



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년06월19일

(11) 등록번호 10-2822899

(24) 등록일자 2025년06월16일

(51) 국제특허분류(Int. Cl.)
H03M 7/30 (2006.01) G06F 16/901 (2019.01)

(52) CPC특허분류
H03M 7/3082 (2013.01)
G06F 16/9024 (2019.01)

(21) 출원번호 10-2024-0008958

(22) 출원일자 2024년01월19일
심사청구일자 2024년01월19일

(30) 우선권주장
1020230182426 2023년12월14일 대한민국(KR)

(56) 선행기술조사문헌
KR1020150125511 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자
이영구
경기도 수원시 영통구 청명북로 33, 431동 1704호
무하마드 우메르
경기도 수원시 영통구 봉영로1482번길 27, B04호

(74) 대리인
두호특허법인

전체 청구항 수 : 총 12 항

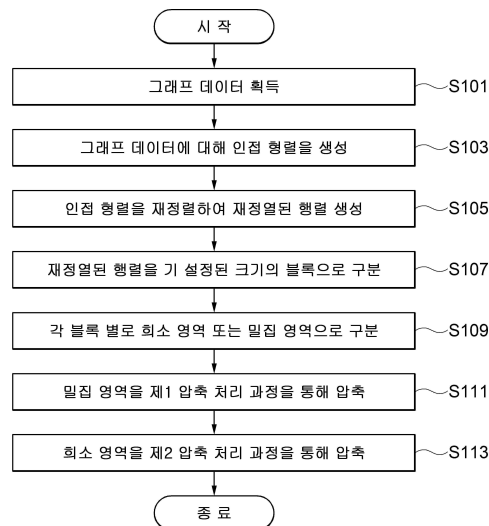
심사관 : 조춘근

(54) 발명의 명칭 영역 기반의 그래프 압축 방법 및 이를 수행하기 위한 컴퓨팅 장치

(57) 요약

영역 기반의 그래프 압축 방법 및 이를 수행하기 위한 컴퓨팅 장치가 개시된다. 개시되는 일 실시예에 따른 그래프 압축 방법은, 하나 이상의 프로세서들, 및 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서, 압축 대상이 되는 그래프 데이터를 획득하는 단계, 그래프 데이터에 대해 인접 행렬을 생성하는 단계, 인접 행렬을 재정렬하여 재정렬된 행렬 생성하는 단계, 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하는 단계를 포함한다.

대표도 - 도2



(52) CPC특허분류

H03M 7/3077 (2013.01)

H03M 7/70 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193818
과제번호	2021-0-00859-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발
연구과제명	초거대 그래프의 지능적 고속 처리를 위한 그래프 DBMS 기술 개발
과제수행기관명	포항공과대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서,

압축 대상이 되는 그래프 데이터를 획득하는 단계;

상기 그래프 데이터에 대해 인접 행렬을 생성하는 단계;

상기 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성하는 단계; 및

상기 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하는 단계를 포함하며,

상기 인접 행렬은, 상기 그래프 데이터에서 두 노드들이 연결되어 있으면 1로 표현하고, 두 노드들이 연결되어 있지 않으면 0으로 표현하는 이진 행렬이고,

상기 재정렬된 행렬도 행과 열의 인덱스 값이 0과 1로 구성된 이진 행렬이며,

상기 압축 처리를 수행하는 단계는,

상기 재정렬된 행렬에서 각 블록 별로 해당 영역의 밀집도에 따라 희소 영역 또는 밀집 영역으로 구분하는 단계를 포함하고,

상기 밀집 영역으로 구분된 영역은 제1 압축 처리 과정을 통해 압축하고, 상기 희소 영역으로 구분된 영역은 제2 압축 처리 과정을 통해 압축하며,

상기 제2 압축 처리 과정은,

상기 희소 영역의 수평 중심선을 기준으로 구분되는 각 수평 영역에 대해 수평 분해를 수행하여 수평 방향 시퀀스를 생성하는 단계; 및

상기 희소 영역의 수직 중심선을 기준으로 구분되는 각 수직 영역에 대해 수직 분해를 수행하여 수직 방향 시퀀스를 생성하는 단계를 포함하는, 그래프 압축 방법.

