



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2018년12월06일
 (11) 등록번호 10-1925506
 (24) 등록일자 2018년11월29일

(51) 국제특허분류(Int. Cl.)
 G06F 19/00 (2018.01)
 (52) CPC특허분류
 G16H 50/70 (2018.01)
 G06F 17/27 (2013.01)
 (21) 출원번호 10-2017-0170011
 (22) 출원일자 2017년12월12일
 심사청구일자 2017년12월12일
 (56) 선행기술조사문헌
 KR1020160042616 A*
 KR1020170079648 A*
 KR101614736 B1*
 *는 심사관에 의하여 인용된 문헌

(73) 특허권자
 한국과학기술정보연구원
 대전광역시 유성구 대학로 245 (어은동)
 (72) 발명자
 김주현
 대전광역시 유성구 온천로 45, 알파동 2204호(봉명동)
 안인성
 대전광역시 유성구 어은로 57, 119동 1201호(어은동, 한빛아파트)
 한영만
 대전광역시 유성구 장대로 103, 102동 901호 (장대동, 한일유엔아이)
 (74) 대리인
 특허법인가산

전체 청구항 수 : 총 11 항

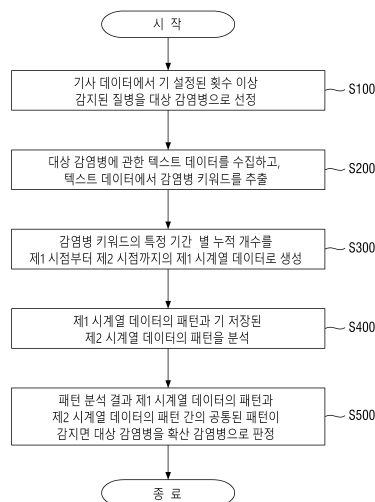
심사관 : 태정범

(54) 발명의 명칭 **감염병 확산 예측 방법 및 장치**

(57) 요약

감염병 확산 예측 방법이 제공된다. 본 발명의 일 실시예에 따른 감염병 확산 예측 방법은 기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하되, 상기 질병은 질병 리스트내에 포함되는 것인, 단계와 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계와 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계와 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계와 상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 단계를 포함할 수 있다.

대표도 - 도4



(52) CPC특허분류

G06Q 50/22 (2018.01)

G06Q 50/26 (2013.01)

G16H 50/80 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 G-17-GT-CA02-S01

부처명 과학기술정보통신부

연구관리전문기관 국가과학기술연구회

연구사업명 2016년도 미래도형 / 실용화형 융합연구단사업

연구과제명 신종 바이러스 감염 대응 융합 솔루션 개발

기 여 율 1/1

주관기관 한국화학연구원

연구기간 2017.08.01 ~ 2018.07.31

명세서

청구범위

청구항 1

확산 예측 서버가 감염병 확산을 예측하는 방법에 있어서,

데이터 베이스 내에 기 저장된 질병 리스트에 포함된 어느 하나의 질병이 기사 데이터 내에 포함된 기사 제목 텍스트에서 기 설정된 횟수 이상 감지된 경우, 상기 질병의 확산 판정의 대상이 되는 대상 감염병으로 선정하는 단계;

상기 기사 제목 텍스트에 상기 선정된 대상 감염병을 포함하지 않은 기사 데이터를 제외한 후, 상기 기사 데이터 내에 포함되는 기사 내용 텍스트를 수집하고, 상기 수집된 기사 내용 텍스트에서 감염병 키워드를 추출하는 단계;

상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계;

상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계; 및

상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 단계를 포함하는,

감염병 확산 예측 방법.

청구항 2

제1 항에 있어서,

기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하는 단계는,

상기 질병이 상기 기사 데이터에서 감지되지 않은 경우, 상기 질병을 상기 질병 리스트에 갱신하는 단계를 포함하는,

감염병 확산 예측 방법.

청구항 3

제1 항에 있어서,

상기 기사 데이터는,

적어도 2이상의 데이터 소스로부터 수집된 것인,

감염병 확산 예측 방법.

청구항 4

제1 항에 있어서,

상기 특정 기간은,

24시간을 단위로 하는 것인,

감염병 확산 예측 방법.

청구항 5

제1 항에 있어서,

상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감

염병 키워드를 추출하는 단계는,
 상기 기사 데이터에서 날짜 데이터, 제목 데이터 및 내용 데이터를 추출하여 분류화 하는 단계를 포함하는,
 감염병 확산 예측 방법.

청구항 6

제1 항에 있어서,
 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는,
 상기 대상 감염병에 대하여 유사도가 최고 값인 감염병 키워드를 메인 키워드로 특정하는 단계를 포함하고,
 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는,
 상기 메인 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함하는,
 감염병 확산 예측 방법.

청구항 7

제1 항에 있어서,
 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는,
 상기 대상 감염병에 대하여 기 설정된 임계치 이상인 유사도를 가지는 감염병 키워드를 서브 키워드로 특정하는 단계를 포함하고,
 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는,
 상기 서브 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함하는,
 감염병 확산 예측 방법.

청구항 8

제1 항에 있어서,
 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계는,
 상기 제1 시계열 데이터를 분석하여 특정 패턴이 감지되는 제1 구간을 도출하는 단계; 및
 상기 제2 시계열 데이터를 분석하여 제2 구간을 도출하는 단계를 포함하되,
 상기 제2 구간은,
 상기 제2 시계열 데이터 상의 특정 패턴이 시작되는 시점부터 과거 감염병 확산 시점 사이의 기간인,
 감염병 확산 예측 방법.

청구항 9

제8 항에 있어서,
 상기 제1 구간 및 상기 제2 구간이 같은 기간으로 판정된 경우, 상기 제1 시계열 데이터 상의 특정 시점에서 제1 구간에 해당하는 기간 이후인 시점을 실제 감염병 확산 시점으로 판정하는 단계를 더 포함하는,
 감염병 확산 예측 방법.

청구항 10

하나 이상의 인스트럭션이 저장되는 메모리; 및

상기 저장된 인스트럭션을 실행하는 프로세서를 포함하되,

상기 프로세서는,

데이터 베이스 내에 기 저장된 질병 리스트에 포함된 어느 하나의 질병이 기사 데이터 내에 포함된 기사 제목 텍스트에서 기 설정된 횟수 이상 감지된 경우, 상기 질병의 확산 판정의 대상이 되는 대상 감염병으로 선정하고,

상기 기사 제목 텍스트에 상기 선정된 대상 감염병을 포함하지 않은 기사 데이터를 제외한 후, 상기 기사 데이터 내에 포함되는 기사 내용 텍스트를 수집하고, 상기 수집된 기사 내용 텍스트에서 감염병 키워드를 추출하고,

상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하고,

상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하고,

상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 것인,

전자 장치.

청구항 11

컴퓨팅 장치와 결합하여 실행되는 경우에, 상기 컴퓨팅 장치가,

데이터 베이스 내에 기 저장된 질병 리스트에 포함된 어느 하나의 질병이 기사 데이터 내에 포함된 기사 제목 텍스트에서 기 설정된 횟수 이상 감지된 경우, 상기 질병의 확산 판정의 대상이 되는 대상 감염병으로 선정하는 단계;

상기 기사 제목 텍스트에 상기 선정된 대상 감염병을 포함하지 않은 기사 데이터를 제외한 후, 상기 기사 데이터 내에 포함되는 기사 내용 텍스트를 수집하고, 상기 수집된 기사 내용 텍스트에서 감염병 키워드를 추출하는 단계;

상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계;

상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계; 및

상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 단계를 실행시키도록 기록 매체에 저장된,

컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명은 감염병 확산 예측 방법 및 장치에 관한 것이다. 보다 자세하게는, 감염병과 관련된 기사 데이터를 이용하여 대상 감염병이 확산될지 여부를 판정하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 급격한 기후 변화 및 인구의 활발한 이동 등의 요인으로 인하여 신종 감염병 또는 재출현 감염병의 발생 및 확산이 증가하는 추세이다. 이와 같은 감염병이 확산되면, 사회적, 경제적 측면에서 피해가 막대하기 때문에, 정부나 보건 당국은 감염병이 확산되기전에 이에 대한 대비책을 마련하기 위하여 많은 노력을 한다.

[0003] 그러나, 상기와 같은 감염병의 출현 진조를 사람이나 정부 기관이 일일이 판단하는 것은 용이하지 않은 실정이며, 빅데이터를 이용하여 감염병의 확산 여부를 판단하기 위해서는 데이터를 선별적으로 수집할 수 있는 구체적

인 수단이 존재하지 않는 것이 현실이다.

[0004] 다만, 최근에는 정보통신기술의 발전과 인터넷 서비스의 확산 등으로 인하여 수많은 데이터가 웹 상에 존재하기 때문에, 이를 수집하고 관리 및 분석하는 과정을 구체화한다면, 감염병의 확산을 효과적으로 예측할 수 있는 여지가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국공개특허 제10-2013-0010789호

발명의 내용

해결하려는 과제

[0006] 본 발명이 해결하고자 하는 기술적 과제는, 가까운 과거로부터 미래까지 이슈가 될 만한 감염병을 탐지하는 방법 및 장치를 제공하는 것이다.

[0007] 본 발명이 해결하고자 하는 다른 기술적 과제는, 이슈가 될 만한 감염병을 질병 리스트에 갱신하는 방법 및 장치를 제공하는 것이다.

[0008] 본 발명이 해결하고자 하는 또 다른 기술적 과제는, 상기 탐지된 감염병이 확산될 지 여부를 판단하는 방법 및 장치를 제공하는 것이다.

