 생성하는 과정을 설명하기 위해 제공되는 흐름도이다.
- [0238] 도 11에 따르면, 1110 단계에서, 방송 프로그램 결정부(731)는 수집된 방송 데이터에서 인기 키워드를 포함하는 방송 프로그램을 결정할 수 있다.
- [0239] 이어, 1020 단계에서, 매칭 점수 계산부(732)는 결정된 방송 프로그램 별로 매칭 점수를 부가할 수 있다. 이때, 매칭 점수 계산부(732)는 방송 데이터를 구성하는 항목들 중 인기 키워드가 프로그램명, 출연진, 극중 이름, 및 방송 내용에 포함되는지 여부에 따라 인기 키워드를 포함하는 방송 프로그램 별로 매칭 점수를 부가할 수 있다.
- [0240] 일례로, 매칭 점수 계산부(732)는 방송 데이터를 구성하는 항목들 중 인기 키워드가 프로그램명에 포함되는 경우 100, 출연진에 포함되는 경우 80, 극중 이름에 포함되는 경우 80, 및 방송 내용에 포함되는 경우 50을 매칭 점수로 부가할 수 있다.
- [0241] 그리고, 1130 단계에서, 매칭 점수 계산부(732)는 결정된 방송 프로그램 별로 시간 점수를 부가할 수 있다.
- [0242] 일례로, 매칭 점수 계산부(732)는 인기 키워드가 노출된 시간이 결정된 방송 프로그램의 방송 시간과 일치하는 경우, 일치하는 방송 프로그램의 시간 점수를 50으로 부가할 수 있다. 이때, 매칭 점수 계산부(732)는 인기 키워드의 노출 시간이 방송 시간과 정확히 일치하지 않더라도, 기설정된 오차 범위 내에서 일치하면 시간 점수를 50으로 부가할 수 있다.
- [0243] 그리고, 매칭 점수 계산부(732)는 방송 프로그램의 방송 시간과 인기 키워드의 노출 시간이 일치하지 않으면, 10분 오차 시 마다 50에서 -5점을 감산하여 시간 점수를 부가할 수 있다. 일례로, 방송 프로그램의 방송 시작 시간과 인기 키워드의 노출 시간이 30분 오차인 경우, 매칭 점수 계산부(732)는 시간 점수를 35로 부가할 수 있다.
- [0244] 이어, 1140 단계에서, 매칭 점수 계산부(732)는 방송 프로그램 별로 방송국 가중치를 부가할 수 있다. 일례로, 매칭 점수 계산부(732)는 결정된 방송 프로그램이 공중파 방송인지, 케이블 방송인지, 혹은 라디오방송인지에 따라 가중치를 다르게 부가할 수 있다.
- [0245] 그리고, 1150 단계에서, 매칭 점수 계산부(732)는 방송 프로그램 별로 최종 매칭 점수를 계산할 수 있다. 이때, 매칭 점수 계산부(732)는 위의 수학적 3과 같이, 결정된 방송 프로그램 별 시청률을 이용하여 방송 프로그램 별로 최종 매칭 점수를 계산할 수 있다.
- [0246] 일례로, 매칭 점수 계산부(732)는 부가된 매칭 점수와 시간 점수의 합을 계산하고, 계산된 합에 방송국 가중치와 시청률을 곱하여 결정된 방송 프로그램 별로 최종 매칭 점수를 계산할 수 있다.
- [0247] 그러면, 1160 단계에서, 방송 프로그램 결정부(731)는 계산된 최종 매칭 점수를 기초로 결정된 방송 프로그램들 중 어느 하나의 방송 프로그램을 선택할 수 있다. 일례로, 방송 프로그램 결정부(731)는 계산된 최종 매칭 점수가 가장 높은 방송 프로그램을 선택할 수 있다.
- [0248] 이어, 1170 단계에서, 매칭 점수 계산부(732)는 선택된 방송 프로그램의 방송 시간과 노출 시간 간의 차이값을

계산할 수 있다.