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

청구항 1에 있어서,

상기 구분하는 단계는,

각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 미만인 경우 해당 영역을 희소 영역으로 구분하고, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 이상인 경우 해당 영역을 밀집 영역으로 구분하는,

그래프 압축 방법.

청구항 6

청구항 1에 있어서,

상기 밀집도는,

각 블록에 해당하는 영역에서 상기 인덱스 값 1의 개수에 따라 결정되는, 그래프 압축 방법.

청구항 7

삭제

청구항 8

청구항 1에 있어서,

상기 제1 압축 처리 과정은,

상기 밀집 영역으로 구분된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하는, 그래프 압축 방법.

청구항 9

삭제

청구항 10

청구항 1에 있어서,

상기 제2 압축 처리 과정은,

상기 각 수평 영역의 수평 방향 시퀀스 및 각 수직 영역의 수직 방향 시퀀스에 대해 상기 인덱스 값에 따라 키-값 쌍을 생성하는 단계;

상기 수평 영역 및 상기 수직 영역 중 상기 키-값 쌍의 개수가 가장 적은 영역을 선택하는 단계; 및

상기 선택된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하는 단계를 더 포함하는, 그래프 압축 방법.

청구항 11

청구항 10에 있어서,

상기 키-값 쌍 중에서 키는 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 상기 인덱스 값 1이 있는 위치를 나타내고, 상기 키-값 쌍 중에서 값은 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 상기 인덱스 값 1이 연속되는 횟수를 나타내는, 그래프 압축 방법.

청구항 12

하나 이상의 프로세서들;

메모리; 및

하나 이상의 프로그램들을 포함하고,

상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성

되며,

상기 하나 이상의 프로그램들은,

압축 대상이 되는 그래프 데이터를 획득하기 위한 명령;

상기 그래프 데이터에 대해 인접 행렬을 생성하기 위한 명령;

상기 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성하기 위한 명령; 및

상기 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하기 위한 명령을 포함하며,

상기 인접 행렬은, 상기 그래프 데이터에서 두 노드들이 연결되어 있으면 1로 표현하고, 두 노드들이 연결되어 있지 않으면 0으로 표현하는 이진 행렬이고,

상기 재정렬된 행렬도 행과 열의 인덱스 값이 0과 1로 구성된 이진 행렬이며,

상기 압축 처리를 수행하기 위한 명령은

상기 재정렬된 행렬에서 각 블록 별로 해당 영역의 밀집도에 따라 희소 영역 또는 밀집 영역으로 구분하기 위한 명령을 포함하고,

상기 밀집 영역으로 구분된 영역은 제1 압축 처리 과정을 통해 압축하고, 상기 희소 영역으로 구분된 영역은 제2 압축 처리 과정을 통해 압축하며,

상기 제2 압축 처리 과정을 수행하기 위한 명령은,

상기 희소 영역의 수평 중심선을 기준으로 구분되는 각 수평 영역에 대해 수평 분해를 수행하여 수평 방향 시퀀스를 생성하기 위한 명령; 및

상기 희소 영역의 수직 중심선을 기준으로 구분되는 각 수직 영역에 대해 수직 분해를 수행하여 수직 방향 시퀀스를 생성하기 위한 명령을 포함하는, 컴퓨팅 장치.

청구항 13

삭제

청구항 14

삭제

청구항 15

청구항 12에 있어서,

상기 구분하기 위한 명령은,

각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 미만인 경우 해당 영역을 희소 영역으로 구분하고, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 이상인 경우 해당 영역을 밀집 영역으로 구분하는, 컴퓨팅 장치.

청구항 16

청구항 12에 있어서,

상기 제1 압축 처리 과정은, 상기 밀집 영역으로 구분된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하는, 컴퓨팅 장치.