[0009] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명의 기술분야에서의 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0010] 상기 기술적 과제를 해결하기 위한, 본 발명의 일 실시예에 따른 감염병 확산 예측 방법은, 기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하되, 상기 질병은 질병 리스트내에 포함되는 것인 단계, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계, 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계, 상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 단계를 포함할 수 있다.

[0011] 일 실시예에서, 기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하는 단계는, 상기 질병이 상기 기사 데이터에서 감지되지 않은 경우, 상기 질병을 상기 질병 리스트에 갱신하는 단계를 포함할 수 있다.

[0012] 일 실시예에서, 감염병 확산 예측 방법의 텍스트 데이터는, 적어도 2이상의 데이터 소스로부터 수집된 것일 수 있다.

[0013] 일 실시예에서, 감염병 확산 예측 방법의 특정 기간은, 24시간을 단위로 하는 것 일 수 있다.

[0014] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는, 상기 기사 데이터에서 날짜 데이터, 제목 데이터 및 내용 데이터를 추출하여 분류화 하는 단계를 포함할 수 있다.

[0015] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는, 상기 대상 감염병에 대하여 유사도가 최고 값인 감염병 키워드를 메인 키워드로 특정하는 단계를 포함하고, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는, 상기 메인 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함할 수 있다.

[0016] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트

데이터에서 감염병 키워드를 추출하는 단계는, 상기 대상 감염병에 대하여 기 설정된 임계치 이상인 유사도를 가지는 감염병 키워드를 서브 키워드로 특정하는 단계를 포함하고, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는, 상기 서브 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함할 수 있다.

[0017] 일 실시예에서, 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계는, 상기 제1 시계열 데이터를 분석하여 특정 패턴이 감지되는 제1 구간을 도출하는 단계, 상기 제2 시계열 데이터를 분석하여 제2 구간을 도출하는 단계를 포함하되, 상기 제2 구간은, 상기 제2 시계열 데이터 상의 특정 패턴이 시작되는 시점부터 과거 감염병 확산 시점 사이의 기간일 수 있다.

[0018] 일 실시예에서, 상기 제1 구간 및 상기 제2 구간이 같은 기간으로 판정된 경우, 상기 제1 시계열 데이터 상의 특정 시점에서 제1 구간에 해당하는 기간 이후인 시점을 실제 감염병 확산 시점으로 판정하는 단계를 더 포함할 수 있다.

[0019] 상술한 기술적 과제를 해결하기 위한 본 발명의 다른 실시예에 따른 전자 장치는, 하나 이상의 인스트럭션이 저장되는 메모리, 상기 저장된 인스트럭션을 실행하는 프로세서를 포함하되, 상기 프로세서는, 질병 리스트에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하고, 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하고, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하고, 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하고, 상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 것일 수 있다.

[0020] 상술한 기술적 과제를 해결하기 위한 본 발명의 또 다른 실시예에 따른 감염병 확산 예측 컴퓨터 프로그램은, 컴퓨팅 장치와 결합하여, 질병 리스트에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하는 단계, 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계, 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계, 상기 분석 결과, 상기 제1 시계열 데이터의 패턴과 상기 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 상기 대상 감염병이 확산 감염병인 것으로 판정하는 단계를 실행시키도록 기록 매체에 저장될 수 있다.

도면의 간단한 설명

- [0021] 도 1은 본 발명의 일 실시예에 따른 감염병 확산 예측 시스템의 구성도이다.
- 도 2는 본 발명의 몇몇 실시예에서 참조되는 제1 시계열 데이터를 설명하기 위한 도면이다.
- 도 3은 본 발명의 몇몇 실시예에서 참조되는 제2 시계열 데이터를 설명하기 위한 도면이다.
- 도 4는 본 발명의 일 실시예에 따른 감염병 확산 예측 방법의 순서도이다.
- 도 5는 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S100)를 설명하기 위한 상세 순서도이다.
- 도 6은 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S200)를 설명하기 위한 상세 순서도이다.
- 도 7은 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S400)를 설명하기 위한 상세 순서도이다.
- 도 8 내지 도 11은 본 발명의 몇몇 실시예에서 참조되는, 3차원 키워드 모델 스펙트럼을 설명하기 위한 도면이다.
- 도 12는 도 8 내지 도 11에 도시된 3차원 키워드 모델 스펙트럼의 세부 구성 요소를 설명하기 위한 도면이다.
- 도 13 내지 도 16은 본 발명의 몇몇 실시예에서 참조되는 시계열 데이터를 설명하기 위한 도면이다.
- 도 17은 도 13 내지 도 16에 도시된 시계열 데이터의 실험값 등을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예들을 상세히 설명한다. 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될

수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

- [0023] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있다. 또 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다. 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.
- [0024] 명세서에서 사용되는 "포함한다 (comprises)" 및/또는 "포함하는 (comprising)"은 언급된 구성 요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 구성 요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다.
- [0025] 본 명세서에 대한 설명에 앞서, 본 명세서에서 사용되는 몇몇 용어들에 대하여 명확하게 하기로 한다.
- [0026] 본 명세서에서, 시계열 데이터란, 감염병이 확산되는지 여부를 판정하기 위하여 상기 감염병에 관련된 키워드에 시간적 요소를 반영하여 생성한 데이터를 가리킨다. 구체적으로, 웹 상에 존재하는 감염병에 관련된 기사 데이터를 수집하고, 상기 기사 데이터에서 감염병과 연관된 키워드를 추출한다. 키워드가 추출되면 상기 키워드의 출현 빈도를 특정 기간 별로 나타내고 이는 임의의 기간인 제1 시점부터 제2 시점까지의 시계열 데이터로 생성될 수 있다.
- [0027] 상기 시계열 데이터를 시간에 흐름에 따라 2차원 그래프로 도시하면, 제1 시점부터 제2 시점까지 구간에서 특정한 패턴을 감지할 수 있을 것이다. 본 발명의 몇몇 실시예에서는 상기 특정한 패턴을 감지하여 대상 감염병이 확산되는지 여부를 판단하는 것이다.
- [0028] 한편, 상기 시계열 데이터를 구성하는 세부 구성 요소에 대한 정의는 첨부된 도면을 설명하는 과정에서 필요에 따라 도입함과 동시에 개별적으로 설명하도록 한다.
- [0029] 이하, 본 발명의 몇몇 실시예들에 대하여 첨부된 도면에 따라 상세하게 설명한다.
- [0030] 도 1은 본 발명의 일 실시예에 따른 감염병 확산 예측 시스템의 구성도이다. 이하, 도 1을 참조하여 본 실시예에 따른 감염병 확산 예측 시스템의 세부 구성 및 동작을 설명한다.
- [0031] 감염병 확산 예측 시스템은 제1 시계열 데이터 DB, 제2 시계열 데이터 DB, 사용자 단말(200) 및 확산 예측 서버(100)를 포함할 수 있다. 단, 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 구성 요소가 추가되거나 삭제될 수 있음은 물론이다. 또한, 도 1에 도시된 감염병 확산 예측 시스템의 각각의 구성 요소들은 기능적으로 구분되는 기능 요소들을 나타낸 것으로서, 적어도 하나의 구성 요소가 실제 물리적 환경에서는 서로 통합되는 형태로 구현될 수도 있음에 유의한다.
- [0032] 확산 예측 서버(100)는 제1 시계열 데이터 및 제2 시계열 데이터를 분석하여 사용자 단말(200)에 감염병 확산 여부에 대한 정보를 제공할 수 있다.
- [0033] 한편, 확산 예측 서버(100)는 질병 사전 DB 에서 수집된 질병에 관련한 데이터를 이용하여 질병 리스트를 생성할 수 있다. 질병의 일종인 감염병의 확산 여부를 판단하기에 앞서서, 미리 질병 리스트에서 이슈가 될 만한 감염병을 질병 리스트에서 추출 또는 갱신하여 감염병 확산 여부를 판단하는 시간을 단축할 수 있다.
- [0034] 확산 예측 서버(100)가 제1 시계열 데이터를 생성하기 위하여, 감염병과 관련된 기사 데이터를 수집한다. 상기 기사 데이터를 수집하는 수단을 예로 들면, 크롤러(Crawler) 모듈이 사용될 수 있다.
- [0035] 또한, 상기 수집된 기사 데이터에서 날짜, 제목 및 내용 등의 내용을 추출할 수 있다. 상기 추출된 날짜, 제목 및 내용 등에 관한 데이터가 제1 시계열 데이터에 반영된다. 예를 들어, 상기 기사 데이터에서 날짜, 제목 및 내용 등의 내용을 추출하기 위하여 파서(Parser) 모듈이 사용될 수 있다.
- [0036] 확산 예측 서버(100)가 질병 사전 DB 및 기사 DB에서 추출된 데이터를 이용하여 제1 시계열 데이터를 생성하면, 제2 시계열 데이터 DB(2000)에 기 저장된 제2 시계열 데이터와 비교 분석하여 대상 감염병이 확산될 지 여부를 판정한다. 상기 확산 판정에 대한 자세한 로직은 도 4를 참조하여 자세히 설명한다. 예를 들어, 상기 제2 시계열 데이터는 질병 관리 본부와 같은 기관이 실제 확산된 감염병에 관한 분석 내용을 시간에 따라 데이터 베이스화 한 데이터에 해당한다.

