



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0117691
(43) 공개일자 2020년10월14일

(51) 국제특허분류(Int. Cl.)

G06F 16/29 (2019.01) G06F 16/2458 (2019.01)

G06F 16/248 (2019.01)

(52) CPC특허분류

G06F 16/29 (2019.01)

G06F 16/2458 (2019.01)

(21) 출원번호 10-2019-0040174

(22) 출원일자 2019년04월05일

심사청구일자 2019년04월05일

(71) 출원인

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

이경호

서울특별시 중구 다산로 32, 5동 1002호 (신당동, 남산타운)

김용남

서울특별시 서대문구 거북골로20길 53-13, 401호 (북가좌동)

(74) 대리인

특허법인우인

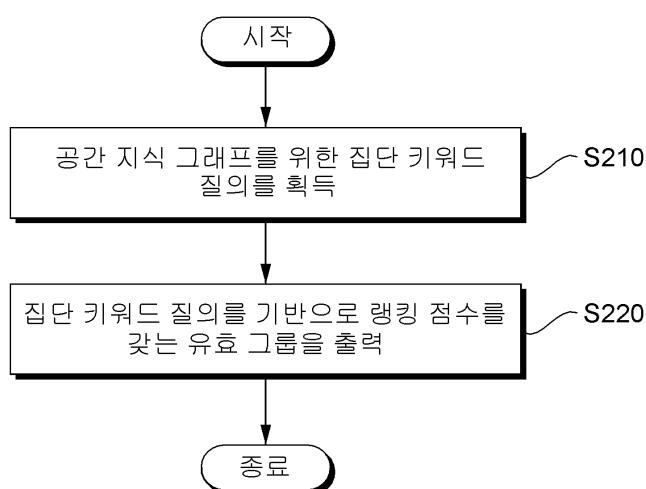
전체 청구항 수 : 총 15 항

(54) 발명의 명칭 공간 지식 베이스 기반의 집단 키워드 질의 처리 방법 및 장치

(57) 요약

본 실시예들은 사용자가 요구하는 위치, 복수의 공간 키워드, 및 출력 결과의 개수를 제한하는 집단 키워드 질의를 통하여 유효 그룹의 키워드 관련도 및 공간적 근접도를 측정하도록 랭킹 함수를 설계함으로써, 집단적으로 질의 키워드를 커버하여 사용자의 요구사항을 높은 확률로 만족시키는 공간 영역을 검색할 수 있는 집단 키워드 질의 처리 방법 및 장치를 제공한다.

대표도 - 도5



(52) CPC특허분류
G06F 16/248 (2019.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	NRF-2016R1A2B4015873
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	한국연구재단 중견연구자지원사업
연구과제명	사물인터넷을 위한 복합 이벤트 처리 기반 개인맞춤 상황인지형 추천 시스템
기 여 율	1/1
과제수행기관명	연세대학교
연구기간	2016.06.01 ~ 2019.05.31

명세서

청구범위

청구항 1

컴퓨팅 디바이스에 의한 집단 키워드 질의 처리 방법에 있어서,
공간 지식 그래프를 위한 집단 키워드 질의를 획득하는 단계; 및
상기 집단 키워드 질의를 기반으로 랭킹 점수를 갖는 유효 그룹을 출력하는 단계
를 포함하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 2

제1항에 있어서,
상기 집단 키워드 질의는 (i) 질의 장소, (ii) 복수의 의미적 공간 키워드, 및 (iii) 상기 유효 그룹의 선별 개수를 포함하며,
상기 유효 그룹은 의미적 장소 정보의 집합이며,
상기 의미적 장소 정보는 상기 공간 지식 그래프의 서브 트리를 의미하고, 상기 공간 지식 그래프는 상기 의미적 공간 키워드를 포함하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 3

제2항에 있어서,
상기 유효 그룹을 출력하는 단계는,
키워드 관련도 및 공간적 근접도를 기준으로 상기 유효 그룹을 평가하는 랭킹 함수를 이용하여 상기 복수의 의미적 공간 키워드 중에서 적어도 일부를 만족하는 의미적 장소 정보의 위치를 검색하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 4

제3항에 있어서,
상기 키워드 관련도는 상기 유효 그룹에 속하는 의미적 장소 정보마다 의미적 공간 키워드에 대한 최단 경로 길이의 합으로 산출되는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 5

제4항에 있어서,
상기 랭킹 함수는 상기 키워드 관련도 및 상기 공간적 근접도의 곱으로 정의되며, 상기 키워드 관련도가 0이 되지 않도록 상기 키워드 관련도에 1을 가산한 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 6

제3항에 있어서,
상기 공간적 근접도는,
(i) 상기 유효 그룹 내의 의미적 장소 정보 및 질의 장소 간의 질의 거리, (ii) 상기 유효 그룹 내의 의미적 장소 정보의 개수, 및 (iii) 상기 유효 그룹 내의 의미적 장소 정보 간의 쌍방 거리를 기반으로 산출되는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 7

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는

상기 공간적 근접도는 제1 거리와 제2 거리의 합으로 정의되며,

상기 제1 거리는 상기 유효 그룹에 속하는 의미적 장소 정보 및 질의 장소 간의 최대 거리이고, 상기 제2 거리는 상기 유효 그룹에 속하는 의미적 장소 정보 간의 최대 쌍방 거리에 의미적 장소 정보의 개수에서 1을 감산한 값을 곱한 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 8

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 공간 지식 그래프를 탐색하는 과정에서 좌표 유무를 기준으로 후보 유효 그룹을 필터링하고 상기 의미적 공간 키워드를 기준으로 후보 유효 그룹을 필터링하여 서브 트리를 추출하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 9

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 질의 거리와 상기 쌍방 거리를 고정된 상태에서 상기 유효 그룹의 선별 개수에 따라 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 상기 질의 거리의 한계를 설정하여 후보 유효 그룹을 필터링하고 상기 랭킹 점수를 갱신하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 10