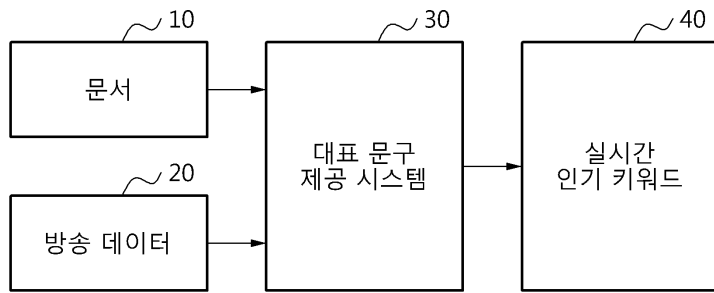
- [0249] 그리고, 1180 단계에서, 차이값이 기설정된 기준값 이하인 경우, 1190 단계에서, 생성부(733)는 선택된 방송 프로그램의 방송 시간을 이용하여 제2 후보 대표 문구를 생성할 수 있다.
- [0250] 이때, 제2 후보 대표 문구가 대표 문구로 결정된 경우, 대표 문구 제공부는 생성된 제2 후보 대표 문구와 인기 키워드를 결합하여 제공할 수 있다. 여기서, 1170 내지 1190 단계의 동작은 앞에서 설명한 940 내지 960 단계와 중복되므로 자세한 설명은 생략하기로 한다.
- [0251] 이상에서 설명한 바와 같이, 본 발명의 일 실시예에 따른 대표 문구 제공 시스템은 문서를 기초로 제1 후보 대표 문구를 생성하고, 방송 데이터를 기초로 제2 후보 대표 문구를 생성할 수 있다. 그리고, 최종 인기도 또는 제2 후보 대표 문구의 시간 점수에 기초하여 제1 후보 대표 문구 및 제2 후보 대표 문구 중 어느 하나를 인기 키워드에 대해 최종적인 대표 문구를 결정할 수 있다. 그리고, 대표 문구 제공 시스템은, 결정된 대표 문구와 인기 키워드를 결합하여 웹 페이지에 노출할 수 있다.
- [0252] 이때, 제1 후보 대표 문구 및 제2 후보 대표 문구 중 어느 하나가 인기 키워드에 대한 대표 문구로 결정되지 않은 경우, 대표 문구 제공 시스템은 제1 후보 대표 문구 및 제2 후보 대표 문구 전부를 인기 키워드와 결합하여 제공할 수 있다. 또한, 대표 문구 제공 시스템은 제1 후보 대표 문구 및 제2 후보 대표 문구 중 어느 하나를 랜덤하게 선택하여 인기 키워드에 대한 최종적인 대표 문구로 결정할 수도 있다. 또한, 대표 문구 제공 시스템은 제1 후보 대표 문구 및 제2 후보 대표 문구 전부를 인기 키워드에 대한 최종적인 대표 문구로 결정하지 않고, 인기 키워드만 웹 페이지에 노출할 수 있다. 즉, 도 5 및 도 8에서, 대표 문구 없이 인기 키워드 목록만을 웹 페이지에 노출할 수 있다.
- [0253] 본 발명의 실시 예에 따른 방법들은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.
- [0254] 이상과 같이 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상의 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.
- [0255] 그러므로, 본 발명의 범위는 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

부호의 설명

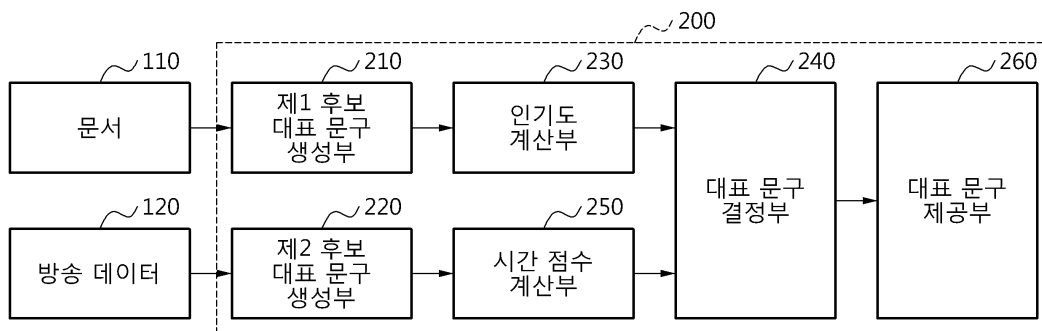
- [0256] 200: 대표 문구 제공 시스템
- 210: 제1 후보 대표 문구 생성부
- 220: 제2 후보 대표 문구 생성부
- 230: 인기도 계산부
- 240: 대표 문구 결정부
- 250: 시간 점수 계산부
- 260: 대표 문구 제공부