청구항 17

삭제

청구항 18

청구항 12에 있어서,

상기 제2 압축 처리 과정을 수행하기 위한 명령은,

상기 각 수평 영역의 수평 방향 시퀀스 및 각 수직 영역의 수직 방향 시퀀스에 대해 상기 인덱스 값에 따라 키-값 쌍을 생성하기 위한 명령;

상기 수직 영역 및 상기 수평 영역 중 상기 키-값 쌍의 개수가 가장 적은 영역을 선택하기 위한 명령; 및

상기 선택된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하기 위한 명령을 더 포함하는, 컴퓨팅 장치.

청구항 19

청구항 18에 있어서,

상기 키-값 쌍 중에서 키는 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 상기 인덱스 값 1이 있는 위치를 나타내고, 상기 키-값 쌍 중에서 값은 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 상기 인덱스 값 1이 연속되는 횟수를 나타내는, 컴퓨팅 장치.

청구항 20

비일시적 컴퓨터 판독 가능한 저장 매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은 하나 이상의 명령어들을 포함하고, 상기 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금,

압축 대상이 되는 그래프 데이터를 획득하는 단계;

상기 그래프 데이터에 대해 인접 행렬을 생성하는 단계;

상기 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성하는 단계; 및

상기 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하는 단계를 수행하도록 하며,

상기 인접 행렬은, 상기 그래프 데이터에서 두 노드들이 연결되어 있으면 1로 표현하고, 두 노드들이 연결되어 있지 않으면 0으로 표현하는 이진 행렬이고,

상기 재정렬된 행렬도 행과 열의 인덱스 값이 0과 1로 구성된 이진 행렬이며,

상기 압축 처리를 수행하는 단계는,

상기 재정렬된 행렬에서 각 블록 별로 해당 영역의 밀집도에 따라 희소 영역 또는 밀집 영역으로 구분하는 단계를 포함하고,

상기 밀집 영역으로 구분된 영역은 제1 압축 처리 과정을 통해 압축하고, 상기 희소 영역으로 구분된 영역은 제2 압축 처리 과정을 통해 압축하며,

상기 제2 압축 처리 과정은,

상기 희소 영역의 수평 중심선을 기준으로 구분되는 각 수평 영역에 대해 수평 분해를 수행하여 수평 방향 시퀀스를 생성하는 단계; 및

상기 최소 영역의 수직 중심선을 기준으로 구분되는 각 수직 영역에 대해 수직 분해를 수행하여 수직 방향 시퀀스를 생성하는 단계를 포함하는, 비밀시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명의 실시예는 그래프 압축 기술과 관련된다.

배경 기술

[0003] 소셜 네트워크, 웹 네트워크, 생물학적 데이터 등과 같은 실제 세계의 데이터는 그래프로 표현될 수 있다. 그래프로 생성되는 데이터의 양은 나날이 증가하고 있는 추세이기 때문에 그래프를 압축할 수 있는 기술 또한 요구된다. 그래프 압축의 목표는 그래프를 최대한 압축하여 저장 공간을 최소화하는 것이다. 대부분의 그래프 압축 알고리즘은 다양한 마이닝 작업을 수행하기 전에 그래프를 압축하여 저장 공간을 줄이지만, 압축을 해제하기 위한 처리 오버헤드가 발생한다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국등록특허공보 제10-2009216호(2019.08.09)

발명의 내용

해결하려는 과제

[0006] 본 발명의 실시예는 그래프를 효율적으로 압축할 수 있는 영역 기반의 그래프 압축 방법 및 이를 수행하기 위한 컴퓨팅 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0008] 개시되는 일 실시예에 따른 그래프 압축 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 방법으로서, 압축 대상이 되는 그래프 데이터를 획득하는 단계; 상기 그래프 데이터에 대해 인접 행렬을 생성하는 단계; 상기 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성하는 단계; 및 상기 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하는 단계를 포함한다.

[0009] 상기 인접 행렬은, 상기 그래프 데이터에서 노드들이 연결되어 있으면 1로 표현하고, 노드들이 연결되어 있지 않으면 0으로 표현하는 이진 행렬이고, 상기 재정렬된 행렬도 인덱스 값이 0과 1로 구성된 이진 행렬일 수 있다.