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 질의 거리와 상기 쌍방 거리를 고정된 상태에서 상기 유효 그룹의 선별 개수에 따라 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 상기 쌍방 거리의 한계를 설정하여 후보 유효 그룹을 필터링하고 상기 랭킹 점수를 갱신하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 11

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 공간 지식 그래프를 탐색하는 과정에서 트리 깊이의 한계를 설정하여 후보 유효 그룹을 필터링하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 12

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 상기 질의 거리의 한계를 동적으로 설정하여 후보 유효 그룹을 필터링하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 13

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 상기 쌍방 거리의 한계를 동적으로 설정하여 후보 유효 그룹을 필터링하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 14

제6항에 있어서,

상기 유효 그룹을 출력하는 단계는,

상기 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹의 랭킹 점수를 참조하여 상기 의미적 장소 정보의 개수 제한을 동적으로 설정하여 후보 유효 그룹을 필터링하는 것을 특징으로 하는 집단 키워드 질의 처리 방법.

청구항 15

하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 집단 키워드 질의 처리 장치에 있어서,

상기 프로세서는 공간 지식 그래프를 위한 집단 키워드 질의를 획득하고,

상기 프로세서는 상기 집단 키워드 질의를 기반으로 랭킹 점수를 갖는 유효 그룹을 출력하는 것을 특징으로 하는 집단 키워드 질의 처리 장치.

발명의 설명

기술 분야

[0001] 본 실시예가 속하는 기술 분야는 공간 지식 베이스 기반의 집단 키워드 질의 처리 방법 및 장치에 관한 것이다.

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 지식 공유 커뮤니티의 증가 및 자동 정보 추출 기술의 발전에 따라 대규모 지식 베이스가 구축되고 있다. 예컨대 DBpedia, YAGO, Freebase 등의 지식 베이스는 방대한 양의 고품질의 의미적 데이터(Semantic Data)를 유지 관리하고 의미적 데이터는 WWW 표준인 RDF 데이터 모델을 준수한다.

[0004] 지식 베이스는 공간 데이터로 보강되고 있다. DBpedia의 11%은 공간 데이터를 포함하고 YAGO2의 59%는 공간 데이터를 포함한다. 공간 데이터의 증가는 지식 그래프를 대상으로 위치 기반 질의를 수행 가능하게 한다.

[0005] GeoSPARQL은 의미적 웹(Semantic Web) 상의 공간 데이터를 위한 표준 질의 언어이다. GeoSPARQL과 같은 구조적인 언어는 사용자가 질의 언어의 문법 및 대상 데이터의 스키마를 숙지해야 하는 문제가 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개공보 제10-2012-0104650호 (2012.09.24)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예들은 사용자가 요구하는 위치, 공간 키워드, 및 출력 결과의 개수를 제한하는 집단 키워드 질의를 통하여 유효 그룹의 키워드 관련도 및 공간적 근접도를 측정하도록 랭킹 함수를 설계함으로써, 집단적으로 질의 키워드를 커버하여 사용자의 요구사항을 높은 확률로 만족시키는 공간 영역을 검색하는 데 주된 목적이 있다.

[0008] 본 발명의 명시되지 않은 또 다른 목적들은 하기의 상세한 설명 및 그 효과로부터 용이하게 추론할 수 있는 범위 내에서 추가적으로 고려될 수 있다.

과제의 해결 수단

[0009] 본 실시예의 일 측면에 의하면, 컴퓨팅 디바이스에 의한 집단 키워드 질의 처리 방법에 있어서, 공간 지식 그래프를 위한 집단 키워드 질의를 획득하는 단계, 및 상기 집단 키워드 질의를 기반으로 랭킹 점수를 갖는 유효 그룹을 출력하는 단계를 포함하는 것을 특징으로 하는 집단 키워드 질의 처리 방법을 제공한다.

[0010] 본 실시예의 다른 측면에 의하면, 하나 이상의 프로세서 및 상기 하나 이상의 프로세서에 의해 실행되는 하나 이상의 프로그램을 저장하는 메모리를 포함하는 집단 키워드 질의 처리 장치에 있어서, 상기 프로세서는 공간 지식 그래프를 위한 집단 키워드 질의를 획득하고, 상기 프로세서는 상기 집단 키워드 질의를 기반으로 랭킹 점수를 갖는 유효 그룹을 출력하는 것을 특징으로 하는 집단 키워드 질의 처리 장치를 제공한다.

발명의 효과

[0011] 이상에서 설명한 바와 같이 본 발명의 실시예들에 의하면, 사용자가 요구하는 위치, 복수의 공간 키워드, 및 출력 결과의 개수를 제한하는 집단 키워드 질의를 통하여 유효 그룹의 키워드 관련도 및 공간적 근접도를 측정하도록 랭킹 함수를 설계함으로써, 집단적으로 질의 키워드를 커버하여 사용자의 요구사항을 높은 확률로 만족시키는 공간 영역을 검색할 수 있는 효과가 있다.

[0012] 여기에서 명시적으로 언급되지 않은 효과라 하더라도, 본 발명의 기술적 특징에 의해 기대되는 이하의 명세서에서 기재된 효과 및 그 잠정적인 효과는 본 발명의 명세서에 기재된 것과 같이 취급된다.

도면의 간단한 설명

[0013] 도 1은 본 발명의 일 실시예에 따른 집단 키워드 질의 처리 장치를 예시한 블록도이다.

도 2는 의미적 장소를 예시한 지도이다.

도 3은 유효 그룹 1 및 유효 그룹 2를 예시한 지도이다.

도 4는 RDF 데이터를 예시한 도면이다.

도 5는 본 발명의 다른 실시예에 따른 집단 키워드 질의 처리 방법을 예시한 흐름도이다.