도면

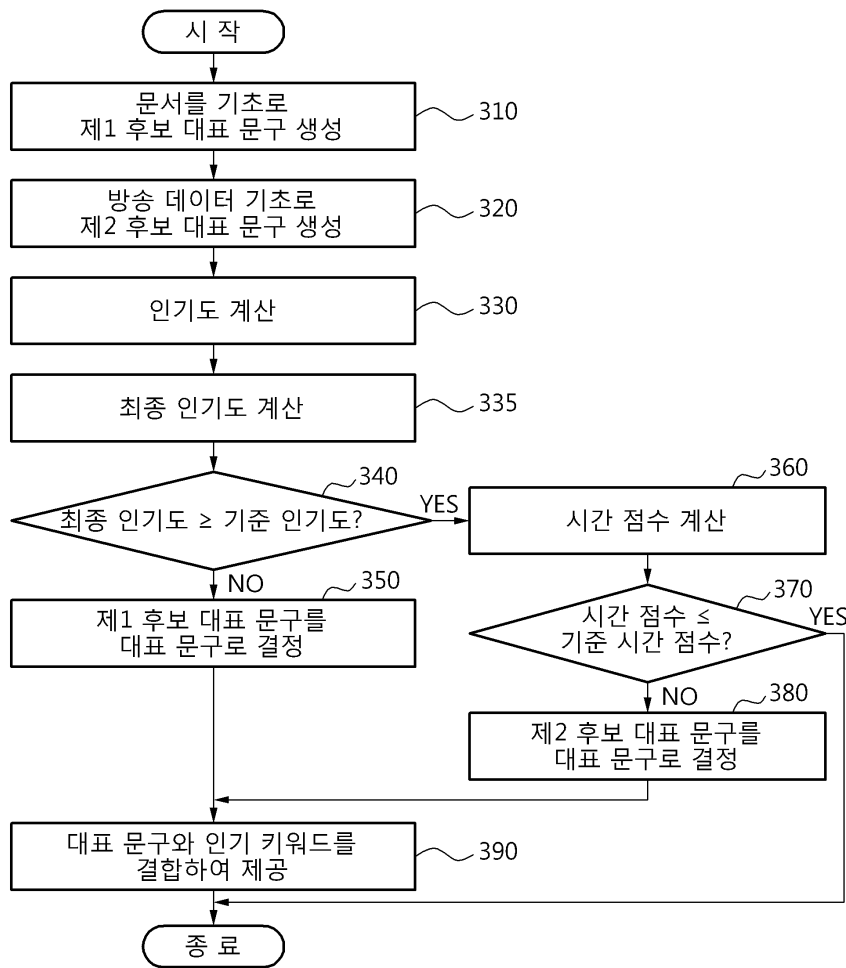
도면1



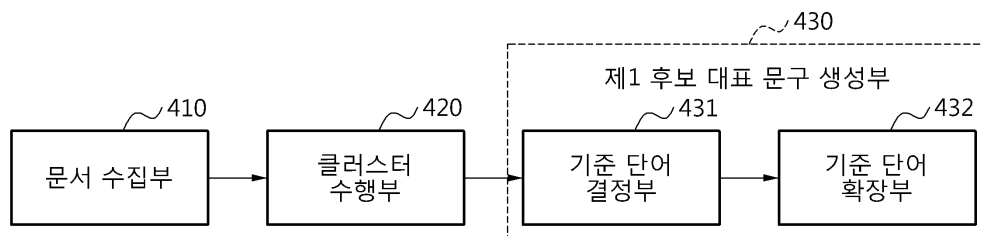
도면2



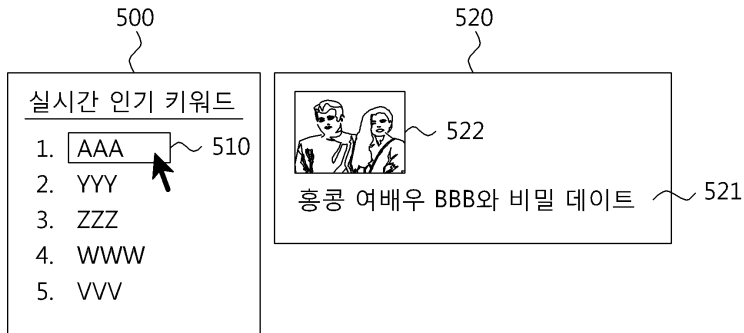