[0010] 상기 블록의 크기는, 상기 재정렬된 행렬의 크기에 따라 설정될 수 있다.

[0011] 상기 압축 처리를 수행하는 단계는, 상기 재정렬된 행렬에서 각 블록 별로 해당 영역의 밀집도에 따라 최소 영역 또는 밀집 영역으로 구분하는 단계를 포함할 수 있다.

[0012] 상기 구분하는 단계는, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 미만인 경우 해당 영역을 최소 영역으로 구분하고, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 이상인 경우 해당 영역을 밀집 영역으로 구분할 수 있다.

[0013] 상기 밀집도는, 각 블록에 해당하는 영역에서 인덱스 값 1의 개수에 따라 결정될 수 있다.

[0014] 상기 압축 처리를 수행하는 단계는, 상기 밀집 영역으로 구분된 영역은 제1 압축 처리 과정을 통해 압축하고, 상기 최소 영역으로 구분된 영역은 제2 압축 처리 과정을 통해 압축할 수 있다.

- [0015] 상기 제1 압축 처리 과정은, 상기 밀집 영역으로 구분된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축할 수 있다.
- [0016] 상기 제2 압축 처리 과정은, 상기 희소 영역의 수평 중심선을 기준으로 구분되는 각 수평 영역에 대해 수평 분해를 수행하여 수평 방향 시퀀스를 생성하는 단계; 및 상기 희소 영역의 수직 중심선을 기준으로 구분되는 각 수직 영역에 대해 수직 분해를 수행하여 수직 방향 시퀀스를 생성하는 단계를 포함할 수 있다.
- [0017] 상기 제2 압축 처리 과정은, 상기 각 수평 영역의 수평 방향 시퀀스 및 각 수직 영역의 수직 방향 시퀀스에 대해 인덱스 값에 따라 키-값 쌍을 생성하는 단계; 상기 수평 영역 및 상기 수직 영역 중 상기 키-값 쌍의 개수가 가장 적은 영역을 선택하는 단계; 및 상기 선택된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하는 단계를 더 포함할 수 있다.
- [0018] 상기 키-값 쌍 중에서 키는 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 인덱스 값 1이 있는 위치를 나타내고, 상기 키-값 쌍 중에서 값은 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 인덱스 값 1이 연속되는 횟수를 나타낼 수 있다.
- [0019] 개시되는 일 실시예에 따른 컴퓨팅 장치는, 하나 이상의 프로세서들; 메모리; 및 하나 이상의 프로그램들을 포함하고, 상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성되며, 상기 하나 이상의 프로그램들은, 압축 대상이 되는 그래프 데이터를 획득하기 위한 명령; 상기 그래프 데이터에 대해 인접 행렬을 생성하기 위한 명령; 상기 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성하기 위한 명령; 및 상기 재정렬된 행렬을 기 설정된 크기의 블록으로 구분하고, 각 블록에 해당하는 영역 별로 압축 처리를 수행하기 위한 명령을 포함한다.
- [0020] 상기 인접 행렬은, 상기 그래프 데이터에서 노드들이 연결되어 있으면 1로 표현하고, 노드들이 연결되어 있지 않으면 0으로 표현하는 이진 행렬이고, 상기 재정렬된 행렬도 인덱스 값이 0과 1로 구성된 이진 행렬일 수 있다.
- [0021] 상기 압축 처리를 수행하기 위한 명령은, 상기 재정렬된 행렬에서 각 블록 별로 해당 영역의 밀집도에 따라 희소 영역 또는 밀집 영역으로 구분하기 위한 명령을 포함할 수 있다.
- [0022] 상기 구분하기 위한 명령은, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 미만인 경우 해당 영역을 희소 영역으로 구분하고, 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 이상인 경우 해당 영역을 밀집 영역으로 구분할 수 있다.
- [0023] 상기 압축 처리를 수행하기 위한 명령은, 상기 밀집 영역으로 구분된 영역은 제1 압축 처리 과정을 통해 압축하고, 상기 희소 영역으로 구분된 영역은 제2 압축 처리 과정을 통해 압축하며, 상기 제1 압축 처리 과정은, 상기 밀집 영역으로 구분된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축할 수 있다.
- [0024] 상기 제2 압축 처리 과정을 수행하기 위한 명령은, 상기 희소 영역의 수평 중심선을 기준으로 구분되는 각 수평 영역에 대해 수평 분해를 수행하여 수평 방향 시퀀스를 생성하기 위한 명령; 및 상기 희소 영역의 수직 중심선을 기준으로 구분되는 각 수직 영역에 대해 수직 분해를 수행하여 수직 방향 시퀀스를 생성하기 위한 명령을 포함할 수 있다.
- [0025] 상기 제2 압축 처리 과정을 수행하기 위한 명령은, 상기 각 수평 영역의 수평 방향 시퀀스 및 각 수직 영역의 수직 방향 시퀀스에 대해 인덱스 값에 따라 키-값 쌍을 생성하기 위한 명령; 상기 수평 영역 및 상기 수직 영역 중 상기 키-값 쌍의 개수가 가장 적은 영역을 선택하기 위한 명령; 및 상기 선택된 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축하기 위한 명령을 더 포함할 수 있다.
- [0026] 상기 키-값 쌍 중에서 키는 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 인덱스 값 1이 있는 위치를 나타내고, 상기 키-값 쌍 중에서 값은 상기 수평 방향 시퀀스 또는 상기 수직 방향 시퀀스에서 인덱스 값 1이 연속되는 횟수를 나타낼 수 있다.