- [0037] 확산 예측 서버(100)는 제1 시계열 데이터와 제2 시계열 데이터의 공통된 패턴을 감지하여 대상 감염병이 확산되는지 여부를 판단하고, 사용자 단말(200)에 제공할 수 있다. 또한, 상기 대상 감염병에 관련된 키워드를 3차원 키워드 스트럭처로 시각화 하여 사용자 단말(200)을 통하여 사용자에게 제공할 수 있다.
- [0038] 도 2는 본 발명의 몇몇 실시예에서 참조되는 제1 시계열 데이터를 설명하기 위한 도면이다. 이하, 도 2를 참조하여 제1 시계열 데이터의 세부 구성 요소를 설명한다.
- [0039] 제1 시계열 데이터는, 웹 상의 기사 데이터에서 추출된 감염병에 관련된 키워드에 시계열적 요소를 반영하여, 제1 시계열 데이터 DB(1000)에 저장된다. 예를 들어, 질병 사전 DB에서 제공되는 질병 리스트에 포함된 감염병을 특정하고, 상기 감염병에 관한 기사 데이터를 수집한다. 상기 수집된 기사 데이터는 기사 DB에 저장된다.
- [0040] 예를 들어 제1 시계열 데이터는, 웹 상에 존재하는 ProMED mail, World Health Organization, CIRAP, Centre for Health Protection 등의 데이터 소스에서 수집된 기사 데이터를 이용하여 생성될 수 있다. 이는, 다양한 데이터 소스를 이용하여 감염병에 관련된 키워드를 충분히 수집할 수 있도록 한다.
- [0041] 도 3은 본 발명의 몇몇 실시예에서 참조되는 제2 시계열 데이터를 설명하기 위한 도면이다. 이하, 도 3을 참조하여 제2 시계열 데이터의 세부 구성 요소를 설명한다.
- [0042] 제2 시계열 데이터는, 실제 감염병이 확산된 경우, 확산되기 이전의 텍스트 데이터에서 감염병에 관련된 키워드를 추출하고, 상기 키워드에 시계열적 요소를 반영하여 생성된 데이터를 가리킨다. 상기 제2 시계열 데이터는 제2 시계열 데이터 DB(2000)에 저장된다.
- [0043] 예를 들어, 제2 시계열 데이터는, 질병 관리 본부 또는 사설 기관의 DB에 저장된 데이터를 가리킬 수 있다. 제2 시계열 데이터에는, 과거 실제로 확산된 감염병의 명칭 및 날짜가 포함되어 있고, 감염병이 확산되기 전조의 패턴이 제2 시계열 데이터에 반영되어 있다.
- [0044] 도 4는 본 발명의 일 실시예에 따른 감염병 확산 예측 방법의 순서도이다. 단, 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 단계가 추가되거나 삭제될 수 있음은 물론이다. 이하, 도 4를 참조하여 본 실시예에 따른 감염병 확산 예측 방법을 설명한다.
- [0045] 이하, 본 발명의 실시예에 따른 감염병 확산 예측 방법의 각 단계는, 전자 장치에 의해 수행될 수 있다. 예를 들어, 상기 전자 장치는 감염병 확산 예측 장치일 수 있다. 다만, 설명의 편의를 위해, 상기 감염병 확산 예측 방법에 포함되는 각 단계의 동작 주체는 그 기체가 생략될 수도 있다. 또한, 감염병 확산 예측 방법 방법의 각 단계는 감염병 확산 예측 방법 소프트웨어가 프로세서에 의해 실행됨으로써, 감염병 확산 예측 방법 장치에서 수행되는 인스트럭션일 수 있다.
- [0046] 기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정한다(S100). 이는, 최근 들어서 자주 언급된 질병을 감염병으로 간주하고, 상기 감염병의 확산 여부를 판정을 하기 위함이다.
- [0047] 상기 감지된 질병은 질병 리스트 내에 포함된 것일 수 있다. 이는 질병 사전 DB를 통하여 질병 리스트를 생성하고, 상기 질병 리스트에 포함된 질병이 기사 제목 또는 기사 내용에서 자주 언급되면 상기 질병을 대상 감염병으로 선정할 수 있다.
- [0048] 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출한다(S200).
- [0049] S200 단계에서는, 대상 감염병과 관련된 키워드를 추출하기 위하여, 웹 상에 존재하는 기사 데이터에서 텍스트에 대한 데이터를 수집한다. 예를 들어, 수집된 텍스트 데이터에서 텍스트 마이닝 기법을 통하여 대상 감염병에 관련된 단어를 추출할 수 있다.
- [0050] 일 실시예에서, 상기 텍스트 데이터는 적어도 2 이상의 데이터 소스로부터 수집된 것일 수 있다. 추출되는 키워드의 다양성 및 정확성을 향상 시키기 위해 여러 가지의 데이터 소스에서 텍스트 데이터를 수집할 필요가 있다. 예를 들어, 상기 데이터 소스는 ProMED mail, World Health Organization, CIRAP, Centre for Health Protection 등의 다양한 질병의 관련 기사를 접할 수 있는 웹 상에 존재하는 빅 데이터에 해당할 수 있다.
- [0051] 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성한다(S300). 제1 시점부터 제2 시점까지 구간 내에서 시간에 따른 특정 패턴을 도출하기 위하여 상기 감염병 키워드에 시적 요소를 반영할 수 있다.

- [0052] 일 실시예에서, 상기 특정 기간은 24시간을 단위로 하는 것일 수 있다. 상기 특정 기간은 사용자의 설정 등에 의하여 조정이 가능할 수 있으나, 일반적으로 기사 데이터가 일 별로 수집되는 것이 통상적이므로, 상기 특정 기간을 24시간으로 설정할 수 있다.
- [0053] 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석한다(S400). 상기 제2 시계열 데이터는 도 1 내지 도 2를 참조하여 상술한 바와 같이 제1 시계열 데이터의 패턴과 비교하기 위하여 실제 확산된 감염병에 관한 분석 내용을 질병 관리 본부 또는 사설 기관의 DB에 기 저장된 데이터일 수 있다.
- [0054] 상기 분석 대상이 되는 패턴에 대하여 설명한다. 본 발명의 몇몇 실시예에서 분석 대상이 되는 패턴에는 다양한 요소가 포함될 수 있음에 유의한다.
- [0055] 예를 들어, 상기 패턴은 특정 감염병 키워드의 일 별 누적 횟수에 관한 패턴을 가리킬 수 있다.
- [0056] 또한, 다른 예로서, 상기 패턴은 특정 감염병 키워드의 관련도 등을 수치화 한 것을 가리킬 수 있다.
- [0057] 상기 분석 결과, 제1 시계열 데이터의 패턴과 제2 시계열 데이터의 패턴 간의 공통된 패턴이 감지되면, 대상 감염병이 확산 감염병인 것으로 판정한다(S500). 이는, 상기 제1 시계열 데이터의 패턴과 제2 시계열 데이터의 패턴 간의 유사성을 확인하여 미리 저장된 과거 데이터를 기반으로 대상 감염병이 확산될지 여부를 판단하는 것이다.
- [0058] 도 5는 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S100)를 설명하기 위한 상세 순서도이다. 단, 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 단계가 추가되거나 삭제될 수 있음은 물론이다. 이하, 도 5를 참조하여 S100 단계의 세부 구성을 설명한다.
- [0059] 특정 질병에 관한 기사가 기사 데이터 내에 존재하는지 여부를 판단한다(S101). 기사 데이터를 이용하여 특정 질병이 감염병인지 여부를 판단하고, 상기 감염병이 확산되는지 여부를 판단하기 때문이다.
- [0060] 상기 특정 질병에 관한 기사가 기사 데이터 내에 존재하지 않는다면(S102), 상기 특정 질병을 리스트화 하여 추후에 사용할 필요가 있다. 이를 위하여, 상기 특정 질병을 질병 리스트에 갱신할 수 있다(S104).
- [0061] 상기 특정 질병에 관한 기사가 기사 데이터 내에 존재하는지 여부를 판단하고(S102), 기 설정된 횟수 이상이 감지되었는지를 판단한다(S103). 상기 횟수는 사용자의 설정에 따라 다양한 수치를 적용할 수 있다.
- [0062] 만약, 상기 특정 질병에 관한 기사가 상기 기사 데이터에서 기 설정된 횟수 이상 감지되면, 상기 특정 질병을 확산 판단의 대상이 되는 대상 감염병으로 선정한다(S105).
- [0063] 반대로, 상기 특정 질병에 관한 기사가 기사 데이터에서 기 설정된 횟수 이상으로 감지되지 않는다면, 상기 특정 질병을 대상 감염병으로 선정하지 않을 것이다(S106).
- [0064] 일 실시예에서, 기사 데이터에서 기 설정된 횟수 이상 감지된 질병을 대상 감염병으로 선정하는 단계는, 상기 질병이 상기 기사 데이터에서 감지되지 않은 경우, 상기 질병을 상기 질병 리스트에 갱신하는 단계를 포함할 수 있다. 이는, 뉴스 기사 등에서 신종 질병이 언급되는 경우, 상기 신종 질병을 리스트화 하고, 추후에 대상 감염병으로 선정하여, 대상 감염병이 확산되는지 여부를 신속히 판단할 수 있는 장점이 있다.
- [0065] 도 6은 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S200)를 설명하기 위한 상세 순서도이다. 단, 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 단계가 추가되거나 삭제될 수 있음은 물론이다. 이하, 도 6을 참조하여, S200 단계의 세부 구성을 설명한다.
- [0066] 본 실시예는 대상 감염병에 관련된 감염병 키워드를 추출하여 이에 대한 유사도에 따라 패턴 분석의 대상이 되는 감염병 키워드를 선정하기 위하여 적용됨을 유의한다.
- [0067] 수집된 텍스트 데이터에서 감염병 키워드를 추출하면(S201), 상기 추출된 대상 감염병에 대한 감염병 키워드의 유사도를 산정한다(S202). 상기 유사도는 상기 대상 감염병과 감염병 키워드 간의 관련 정도를 나타낼 수 있다.
- [0068] 유사도 산정이 완료되면, 유사도의 수치에 따른 감염병 키워드를 선별한다. 특정 감염병 키워드의 유사도가 최고 값에 해당하면(S203), 상기 추출된 감염병 키워드를 메인 키워드로 특정한다(S204). 따라서, 추후 패턴 분석의 대상이 되는 제1 시계열 데이터의 구성은 상기 메인 키워드가 될 수 있다. 상기 메인 키워드의 특정 후, 메인 키워드를 이용하여 제1 시계열 데이터를 생성한다(S304).
- [0069] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트

데이터에서 감염병 키워드를 추출하는 단계는, 상기 대상 감염병에 대하여 유사도가 최고 값인 감염병 키워드를 메인 키워드로 특정하는 단계를 포함하고, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는, 상기 메인 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함할 수 있다. 이는, 대상 감염병과 가장 관련도가 높은 감염병 키워드를 이용하여 명확한 패턴을 도출할 수 있고, 상기 명확한 패턴에 의하여 감염병 확산 판단의 정확성을 향상시킬 수 있다.