도면3



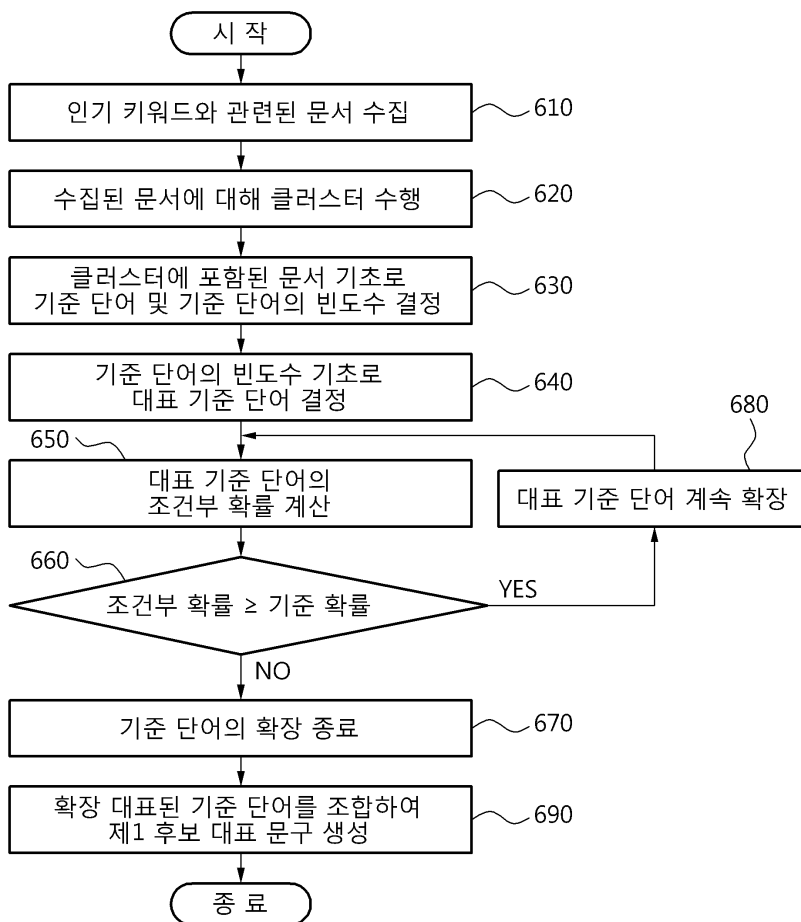
도면4



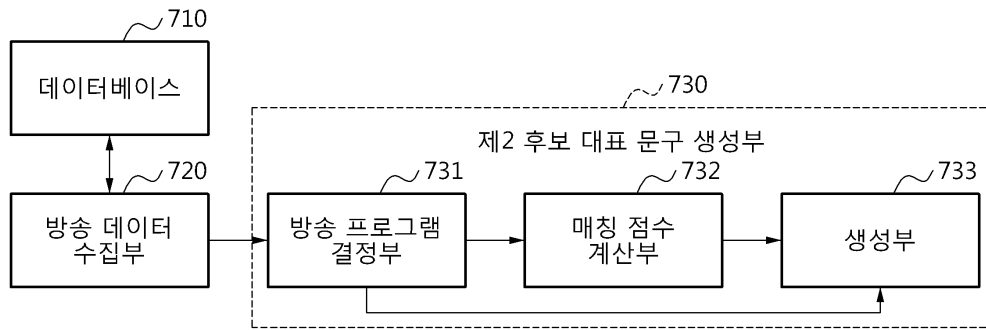
도면5



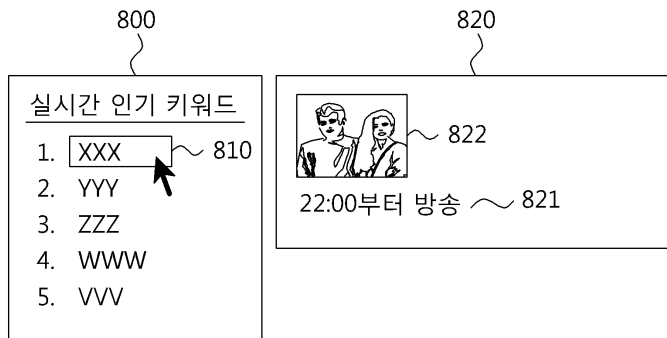