발명의 효과

- [0028] 개시되는 실시예에 의하면, 그래프 데이터에 대해 인접 행렬을 생성하고 인접 행렬을 재정렬 한 후, 희소 영역에 대해서는 수평 분해 및 수직 분해를 수행하여 압축함으로써, 그래프 데이터를 효율적으로 압축하여 저장하는 데이터 양을 축소시킬 수 있게 된다.

도면의 간단한 설명

- [0030] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도
- 도 2는 본 발명의 일 실시예에 따른 그래프 압축 방법을 설명하기 위한 흐름도
- 도 3은 그래프 데이터에 대해 인접 행렬을 생성하는 일 예를 나타낸 도면
- 도 4는 본 발명의 일 실시예에서 인접 행렬을 슬래시 번 알고리즘에 따라 재정렬하여 재정렬된 행렬을 생성하는 상태를 나타낸 도면
- 도 5는 본 발명의 일 실시예에 따른 제2 압축 처리 과정을 개략적으로 나타낸 도면
- 도 6은 본 발명의 일 실시예에 따른 그래프 압축 장치의 구성을 나타낸 블록도

발명을 실시하기 위한 구체적인 내용

- [0031] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.
- [0032] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.
- [0033] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.
- [0035] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술된 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0036] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 그래프를 압축하기 위한 장치일 수 있다.
- [0037] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0038] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