도 6은 본 발명의 다른 실시예에 따른 집단 키워드 질의 처리 방법에서 랭킹 점수를 갖는 유효 그룹을 출력하는 단계를 예시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하, 본 발명을 설명함에 있어서 관련된 공지기능에 대하여 이 분야의 기술자에게 자명한 사항으로서 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하고, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다.

[0015] 도 1은 본 발명의 일 실시예에 따른 집단 키워드 질의 처리 장치를 예시한 블록도이다.

[0016] 집단 키워드 질의 처리 장치(110)는 적어도 하나의 프로세서(120), 컴퓨터 판독 가능한 저장매체(130) 및 통신 버스(170)를 포함한다.

[0017] 프로세서(120)는 집단 키워드 질의 처리 장치(110)로 동작하도록 제어할 수 있다. 예컨대, 프로세서(120)는 컴퓨터 판독 가능한 저장 매체(130)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 컴퓨터 실행 가능 명령어는 프로세서(120)에 의해 실행되는 경우 집단 키워드 질의 처리 장치(110)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.

[0018] 컴퓨터 판독 가능한 저장 매체(130)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능한 저장 매체(130)에 저장된 프로그램(140)은 프로세서(120)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독한 가능 저장 매체(130)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하

나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 집단 키워드 질의 처리 장치(110)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

- [0019] 통신 버스(170)는 프로세서(120), 컴퓨터 판독 가능한 저장 매체(140)를 포함하여 집단 키워드 질의 처리 장치(110)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0020] 집단 키워드 질의 처리 장치(110)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(150) 및 하나 이상의 통신 인터페이스(160)를 포함할 수 있다. 입출력 인터페이스(150) 및 통신 인터페이스(160)는 통신 버스(170)에 연결된다. 입출력 장치(미도시)는 입출력 인터페이스(150)를 통해 집단 키워드 질의 처리 장치(110)의 다른 컴포넌트들에 연결될 수 있다.
- [0021] 집단 키워드 질의 처리 장치(110)는 공간 지식 베이스 또는 공간 지식 그래프를 대상으로 키워드 기반의 공간 질의인 kSP(top-k Semantic Place Retrieval)를 활용한다. kSP 질의는 공간 정점을 루트로 하는 서브 트리인 의미적 장소(Semantic Place, SP)를 찾는다.
- [0022] 의미적 장소는 사용자 질의의 모든 키워드를 포함하면서 랭킹 함수의 값을 최소화하여야 한다. 랭킹 함수는 그룹 내 하나의 공간 오브젝트에서 출발하여 기타 공간 오브젝트를 모두 방문하는데 소요하는 최단 거리, 즉 허브 기반 거리를 측정한다.
- [0023] 하지만 실제 공간에서 하나의 의미적 장소가 복수의 키워드를 모두 포함할 확률은 희박하고, 의미적 장소가 사용자가 요구한 위치 근처에 존재할 확률은 더욱 희박하다. 만약 출력한 그룹에 속하는 장소가 서로 인접하더라도 사용자가 요구하는 위치와는 멀리 떨어져 있을 수 있다.
- [0024] 비 의미적 데이터를 대상으로 검색하는 방식은 사용자가 '음식점'이라고 키워드를 입력했을 때 '한식집'이라는 키워드를 포함하고 있는 장소를 후보로 선정할 수 없어서 결과의 정확도가 높지 않다.
- [0025] 공간 정보가 사용자의 모든 키워드를 포함하지 않는 문제를 해결하기 위해서 집단 키워드 질의 처리 장치(110)는 사용자가 요구하는 위치, 복수의 공간 키워드, 및 출력 결과의 개수를 제한하는 집단 키워드 질의를 통하여 유효 그룹의 키워드 관련도 및 공간적 근접도를 측정하도록 랭킹 함수를 설계함으로써, 집단적으로 질의 키워드를 커버하여 사용자의 요구사항을 높은 확률로 만족시키는 공간 영역을 검색한다.
- [0026] 도 2 내지 도 4를 참조하여 장소 검색 시나리오를 설명하기로 한다.
- [0027] 도 2는 의미적 장소를 예시한 지도이고, 도 3은 유효 그룹 1 및 유효 그룹 2를 예시한 지도이고, 도 4는 RDF 데이터를 예시한 도면이다.
- [0028] 뉴욕 시티와 같은 특정 공간 영역에서 식당(Restaurant), 호텔(Hotel), 공원(Park), 및 박물관(Museum)을 찾을 때, 각각의 장소는 서로 인접하고 지하철역에서 멀지 않아야 한다.
- [0029] 도 2에서는 RDF(Resource Description Framework) 그래프에 기반한 키워드 {Restaurant, Hotel, Park, Museum}를 적어도 하나 포함하는 장소(p)가 41개 도시되어 있다.
- [0030] 도 3에 도시된 유효 그룹 1(vg1)은 p1(Hotel), p2(Restaurant), p3(Park), p4(Museum)를 포함하고, 유효 그룹 2(vg2)는 p5(Hotel), p6(Park), p7(Restaurant, Museum)을 포함한다.
- [0031] 적합한 장소를 선택하기 위해서 집단 키워드 질의 처리 장치(110)는 키워드 관련도 및 공간적 근접도를 기준으로 유효 그룹을 평가한다. 집단 키워드 질의 처리 장치(110)는 지식 베이스 내의 정점 사이의 관계를 분석하고 허브 기반의 최대합(MaxSum) 거리를 산출하고, 질의 위치를 인식하여 공간적 근접도를 산출한다.

표 1

	vg_1	vg_2
semantic places	$\{p_1, p_2, p_3, p_4\}$	$\{p_5, p_6, p_7\}$
query distance	0.011599	0.011598
cardinality	4	3
pairwise distance	0.008246	0.010891
hub-based MaxSum	0.036337	0.033380

[0032]

[0033] 표 1을 참조하면, 유효 그룹 2는 유효 그룹 1보다 작은 허브 기반의 최대합 거리 값을 갖는다. 유효 그룹 1은 유효 그룹 2보다 작은 쌍방 거리 값을 갖는다. 유효 그룹 1에서 장소의 개수(Cardinality)는 유효 그룹 2보다 크다.