- [0070] 유사도 산정이 완료되고, 산정된 유사도의 값이 최고 값에 해당하지 않는 경우를 설명한다. 상기 유사도의 값이 터무니 없이 낮은 경우에는 무의미한 데이터에 해당할 것이다. 따라서, 본 실시예에서는 유사도 값에 대한 임계치를 설정하여 기 설정된 임계치 이상인 경우에만 분석 대상인 키워드로 특정할 수 있다. 만약, 유사도가 기 설정된 임계치 이상에 해당된다면, 상기 추출된 감염병 키워드를 서브 키워드로 특정할 수 있다(S205). 상기 서브 키워드의 특정 후, 서브 키워드를 이용하여 제1 시계열 데이터를 생성한다(S305).
- [0071] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는, 상기 대상 감염병에 대하여 기 설정된 임계치 이상인 유사도를 가지는 감염병 키워드를 서브 키워드로 특정하는 단계를 포함하고, 상기 감염병 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계는, 상기 서브 키워드의 특정 기간 별 누적 개수를 제1 시점부터 제2 시점까지의 제1 시계열 데이터로 생성하는 단계를 포함할 수 있다.
- [0072] 일 실시예에서, 상기 기사 데이터에서 상기 대상 감염병에 관한 텍스트 데이터를 수집하고, 상기 수집된 텍스트 데이터에서 감염병 키워드를 추출하는 단계는, 상기 기사 데이터에서 날짜 데이터, 제목 데이터 및 내용 데이터를 추출하여 분류화 하는 단계를 포함할 수 있다. 이는, 상기 감염병 키워드의 유사도 측정 및 시간적 요소를 반영하고, 방대한 데이터의 관리를 용이하게 할 수 있다.
- [0073] 도 7은 도 4에 도시된 감염병 확산 예측 방법을 구성하는 세부 단계(S400)를 설명하기 위한 상세 순서도이다. 단, 이는 본 발명의 목적을 달성하기 위한 바람직한 실시예일 뿐이며, 필요에 따라 일부 단계가 추가되거나 삭제될 수 있음은 물론이다. 이하, 도 7을 참조하여, S400 단계의 세부 구성을 설명한다.
- [0074] 제1 시계열 데이터를 분석하여 특정 패턴을 감지하고, 특정 패턴이 감지되는 구간을 제1 기간으로 특정한다(S401). 상기 제1 기간을 기준으로 대상 감염병이 확산될 지 여부를 판단한다.
- [0075] 제2 시계열 데이터에서 특정 패턴의 시작 시점 및 과거 감염병 확산 시점을 판단한다(S402). 이는, 실제 감염병이 확산된 케이스를 분석한 제2 시계열 데이터를 이용하여 특정 패턴이 감지된 후 어느 정도 이후에 감염병이 확산되었는지 여부를 판단하기 위한 것이다.
- [0076] 제2 시계열 데이터에서 특정 패턴의 시작 시점부터 과거 감염병 확산 시점의 사이 구간을 제2 기간으로 특정하고(S403), 제1 기간과 상기 제2 기간을 비교한다(S404). 만약, 제1 시계열 데이터의 특정 패턴과 제2 시계열 데이터의 특정 패턴 간에 공통점이 발견되고, 상기 제1 기간과 상기 제2 기간이 같은 기간에 해당한다면, 제1 시계열 데이터 상에서 예측되는 감염병 확산 시점은 제2 기간 종료 직후로 추정할 수 있을 것이다.
- [0077] 제1 기간과 제2 기간이 같은 기간에 해당한다면(S405), 제1 시계열 데이터 상의 특정 시점에서 제1 기간 이후인 시점을 실제 감염병 확산 시점으로 판정한다(S406).
- [0078] 제1 기간과 제2 기간이 같은 기간에 해당하지 않는다면(S405), 제1 시계열 데이터 상의 특정 시점에서 제1 기간 이후인 시점과 일정 범위 내인 시점을 실제 감염병 확산 시점으로 추정할 수 있다(S407).
- [0079] 일 실시예에서, 상기 제1 시계열 데이터의 패턴과 기 저장된 제2 시계열 데이터의 패턴을 분석하는 단계는, 상기 제1 시계열 데이터를 분석하여 특정 패턴이 감지되는 제1 구간을 도출하는 단계, 상기 제2 시계열 데이터를 분석하여 제2 구간을 도출하는 단계를 포함하되, 상기 제2 구간은 상기 제2 시계열 데이터 상의 특정 패턴이 시작되는 시점부터 과거 감염병 확산 시점 사이의 기간일 수 있다.
- [0080] 일 실시예에서, 상기 제1 구간 및 상기 제2 구간이 같은 기간으로 판정된 경우, 상기 제1 시계열 데이터 상의 특정 시점에서 제1 구간에 해당하는 기간 이후인 시점을 실제 감염병 확산 시점으로 판정하는 단계를 포함할 수 있다. 이는, 제1 시계열 데이터에서 제2 시계열 데이터 상의 특정 패턴과 같은 패턴이 감지되면 같은 기간 이후에 대상 감염병이 확산되는 것으로 추정할 수 있도록 한다.
- [0081] 도 8 내지 도 11은 본 발명의 몇몇 실시예에서 참조되는, 3차원 키워드 모델 스트럭처를 설명하기 위한 도면이

다. 이하, 도 8 내지 도 11을 참조하여 3차원 키워드 모델 스트럭처를 설명한다.

- [0082] 도 8을 참조하면, 기사 데이터에서 추출된 감염병 키워드에 대한 데이터가 3차원 키워드 모델 스트럭처에 시각화 된 것을 확인할 수 있다. 3차원 그래프의 중심축과의 거리는 대상 감염병과 감염병 키워드에 대한 관련도를 가리킬 수 있다.
- [0083] 도 9를 참조하면, 상기 3차원 그래프 좌표상에 감염병 키워드에 대한 텍스트가 표시된 것을 확인할 수 있다. 이와 같은 텍스트 시각화 과정에 의하여 어떠한 감염병 키워드가 유의미한 것인지 사용자에게 알릴 수 있다. 이를 확인하여, 다시 사용자의 선택에 따라 감염병 키워드를 필터링 하여 대상 감염병의 확산 여부 판단의 정확성을 높일 수 있다.
- [0084] 도 10을 참조하면, 도 9에서 도시된 텍스트에서 유사도가 기 설정된 임계치 이상인 텍스트만 시각화 된 것을 확인할 수 있다. 본 시각화 과정에서는 Word2Vec 알고리즘을 사용한 텍스트 데이터 분석 모듈의 최적 파라미터 계산을 통한 성능 향상을 확인할 수 있었다.
- [0085] 마지막으로 도 11을 참조하면, 본 발명의 몇몇 실시예가 적용되어 사용자에게 제공되는 인터페이스 화면을 확인할 수 있다. 이는, 도 12를 동시에 참조하여 설명한다.
- [0086] 예를 들어, 중국에서 감염병의 확산이 염려되는 조류 인플루엔자 H7N9과 관련성이 깊은 단어들을 검색한 결과 Guangdong, China, Shanghai, Jiangsu 등 중국의 지역명과 조류 인플루엔자인 H5N6, H9N2 등의 결과가 도출되었다.
- [0087] 도 13 내지 도 16은 본 발명의 몇몇 실시예에서 참조되는 시계열 데이터를 설명하기 위한 도면이다. 이는, 도 17을 동시에 참조하여, 시계열 데이터의 실험값 등과 함께 설명한다.
- [0088] 도 13 내지 도 16에 도시된 바와 같이, 2차원 그래프의 가로축에는 시간 요소가 반영된다. 다만, 상기 2차원 그래프의 세로축에는 감염병 키워드의 누적 횟수 및 유사도 외의 감염병 키워드의 패턴을 비교하기 위한 다양한 수치가 반영될 수 있음을 유의한다.
- [0089] 먼저 도 13을 참조하여 2014년도의 제1 시계열 데이터와 제2 시계열 데이터의 패턴을 비교하여 설명한다. 예를 들어, 한국에서 확산될 가능성이 있는 인플루엔자 감염병이 확산되는지 여부를 판단하기 위하여, 도 13에 도시된 바와 같이, 제1 시계열 데이터(KOR_INFLUENZA)가 생성되었고, 생성된 제1 시계열 데이터의 패턴과 비교하기 위하여 복수의 제2 시계열 데이터(Guangdong, Zhejiang, Jiangsu, China)가 생성된 것을 확인할 수 있다.
- [0090] 도 13을 참조하면, 상기 복수의 제2 시계열 데이터의 2013.11.27 부터 2014.2.5 까지 기간의 수치는 대부분 피크치를 나타낸다. 또한 제1 시계열 데이터의 2014.2.5 부터 2014.4.16까지 기간의 수치는 피크치를 나타낸다.
- [0091] 상기 제1 시계열 데이터와 상기 제2 시계열 데이터의 피크치가 나타나는 패턴을 비교하여 제1 시계열 데이터의 감염병 확산 여부를 판단할 수 있다. 예를 들어, 상기 복수의 제2 시계열 데이터의 패턴이 종료되는 시점인 2014.2.5에 실제 감염병이 확산되었다면, 제1 시계열 데이터의 패턴이 종료되는 시점인 2014.4.16에는 한국에서 인플루엔자 감염병이 확산되는 것을 예측할 수 있을 것이다.
- [0092] 다음으로, 도 14 내지 도 16를 참조하면, 2015년도, 2016년도 및 2017년도 에서 실제 먼 과거의 제2 시계열 데이터와 가까운 과거의 제1 시계열 데이터를 비교하여 대상 감염병이 확산되었는지 여부를 판단하기 위한 그래프를 확인할 수 있다.
- [0093] 도 17을 참조하면, 상기 시계열 데이터의 실험값의 다양한 도출 결과를 확인할 수 있다.
- [0094] 도 17에서는, 제2 시계열 데이터의 실험값을 지역별로 키워드의 발생 빈도를 색상의 농도로 표현한 것을 확인할 수 있다.
- [0095] 2개의 인공 신경망 모델을 사용하여 기사 데이터를 기반으로 학습시킨 후에 기사 내에 포함된 모든 단어를 벡터화 하고, 대상 감염병과 Euclidean distance가 가까운 몇몇 단어를 추출하여 이를 지역별로 나타낼 수 있다. 구체적으로, 연관성이 높은 단어의 출현 빈도를 시계열 데이터로 생성하고 교차 상관 관계를 수치화 한 결과를 확인할 수 있다. 상기 결과를 이용하면, 감염병 키워드의 발생 빈도와 실제 감염병의 발생 빈도에 대한 교차를 통해 특정 감염병과 연관이 있는 특정 단어를 포함하는 기사의 발생 빈도가 늘어난 뒤 얼마만큼의 시간이 흐른 후 감염병이 발생하게 되는지 예측할 수 있게 한다.
- [0096] 예를 들어, 상기 인공 신경망 모델은 Contiguous Bag of Words과 Skip-Gram model with Negative Sampling 일