- [0039] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0040] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.
- [0042] 도 2는 본 발명의 일 실시예에 따른 그래프 압축 방법을 설명하기 위한 흐름도이다. 도시된 흐름도에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.
- [0043] 도 2를 참조하면, 컴퓨팅 장치(12)는 압축 대상이 되는 그래프 데이터를 획득할 수 있다(S 101). 일 실시예에서, 그래프 데이터는 소셜 네트워크(social network) 데이터, 전자 메시지를 교환하는 네트워크 데이터, 생태계 네트워크 데이터, 단백질 네트워크 데이터 등을 포함할 수 있다. 또한, 그래프 데이터는 무방향 그래프(undirected graph)일 수도 있고, 유방향 그래프(directed graph)일 수도 있다.
- [0044] 다음으로, 컴퓨팅 장치(12)는 그래프 데이터에 대해 인접 행렬(adjacency matrix)을 생성할 수 있다(S 103). 그래프 데이터는 복수 개의 노드(node) 및 노드들 간의 연결 관계인 에지(edge)로 구성된다. 인접 행렬은 그래프 데이터의 구조를 나타낸 것으로, 그래프 데이터에서 두 노드가 연결되어 있으면 1로 표현하고, 연결되어 있지 않으면 0으로 표현하는 이진 행렬을 의미할 수 있다. 도 3은 그래프 데이터에 대해 인접 행렬을 생성하는 일 예를 나타낸 도면이다.
- [0045] 다음으로, 컴퓨팅 장치(12)는 인접 행렬을 압축 친화적인 형태로 재정렬하여 재정렬된 행렬을 생성할 수 있다(S 105). 인접 행렬의 경우, 두 노드 간 연결되어 있으면 1, 연결되어 있지 않으면 0으로 표현되기 때문에 그래프 데이터의 인접 행렬은 대칭되는 형태를 가지게 되고, 이는 그래프 압축률이 좋지 않은 형태가 된다. 따라서, 인접 행렬을 압축 친화적인 형태로 재정렬할 필요가 있게 된다. 이에, 개시되는 실시예에서는 인접 행렬을 기 설정된 알고리즘에 따라 재정렬하여 재정렬된 행렬을 생성할 수 있다.
- [0046] 일 실시예에서, 컴퓨팅 장치(12)는 인접 행렬을 슬래시 번(Slash Burn) 알고리즘에 따라 재정렬하여 재정렬된 행렬을 생성할 수 있다. 슬래시 번은 그래프 데이터에서 노드 재배치를 위한 기법으로, 실 세계 그래프 데이터의 멱법칙(power-law) 특성을 활용한다. 멱법칙 특성을 따르는 그래프 데이터는 높은 차수(degree)를 갖는 소수의 허브 노드들과 낮은 차수를 갖는 대부분의 노드들로 구성되는데, 그래프 데이터에서 허브 노드와 허브 노드를 연결하는 에지를 제거하면 그래프 데이터는 여러 개의 작은 연결 요소(connected component)들로 분리된다는 특징을 가진다. 이 연결 요소들은 가장 큰 크기를 가지는 거대 연결 요소(giant connected component)와 스포크(spoke)라고 불리는 나머지 비 거대 연결 요소들로 나눌 수 있는데, 거대 연결 요소는 원래 그래프 데이터에 비해 훨씬 작은 크기를 가진다는 특성이 있다.
- [0047] 슬래시 번 알고리즘은 이러한 실 세계 그래프 데이터의 특징을 이용하여 허브 노드와 스포크 노드들을 적절히 재배치하고, 거대 연결 요소에 대해 동일한 과정을 재귀적으로 수행하여 인접 행렬의 0이 아닌 값들이 집중되도록 한다. 슬래시 번 알고리즘은 기 공지된 기술이므로 이에 대한 자세한 설명은 생략하기로 한다. 한편, 여기서는 슬래시 번 알고리즘을 사용하는 것으로 설명하였으나, 이에 한정되는 것은 아니며 그 이외에 인접 행렬을 재정렬할 수 있는 다양한 알고리즘을 사용할 수 있다.
- [0048] 도 4는 본 발명의 일 실시예에서 인접 행렬을 슬래시 번 알고리즘에 따라 재정렬하여 재정렬된 행렬을 생성하는 상태를 나타낸 도면이다. 재정렬된 행렬을 인접 행렬에서 0이 아닌 값 즉, 1을 갖는 값들이 집중되도록 함으로써 압축 친화적인 행렬이 된다. 이때, 재정렬된 행렬도 행과 열의 인덱스 값이 0과 1로 구성된 이진 행렬이 될 수 있다.
- [0049] 다음으로, 컴퓨팅 장치(12)는 재정렬된 행렬을 기 설정된 크기의 블록으로 구분할 수 있다(S 107). 여기서, 블

록의 크기는 재정렬된 행렬의 크기에 따라 설정될 수 있다. 일 실시예에서, 컴퓨팅 장치(12)는 재정렬된 행렬을 $N \times N$ 크기(행과 열이 동일한 크기)의 블록으로 각각 구분할 수 있으나, 이에 한정되는 것은 아니다. 이때, 블록은 재정렬된 행렬을 복수 개의 블록으로 구분할 수 있는 크기로 설정될 수 있다.