[0034] 집단 키워드 질의 처리 장치(110)는 최단 경로의 길이(Length of Shortest Path, LSP)를 누적하여 키워드 관련도를 산출한다. 도 4를 참조하면, 유효 그룹 1의 키워드 관련도는 $(1+1+1+1)+1=5$ 이고, 유효 그룹 2의 키워드 관련도는 $(1+1+2)+1=5$ 이다. p7의 최단 경로의 길이는 2이고 나머지는 1이다.

[0035] 집단 키워드 질의 처리 장치(110)는 의미적 장소, 질의 거리, 장소의 개수, 쌍방 거리를 종합적으로 평가하여 키워드 관련도 및 공간적 근접도의 관계를 기반으로 유효 그룹 2를 출력한다. 사용자 입장에서 이동 거리가 가까운 유효 그룹 2가 유효 그룹 1보다 정답에 가깝다.

[0036] 도 5는 본 발명의 다른 실시예에 따른 집단 키워드 질의 처리 방법을 예시한 흐름도이다. 집단 키워드 질의 처리 방법은 컴퓨팅 디바이스에 의하여 수행될 수 있으며, 집단 키워드 질의 처리 장치와 동일한 방식으로 동작한다.

[0037] 집단 키워드 질의 처리 방법은 하나의 공간 지식 베이스 및 질의를 입력으로 받고, Top k 랭킹 점수를 갖는 유효 그룹을 출력한다.

[0038] 단계 S210에서 프로세서는 공간 지식 그래프를 위한 집단 키워드 질의를 획득한다. 집단 키워드 질의(q)는 (i) 질의 장소(q, λ), (ii) 복수의 의미적 공간 키워드(q, φ), 및 (iii) 유효 그룹의 선별 개수(k)를 포함한다.

[0039] 단계 S210에서 프로세서는 집단 키워드 질의를 기반으로 랭킹 점수를 갖는 유효 그룹을 출력한다. 유효 그룹은 의미적 장소 정보의 집합으로 각각의 의미적 장소 정보가 모여 전체 키워드의 조건을 최대한 만족시킨다. 의미적 장소 정보는 공간 지식 그래프의 서브 트리(\mathcal{T})를 의미하고, 공간 지식 그래프는 의미적 공간 키워드를 포함한다.

[0040] 유효 그룹을 출력하는 단계(S210)는, 키워드 관련도 및 공간적 근접도를 기준으로 유효 그룹을 평가하는 랭킹 함수를 이용하여 복수의 의미적 공간 키워드 중에서 적어도 일부를 만족하는 의미적 장소 정보의 위치를 검색한다. 낮은 랭킹 점수를 갖는 유효 그룹은 사용자의 요구사항에 매칭한 확률이 높다.

[0041] 랭킹 함수 f는 수학식 1과 같이 표현된다.

수학식 1

[0042]

$$f(\cdot) = kr(vg, q) \times sd(vg, q)$$

[0043] 수학식 1에서 랭킹 함수는 키워드 관련도 및 공간적 근접도의 곱으로 정의된다. kr은 키워드 관련도(Keyword Relevance)이고, sd는 공간적 근접도(Spatial Distance)이다. vg은 유효 그룹(Valid Group)이고, q는 질의(query)이다.

[0044] 키워드 관련도는 수학적식 2와 같이 표현된다.

수학적식 2

$$kr(vg, q) = 1 + \sum_{\tau_i \in q} \sum_{rsp_j \in vg} lsp(rsp_j, \tau_i)$$

[0045]

[0046] 키워드 관련도는 유효 그룹에 속하는 의미적 장소 정보마다 의미적 공간 키워드에 대한 최단 경로의 길이(LSP)의 합으로 산출된다. 최단 경로의 길이는 RDF 그래프에서 관련된 의미적 장소에 매칭하는 깊이를 의미한다. 키워드 관련도가 0이 되지 않도록 키워드 관련도에 1을 가산한다.

[0047] 공간적 근접도는 (i) 유효 그룹 내의 의미적 장소 정보 및 질의 장소 간의 질의 거리, (ii) 유효 그룹 내의 의미적 장소 정보의 개수, 및 (iii) 유효 그룹 내의 의미적 장소 정보 간의 쌍방 거리를 기반으로 산출된다. 공간적 근접도는 수학적식 3과 같이 표현된다.

수학적식 3

$$sd(vg, q) = \max_{rsp_i \in vg} dist(rsp_i, q) + (|vg| - 1) \times diameter(vg)$$

[0048]

[0049] 공간적 근접도는 제1 거리와 제2 거리의 합으로 정의된다. 제1 거리는 유효 그룹에 속하는 의미적 장소 정보 및 질의 장소 간의 최대 거리이고, 제2 거리는 유효 그룹에 속하는 의미적 장소 정보 간의 최대 쌍방 거리에 의미적 장소 정보의 개수에서 1을 감산한 값을 곱한 값이다. |vg|는 의미적 장소 정보의 개수(Cardinality)이다.

[0050] 최대 쌍방 거리는 수학적식 4와 같이 표현된다.

수학적식 4

$$diameter(vg) = \max_{rsp_i, rsp_j \in vg} dist(rsp_i, rsp_j)$$

[0051]

[0052] 최대 쌍방 거리는 유효 그룹 내의 복수의 장소 간의 거리 중에서 최대값을 의미한다.

[0053] 프로세서는 4개의 요소를 고려하여 랭킹 함수를 산출한다. 프로세서는 키워드 관련도, 질의 거리, 쌍방 거리, 및 의미적 장소 정보의 개수를 기준으로 순위를 판단한다.