도면6



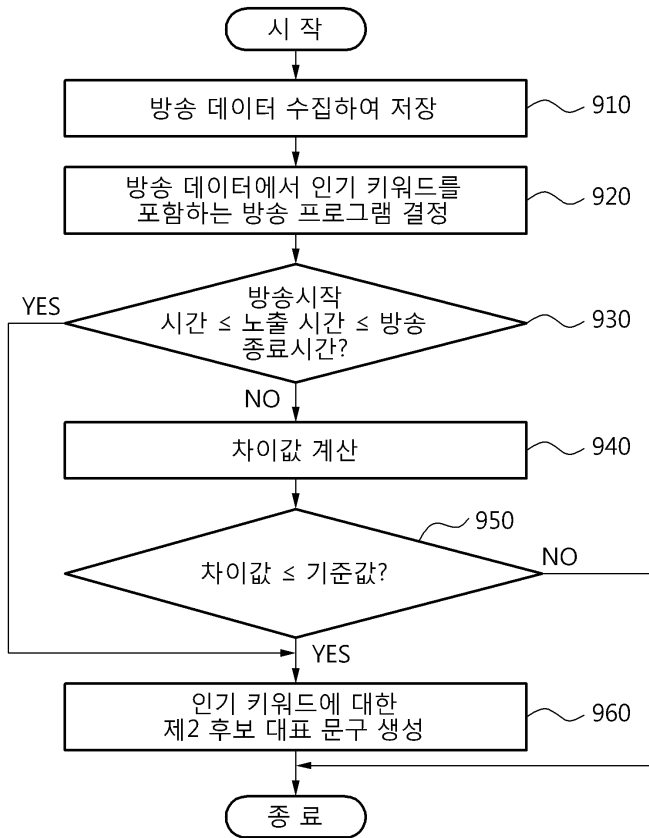
도면7



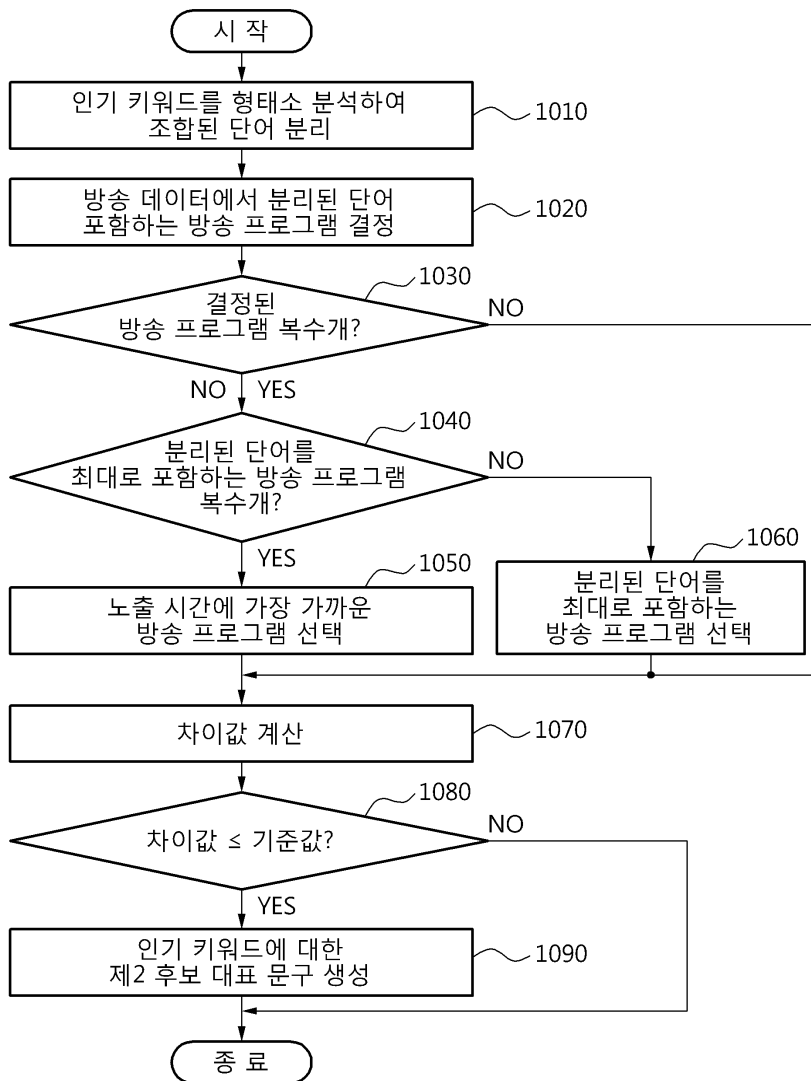
도면8



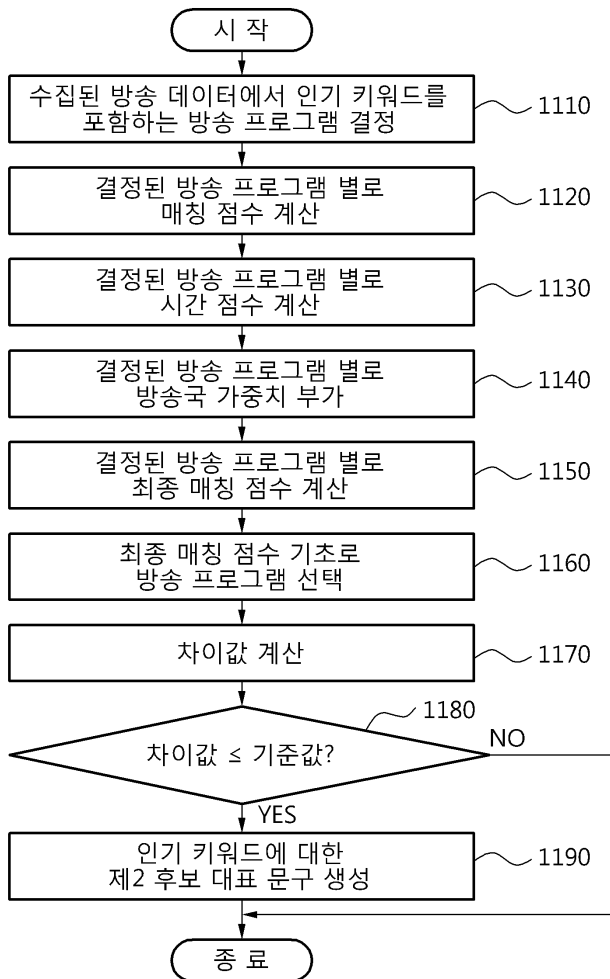
도면9



도면10



도면11



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 29

【변경전】

상기 생성부

【변경후】

상기 제2 후보 대표 문구 생성부