수 있다.

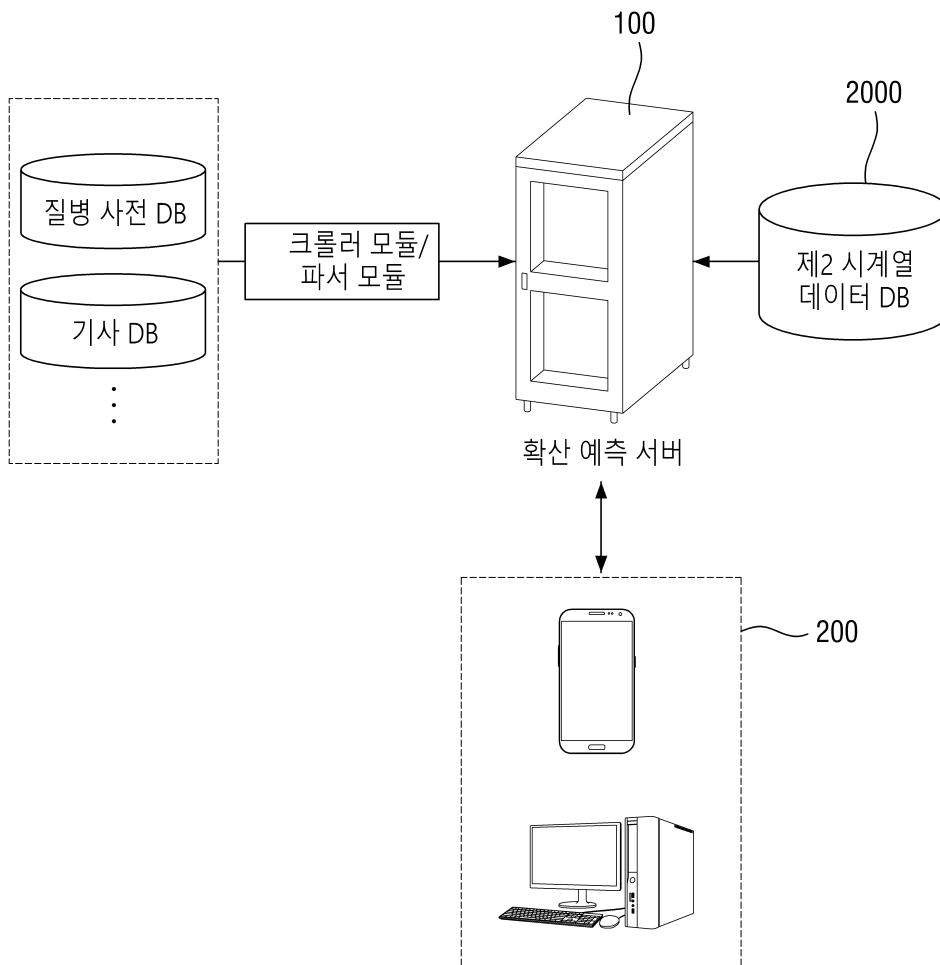
[0097] 지금까지 도 1 내지 도 17을 참조하여 설명된 본 발명의 개념은 컴퓨터가 읽을 수 있는 매체 상에 컴퓨터가 읽을 수 있는 코드로 구현될 수 있다. 상기 컴퓨터로 읽을 수 있는 기록 매체는, 예를 들어 이동형 기록 매체(CD, DVD, 블루레이 디스크, USB 저장 장치, 이동식 하드 디스크)이거나, 고정식 기록 매체(ROM, RAM, 컴퓨터 구비형 하드 디스크)일 수 있다. 상기 컴퓨터로 읽을 수 있는 기록 매체에 기록된 상기 컴퓨터 프로그램은 인터넷 등의 네트워크를 통하여 다른 컴퓨팅 장치에 전송되어 상기 다른 컴퓨팅 장치에 설치될 수 있고, 이로써 상기 다른 컴퓨팅 장치에서 사용될 수 있다.

[0098] 도면에서 동작들이 특정한 순서로 도시되어 있지만, 반드시 동작들이 도시된 특정한 순서로 또는 순차적 순서로 실행되어야만 하거나 또는 모든 도시된 동작들이 실행되어야만 원하는 결과를 얻을 수 있는 것으로 이해되어서는 안 된다. 특정 상황에서는, 멀티태스킹 및 병렬 처리가 유리할 수도 있다. 더욱이, 위에 설명한 실시예들에서 다양한 구성들의 분리는 그러한 분리가 반드시 필요한 것으로 이해되어서는 안 되고, 설명된 프로그램 컴포넌트들 및 시스템들은 일반적으로 단일 소프트웨어 제품으로 함께 통합되거나 다수의 소프트웨어 제품으로 패키징될 수 있음을 이해하여야 한다.

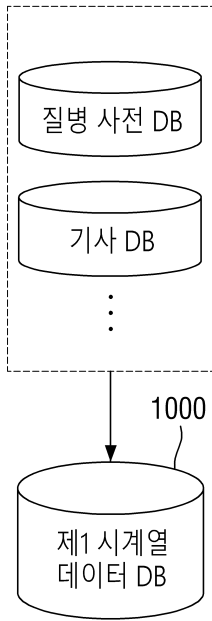
[0099] 이상 첨부된 도면을 참조하여 본 발명의 실시예들을 설명하였지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로 이해해야만 한다.