[0054] 프로세서는 질의 거리와 쌍방 거리를 고정된 상태에서 유효 그룹의 선별 개수에 따라 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 질의 거리의 한계를 설정할 수 있다. 프로세서는 질의 거리의 한계에 따라 후보 유효 그룹을 필터링하고 랭킹 점수를 갱신하여 유효 그룹을 출력할 수 있다.

[0055] 프로세서는 질의 거리와 상기 쌍방 거리를 고정된 상태에서 상기 유효 그룹의 선별 개수에 따라 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 쌍방 거리의 한계를 설정할 수 있다. 프로세서는 쌍방 거리의 한계에 따라 후보 유효 그룹을 필터링하고 랭킹 점수를 갱신하여 유효 그룹을 출력할 수 있다.

[0056] 도 6은 집단 키워드 질의 처리 방법에서 랭킹 점수를 갖는 유효 그룹을 출력하는 단계를 예시한 흐름도이다.

[0057] 집단 키워드 질의 처리 방법은 정확도에 집중한 제1 검색 방식과 속도에 집중한 제2 검색 방식으로 동작할 수 있다.

[0058] 제1 검색 방식은 지식 베이스 내의 후보 서브 트리를 검색하여 정확한 Top k 결과를 출력한다. 제1 검색 방식은 모든 후보 서브 트리를 검색하지 않고, 랭킹 함수를 분할하여 질의 거리 한계와 쌍방 거리 한계를 결정하는 중

요 의미적 장소를 찾는다. 중요 의미적 장소에 기반하여 검색 공간을 줄이는 필터링을 수행한다. 중요 의미적 장소는 질의 키워드를 적어도 하나를 포함하고 있는 의미적 장소를 나타낸다.

- [0059] 단계 S310에서 프로세서는 공간 지식 그래프를 탐색하는 과정에서 좌표 유무를 기준으로 후보 유효 그룹을 필터링하고 의미적 공간 키워드를 기준으로 후보 유효 그룹을 필터링하여 서브 트리를 추출한다.
- [0060] 단계 S320에서 프로세서는 질의 거리와 쌍방 거리를 고정된 상태에서 유효 그룹의 선별 개수에 따라 마지막 순위에 해당하는 유효 그룹을 선택한다.
- [0061] 단계 S330에서 프로세서는 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 질의 거리의 한계를 설정한다.
- [0062] 단계 S340에서 프로세서는 선택된 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 쌍방 거리의 한계를 설정한다. 단계 S350에서 프로세서는 후보 유효 그룹을 필터링하고 랭킹 점수를 갱신한다.
- [0063] 제2 검색 방식은 추가적으로 트리 깊이 한계를 통해 필터링을 수행한다. 적당한 트리 깊이 한계는 정확도를 희생하면서 검색 공간을 획기적으로 줄일 수 있다. 예를 들어, 트리 깊이 한계를 8로 설정했을 때 DBpedia의 경우 검색 공간이 감소하여 질의 처리 시간은 50% 이상 감소하였으나 정확도는 4% 정도 감소하였다.
- [0064] 제2 검색 방식은 알고리즘의 조기 종료를 위한 질의 거리, 쌍방 거리, 및 장소의 개수를 동적으로 제한하여 질의 처리 시간을 단축한다.
- [0065] 표 2는 제2 검색 방식의 알고리즘을 예시한다.

표 2

Algorithm 1 iSCK

Input: \mathcal{G} := knowledge graph, q := query, tdb := tree depth bound

Output: kVG := top- k valid groups

```

1:  $kVG = \emptyset, \delta = +\infty$ 
2: for each  $\mathcal{T}_i = (\mathcal{V}^\circ, \mathcal{E}^\circ) \in \mathcal{G}$  do
3:   if  $\mathcal{T}_i.root \neq pv$  or  $\forall v \in \mathcal{V}^\circ v.\varphi \cap q.\varphi = \emptyset$  then
4:     delete  $\mathcal{T}_i$  from  $\mathcal{G}$  ▷ Pruning Rule 1
5:  $RSP \leftarrow \mathcal{G}$ 
6: for each  $\tau_i \in q.\varphi, rsp_j \in RSP$  do
7:   if  $\tau_i \in rsp_j.\varphi$  and  $lsp(rsp_j, \tau_i) > tdb$  then
8:     delete  $\tau_i$  from  $rsp_j.\varphi$  ▷ Pruning Rule 4
9: delete  $rsp$  from  $RSP$  whose  $rsp.\varphi = \emptyset$ 
10:  $RSP_{qd}^\oplus \leftarrow RSP$ 
11: while  $RSP_{qd}^\oplus \neq \emptyset$  do
12:    $RSP^\oplus \leftarrow RSP$ 
13:    $rsp_{qd} \leftarrow \text{select NN of } q \text{ from } RSP_{qd}^\oplus$ 
14:   if  $dist(rsp_{qd}, q) > \delta / (lsp(rsp_{qd}, q) + 1)$  then
15:     delete  $rsp_{qd}$  and continue ▷ Lemma 2
16:   if  $rsp_{qd}.\varphi \supseteq q.\varphi$  then
17:     generate a  $vg \{rsp_{qd}\}$ , maintain  $kVG$ , and continue
18:   for each  $rsp_i \in RSP_{qd}^\oplus$  do
19:     if  $rsp_i \notin circle(q, dist(q, rsp_{qd}))$  then
20:       delete  $rsp_i$  from  $RSP^\oplus$  ▷ Pruning Rule 2
21:    $RSP_{pd}^\oplus \leftarrow RSP^\oplus$ 
22:   while  $\exists \text{unselected}(rsp_{pi}, rsp_{pj}) \in RSP_{pd}^\oplus$  do
23:      $rsp_{pi}, rsp_{pj} \leftarrow \text{choose an unselected } rsp \text{ pair}$ 
24:     if  $(rsp_{qd}.\varphi \cup rsp_{pi}.\varphi) \supseteq q.\varphi$  then
25:       generate a  $vg \{rsp_{qd}, rsp_{pi}\}$ , maintain  $kVG$ , and continue
26:     if  $(rsp_{pi}, rsp_{pj})$  conflicts dynamic constrains for pair-wise distance bound then
27:       set as selected and continue ▷ Lemma 3
28:     for each  $rsp_i \in RSP_{pd}^\oplus$  do
29:       if  $rsp_i \notin circle(rsp_{pi}, \cdot) \cap circle(rsp_{pj}, \cdot)$  then
30:         delete  $rsp_i$  from  $RSP_{pd}^\oplus$  ▷ Pruning Rule 3
31:    $RSP_{vg}^\oplus \leftarrow RSP_{pd}^\oplus$ 
32:    $gsc \leftarrow \text{apply GreedySetCover on } (RSP_{vg}^\oplus - rsp_{qd} -$ 
    $rsp_{pi} - rsp_{pj})$ 
33:    $\phi_{min} \leftarrow gsc \cdot (\log |q.\varphi| + 1) + 3$  ▷ Lemma 4
34:   calculate  $\phi_{max}$ 
35:   while  $\phi_{min} \leq \phi \leq \phi_{max}$  do
36:      $VG_\phi \leftarrow \text{generate uncovered } vg \text{ w.r.t. } \phi$ 
37:     calculate ranking scores of  $VG_\phi$ 
38:     if  $\theta < \delta$  then
39:        $kVG \leftarrow \text{add } vg_i, \delta \leftarrow \theta, \text{maintainTopK}(kVG)$ 
40:      $\theta \leftarrow \theta + 1$ 
41:   delete  $rsp_{qd}$  from  $RSP_{qd}^\oplus$ 
42: return  $kVG$ 