도면

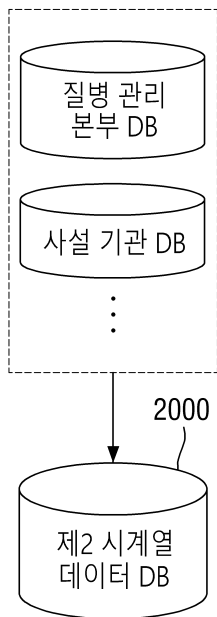
도면1



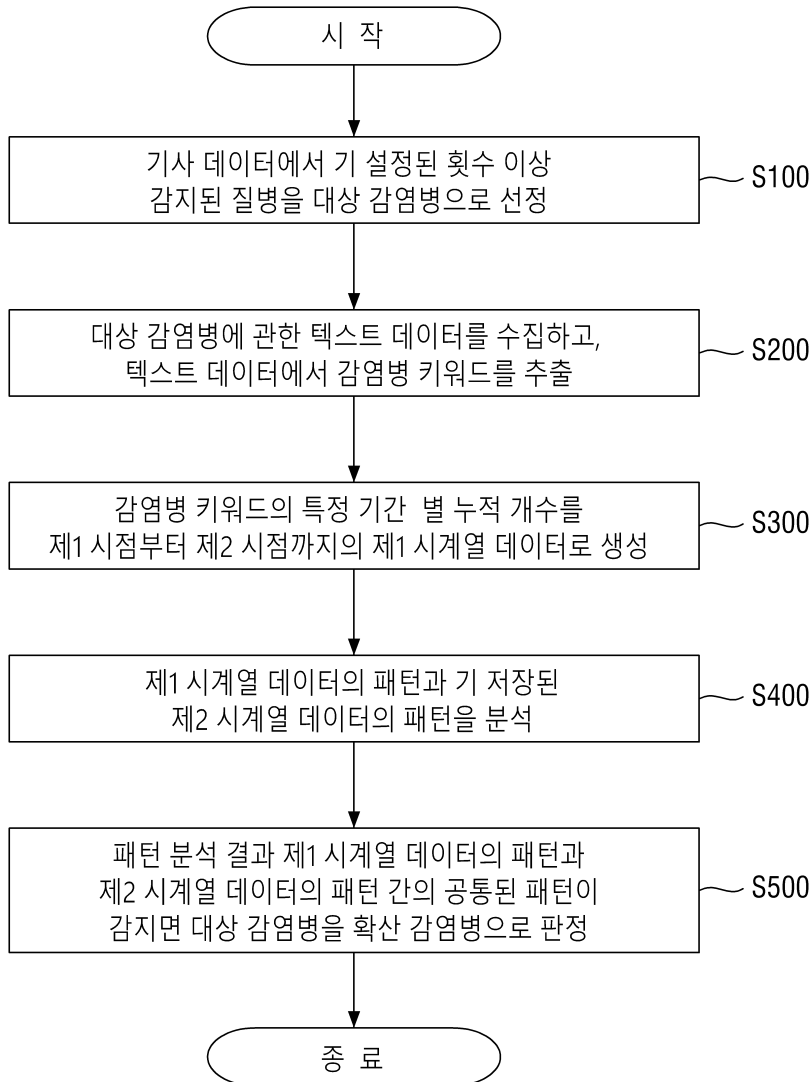
도면2



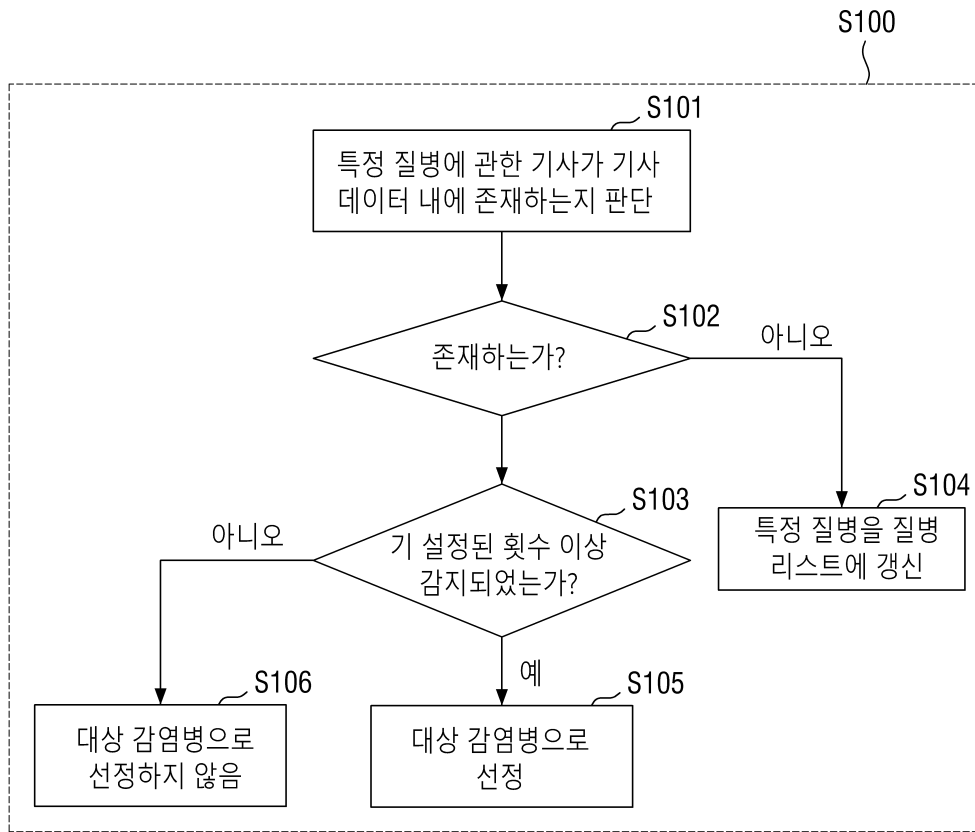
도면3



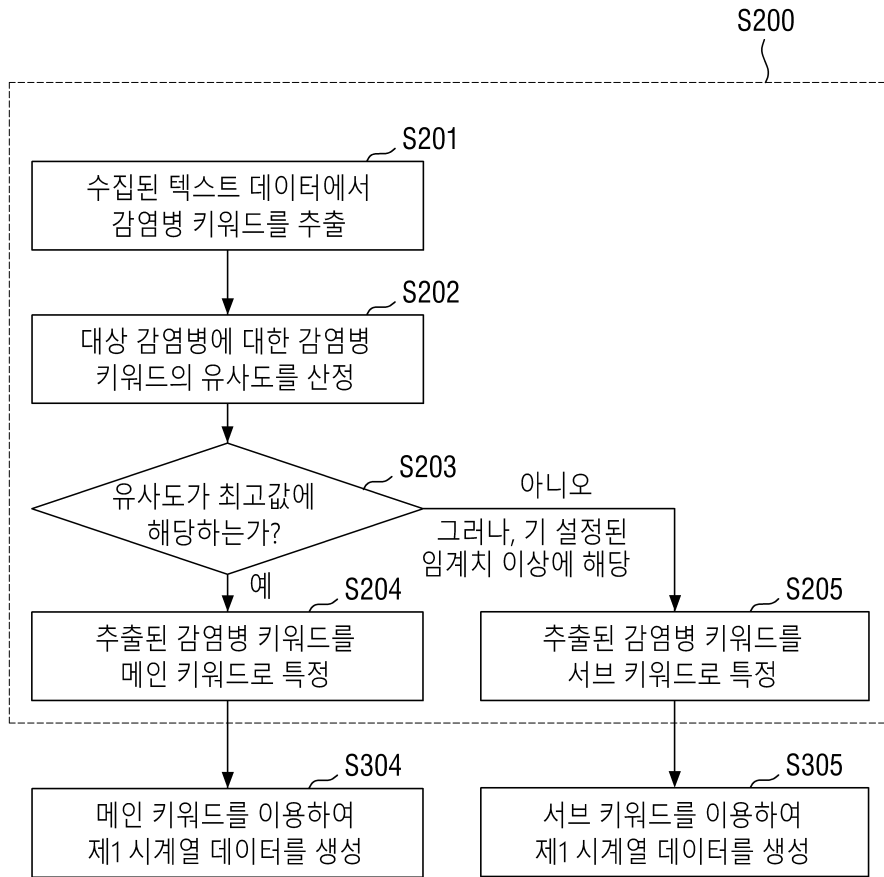
도면4



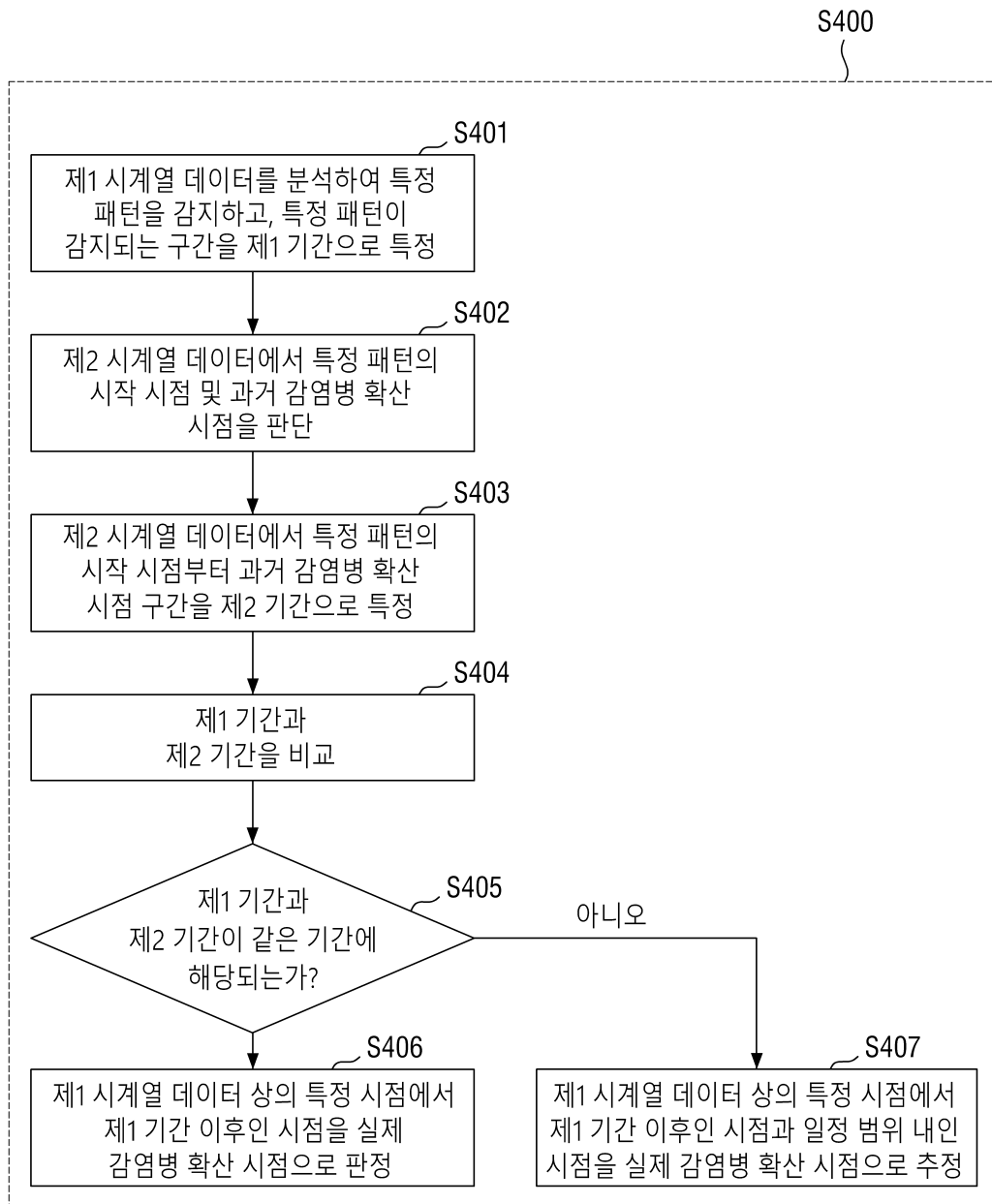
도면5



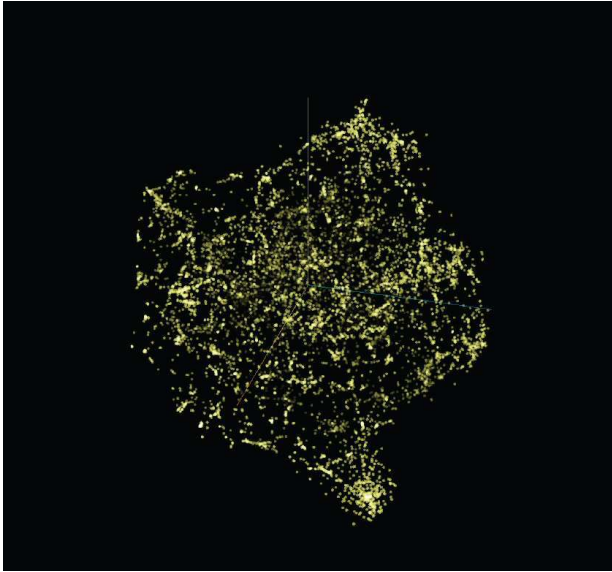
도면6



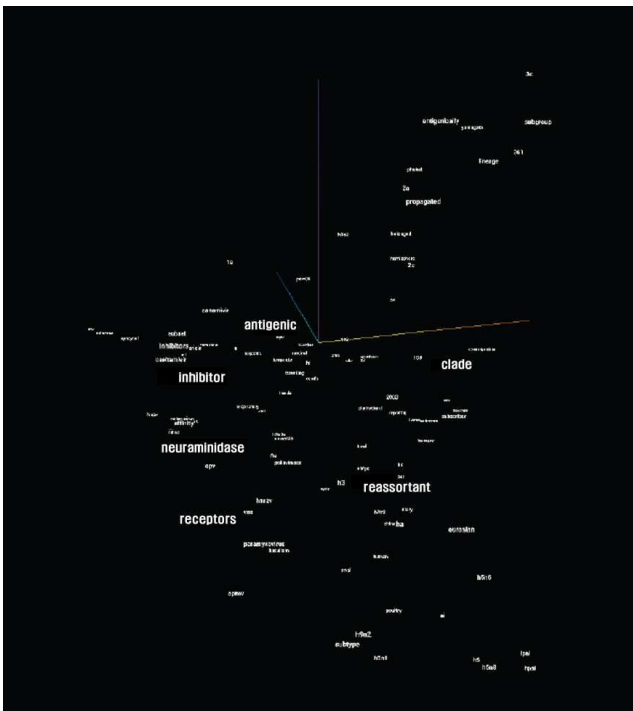
도면7



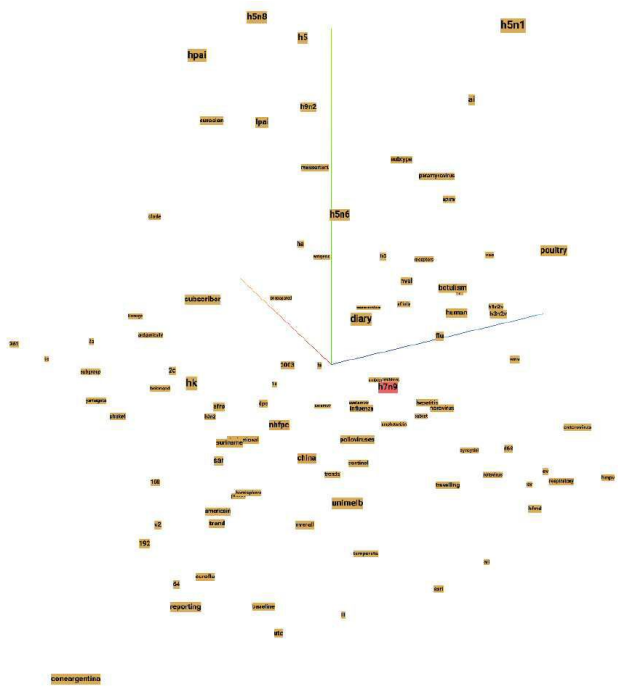