```

[0066]

[0067] 표 2에서 RSP는 의미적 장소이고, T는 서브 트리이고, kVG는 Top k의 해당하는 유효 그룹이고, ϕ 는 현재 의미적 장소 정보의 개수이고, θ 는 유효 그룹의 랭킹 점수이고, δ 는 현재 Top k의 랭킹 점수이다.

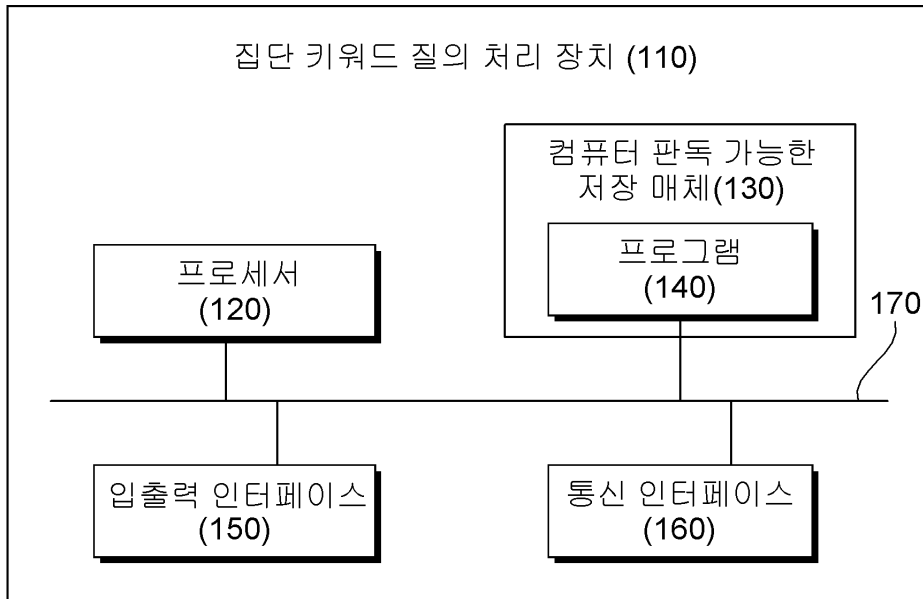
[0068] 제2 검색 방식의 알고리즘 6 내지 9 라인에서 프로세서는 공간 지식 그래프를 탐색하는 과정에서 트리 깊이의 한계를 설정하여 후보 유효 그룹을 필터링하여 유효 그룹을 출력한다.

[0069] 제2 검색 방식의 알고리즘 14 내지 15 라인에서 프로세서는 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 질의 거리의 한계를 동적으로 설정하여 후보 유효 그룹을 필터링하여 유효 그룹을 출력한다.

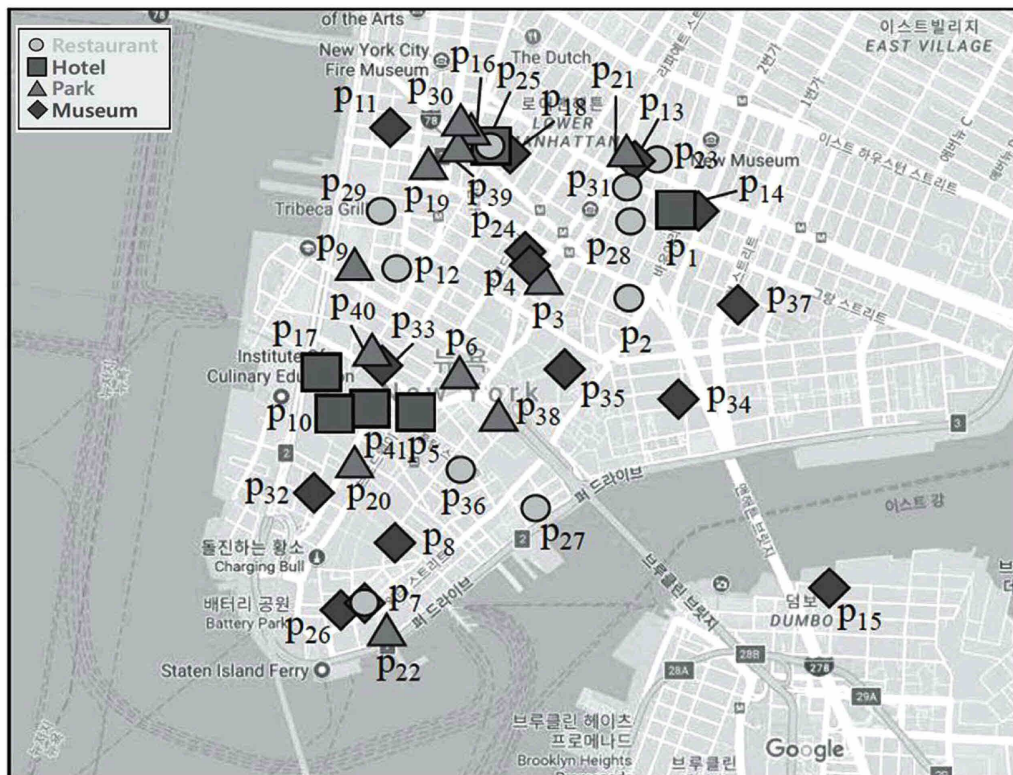
- [0070] 제2 검색 방식의 알고리즘 26 내지 27 라인에서 프로세서는 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹에 속하는 의미적 장소 정보를 기준으로 쌍방 거리의 한계를 동적으로 설정하여 후보 유효 그룹을 필터링하여 유효 그룹을 출력한다.
- [0071] 제2 검색 방식의 알고리즘 32 내지 40 라인에서 프로세서는 유효 그룹의 선별 개수의 범위를 만족하는 현재 마지막 순위에 해당하는 유효 그룹의 랭킹 점수를 참조하여 의미적 장소 정보의 개수 제한을 동적으로 설정하여 후보 유효 그룹을 필터링하여 유효 그룹을 출력한다.
- [0072] 본 실시예에서 정의된 질의는 사용자가 문법을 숙지해야 하는 구조적 언어가 아니라 사용자한테 친숙한 키워드 기반의 질의이다. 키워드 질의는 사용자의 공간 지식 베이스에 대한 접근 장벽을 낮추고, 집단 공간 키워드 질의로 여러 개의 장소를 그룹화하여 사용자의 요구를 만족시키므로 대량의 키워드에 대해서 정확한 결과를 반환할 수 있다.
- [0073] 집단 키워드 질의 처리 장치는 하드웨어, 펌웨어, 소프트웨어 또는 이들의 조합에 의해 로직회로 내에서 구현될 수 있고, 범용 또는 특정 목적 컴퓨터를 이용하여 구현될 수도 있다. 장치는 고정배선형(Hardwired) 기기, 필드 프로그램 가능한 게이트 어레이(Field Programmable Gate Array, FPGA), 주문형 반도체(Application Specific Integrated Circuit, ASIC) 등을 이용하여 구현될 수 있다. 또한, 장치는 하나 이상의 프로세서 및 컨트롤러를 포함한 시스템온칩(System on Chip, SoC)으로 구현될 수 있다.
- [0074] 집단 키워드 질의 처리 장치는 하드웨어적 요소가 마련된 컴퓨팅 디바이스 또는 서버에 소프트웨어, 하드웨어, 또는 이들의 조합하는 형태로 탑재될 수 있다. 컴퓨팅 디바이스 또는 서버는 각종 기기 또는 유무선 통신망과 통신을 수행하기 위한 통신 모듈 등의 통신장치, 프로그램을 실행하기 위한 데이터를 저장하는 메모리, 프로그램을 실행하여 연산 및 명령하기 위한 마이크로프로세서 등을 전부 또는 일부 포함한 다양한 장치를 의미할 수 있다.
- [0075] 도 5 및 도 6에서는 각각의 과정을 순차적으로 실행하는 것으로 기재하고 있으나 이는 예시적으로 설명한 것에 불과하고, 이 분야의 기술자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 5 및 도 6에 기재된 순서를 변경하여 실행하거나 또는 하나 이상의 과정을 병렬적으로 실행하거나 다른 과정을 추가하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이다.
- [0076] 본 실시예들에 따른 동작은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능한 매체에 기록될 수 있다. 컴퓨터 판독 가능한 매체는 실행을 위해 프로세서에 명령어를 제공하는 데 참여한 임의의 매체를 나타낸다. 컴퓨터 판독 가능한 매체는 프로그램 명령, 데이터 파일, 데이터 구조 또는 이들의 조합을 포함할 수 있다. 예를 들면, 자기 매체, 광기록 매체, 메모리 등이 있을 수 있다. 컴퓨터 프로그램은 네트워크로 연결된 컴퓨터 시스템 상에 분산되어 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 본 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드, 및 코드 세그먼트들은 본 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.
- [0077] 본 실시예들은 본 실시예의 기술 사상을 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