도면8



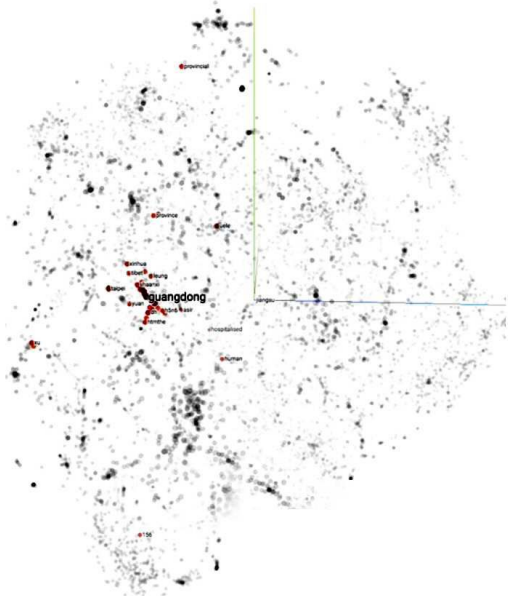
도면9



도면10



도면11



도면12

Search by

h7n9 .* label ▼

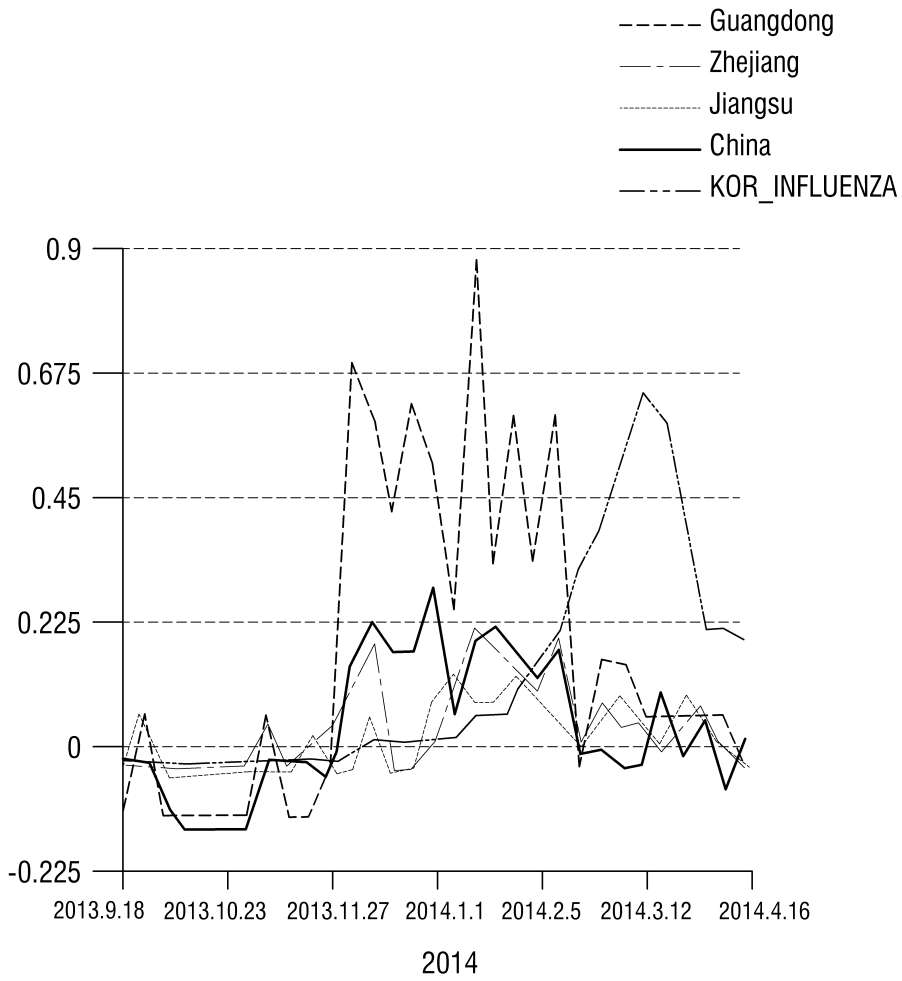
neighbors ? 100

distance COSINE EUCLIDIAN

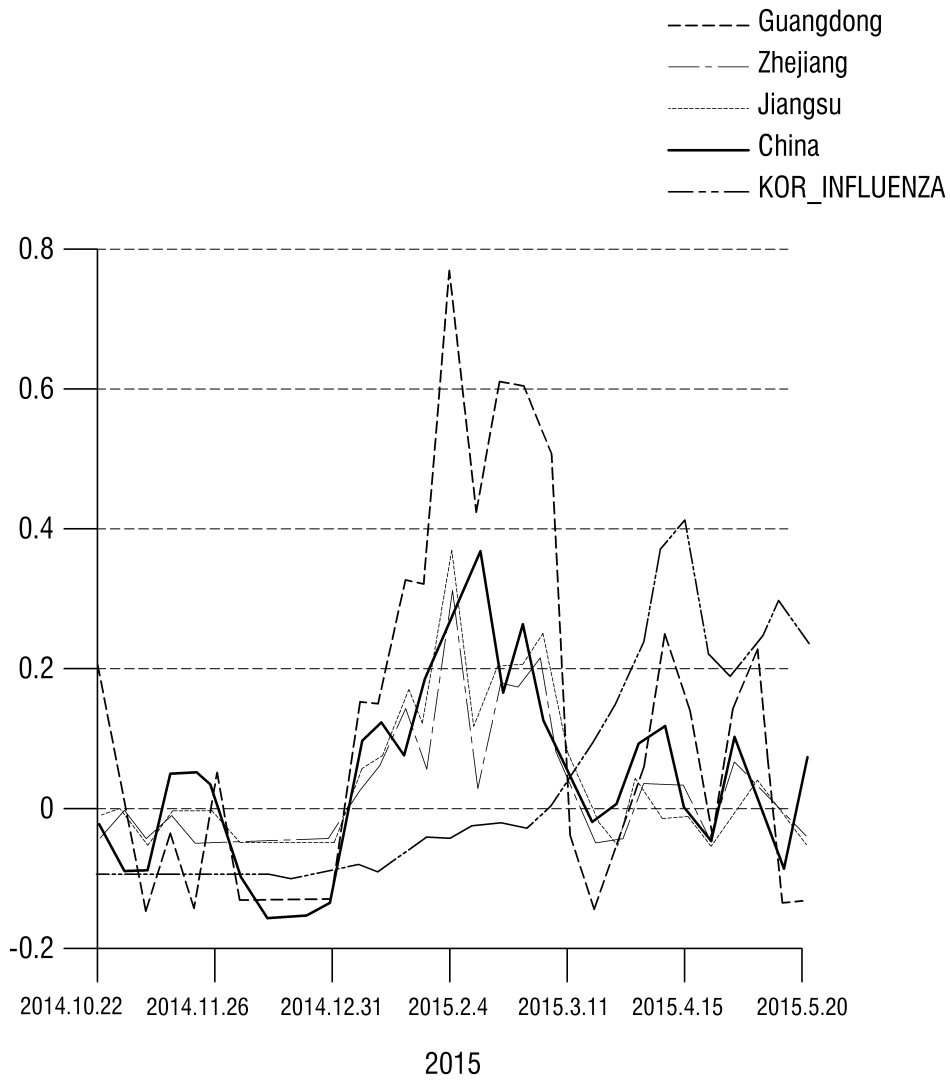
Nearest points in the original space:

guangdong	0.934
h5n6	0.961
china	0.962
shanghai	0.969
jiangsu	0.973
nhfpc	0.978
chinese	0.979
Zhejiang	0.983
planning	0.988
mainland	0.992
h9n2	0.993
ah7n9	0.995
beijing	1.011
dongguan	1.027
anhui	1.034
hongkong	1.036
commission	1.037
Jiangxi	1.046
hunan	1.066
flu	1.073
shenzhen	1.085
human	1.085
h10n8	1.088
h5n1	1.091

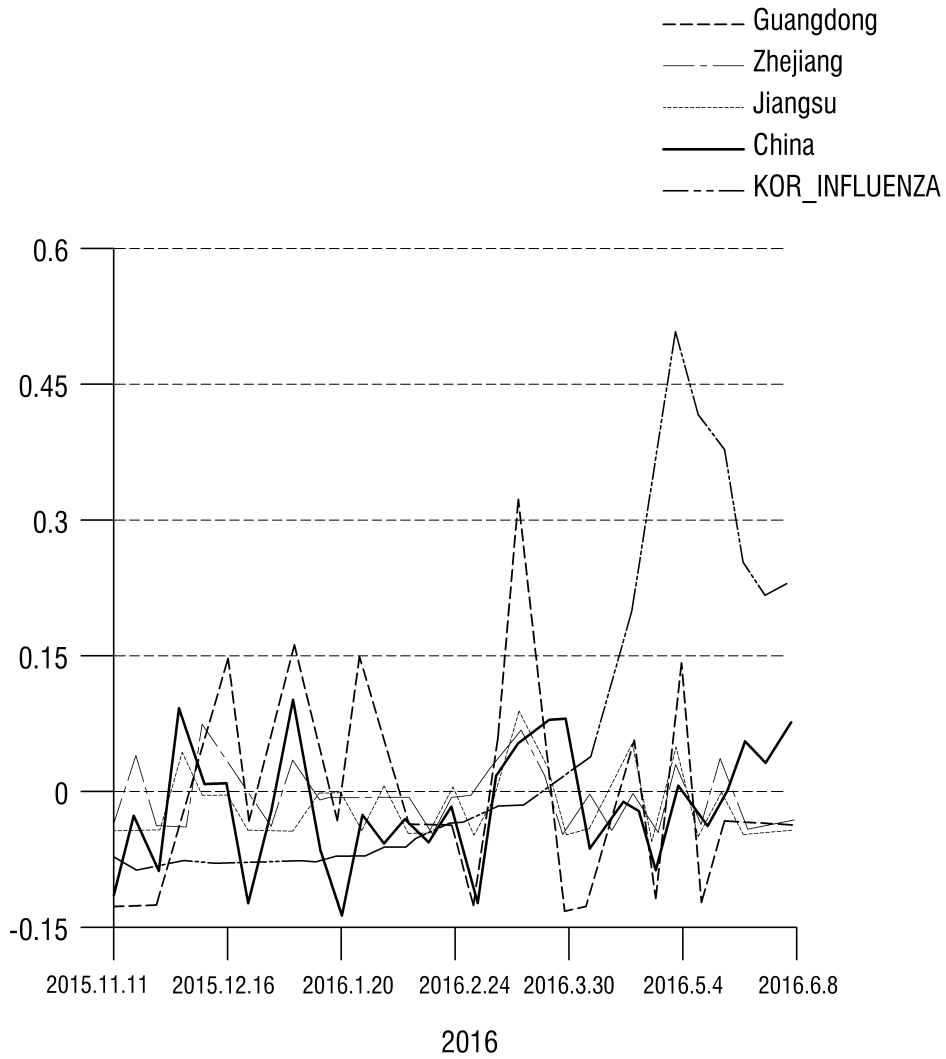
도면13



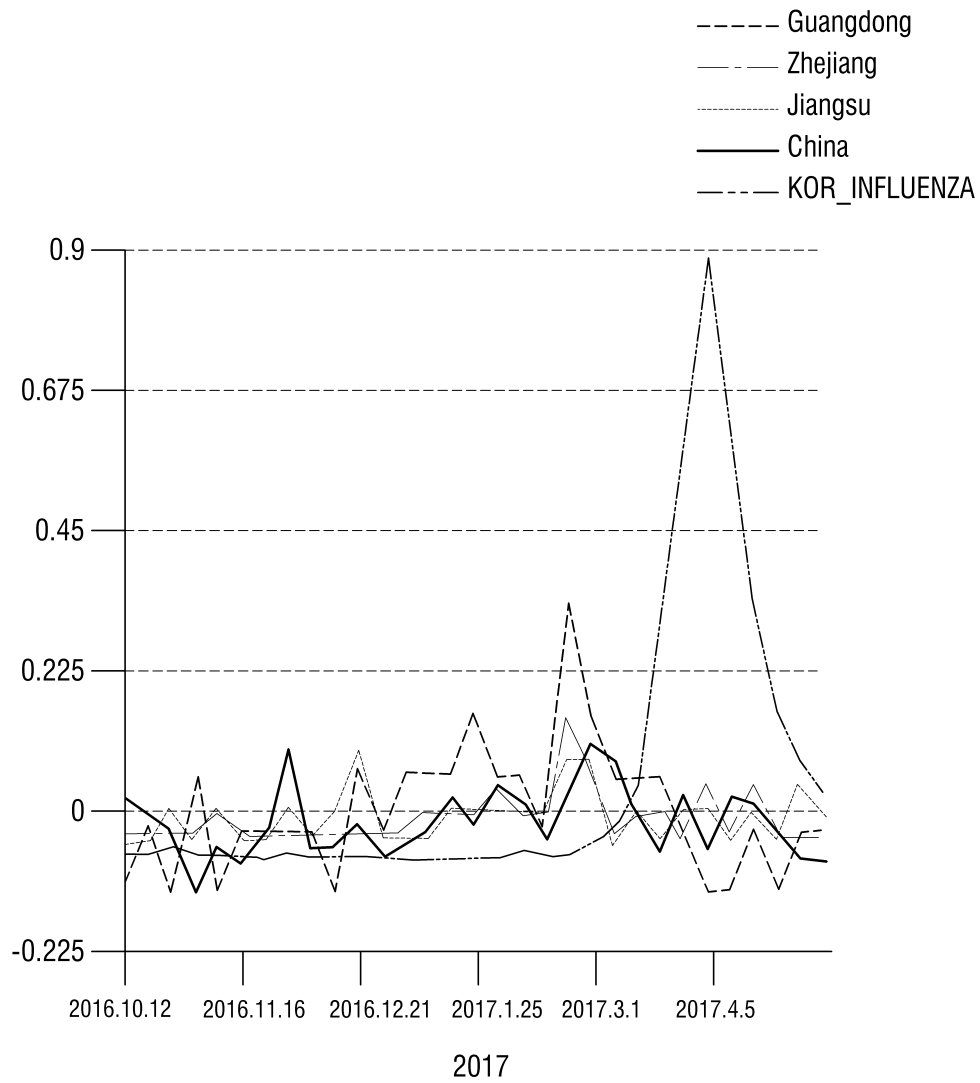
도면14



도면15



도면16



도면17