도면

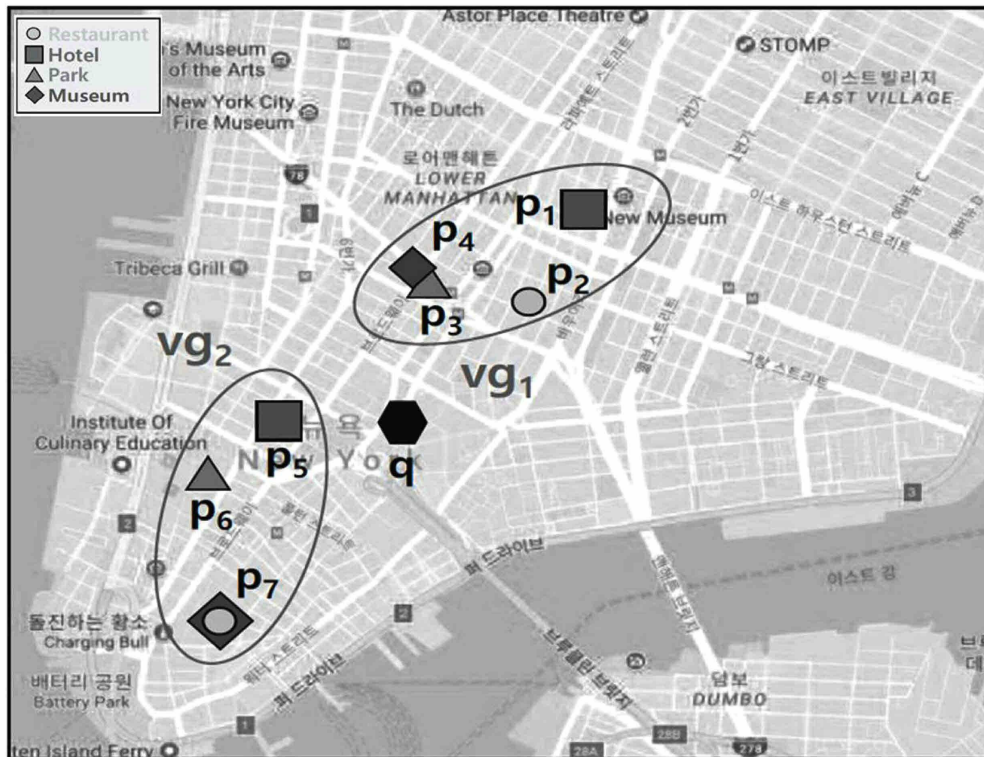
도면1



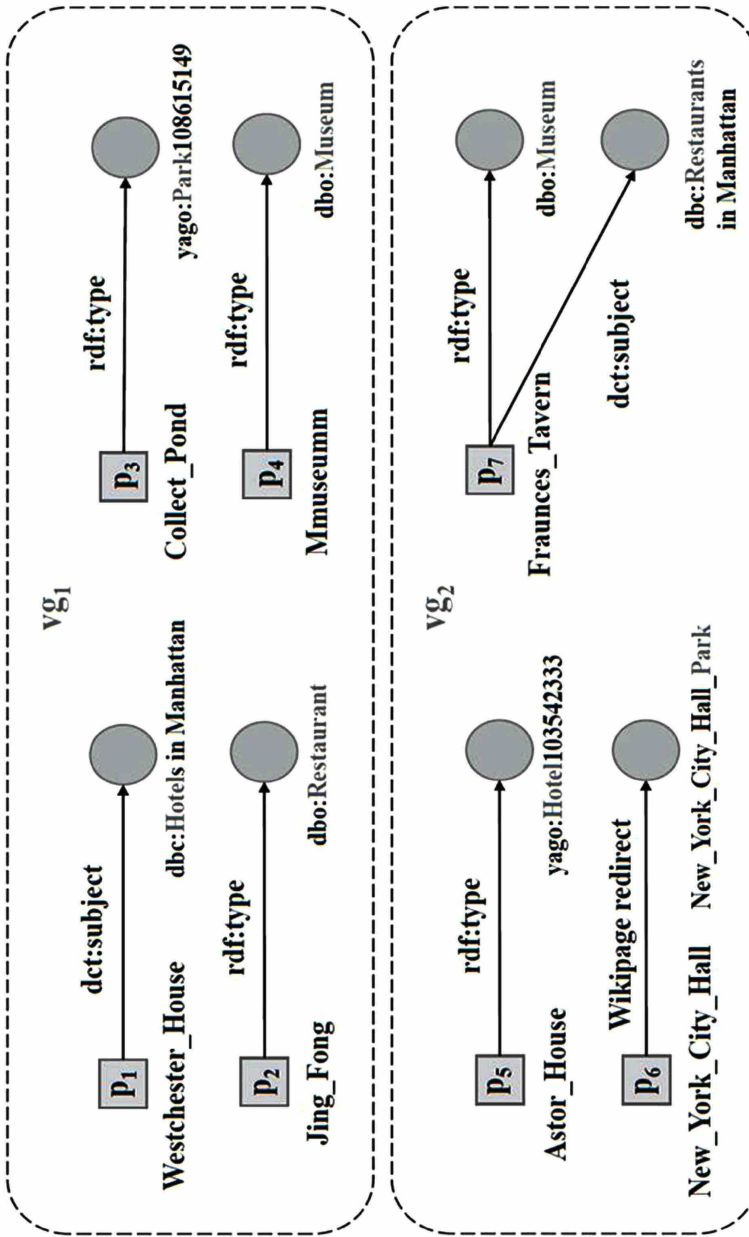
도면2



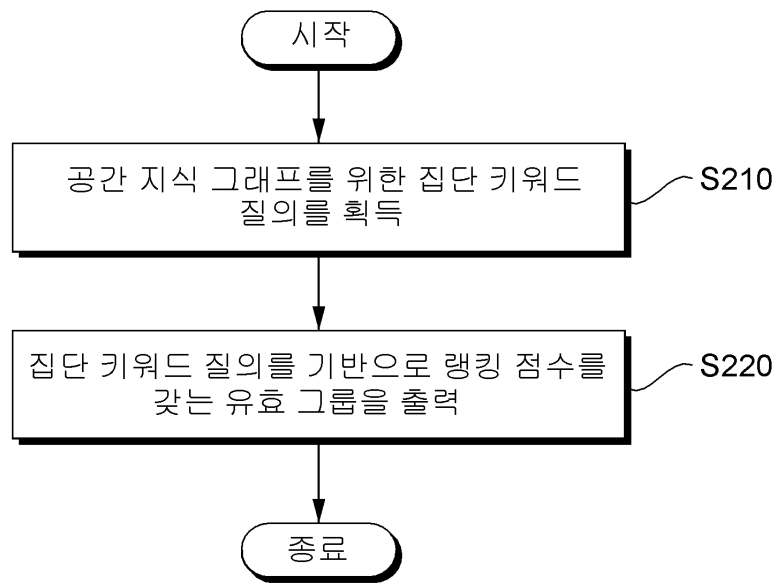
도면3



도면4



도면5



도면6