[0050] 다음으로, 컴퓨팅 장치(12)는 재정렬된 행렬에서 각 블록 별로 해당 영역을 최소 영역 또는 밀집 영역으로 구분할 수 있다(S 109). 구체적으로, 컴퓨팅 장치(12)는 재정렬된 행렬에서 각 블록 별로 해당 블록에 해당하는 영역의 밀집도를 확인할 수 있다. 이때, 밀집도는 블록에 해당하는 영역에서 1의 개수에 따라 결정될 수 있다. 예를 들어, 블록이 4×4 크기인 경우, 재정렬된 행렬에서 각 블록에 해당하는 영역에는 16개의 인덱스 값(0 또는 1의 값)이 포함되게 된다. 이때, 각 블록에 해당하는 영역의 밀집도는 1의 개수에 따라 결정될 수 있다.

[0051] 컴퓨팅 장치(12)는 각 블록에 해당하는 영역의 밀집도(즉, 1의 개수)가 기 설정된 임계 밀집도 미만인 경우 해당 영역을 최소 영역으로 구분할 수 있다. 컴퓨팅 장치(12)는 각 블록에 해당하는 영역의 밀집도가 기 설정된 임계 밀집도 이상인 경우 해당 영역을 밀집 영역으로 구분할 수 있다. 여기서, 임계 밀집도는 블록의 크기에 따라 설정될 수 있다. 예를 들어, 임계 밀집도는 블록의 크기의 $1/3$ 로 설정될 수 있다. 즉, 블록의 크기가 4×4 인 경우, 임계 밀집도는 $16/3$ 으로 설정될 수 있다. 컴퓨팅 장치(12)는 재정렬된 행렬에서 밀집 영역과 최소 영역에 대해 각각 서로 다른 압축 처리 과정을 수행할 수 있다.

[0052] 다음으로, 컴퓨팅 장치(12)는 재정렬된 행렬에서 블록으로 구분된 영역이 밀집 영역인 경우, 제1 압축 처리 과정을 통해 압축할 수 있다(S 111). 구체적으로, 컴퓨팅 장치(12)는 재정렬된 행렬에서 블록으로 구분된 영역이 밀집 영역인 경우, 해당 영역을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축할 수 있다. 여기서, 기 설정된 압축 알고리즘은 공지된 그래프 압축 알고리즘일 수 있다. 압축 알고리즘이 복수 개인 경우, 각 압축 알고리즘에 따라 압축된 데이터 양이 가장 적은 것을 사용할 수 있다. 제1 압축 처리 과정은 밀집 영역으로 구분된 영역을 기 설정된 압축 알고리즘에 따라 바로 압축하는 과정일 수 있다.

[0053] 다음으로, 컴퓨팅 장치(12)는 재정렬된 행렬에서 블록으로 구분된 영역이 최소 영역인 경우, 제2 압축 처리 과정을 통해 압축할 수 있다(S 113). 제2 압축 처리 과정은 제1 압축 처리 과정과는 다른 과정이다. 제2 압축 처리 과정은 최소 영역으로 구분된 영역을 바로 압축하는 것이 아니라 별도의 추가 처리 과정을 거치게 된다. 한편, S 111과 S 113은 반드시 선후 관계를 가지는 것은 아니며 동시에 처리될 수 있다.

[0054] 도 5는 본 발명의 일 실시예에 따른 제2 압축 처리 과정을 개략적으로 나타낸 도면이다. 여기서는 블록의 크기가 4×4 인 경우를 일 예로 하였다. 도 5를 참조하면, 컴퓨팅 장치(12)는 블록으로 구분된 영역이 최소 영역인 경우, 최소 영역에 대해 수평 분해 및 수직 분해를 각각 수행할 수 있다.

[0055] 여기서, 수평 분해는 최소 영역의 수평 중심선(H)을 기준으로 구분되는 각 수평 영역(CV_1 , CV_2)에 대해 수평 방향으로 인덱스 값(0 또는 1의 값)을 순차적으로 배열하여 인덱스 값들의 수평 방향 시퀀스를 생성하는 것일 수 있다. 도 5를 보면, 수평 영역(CV_1)에 대해서는 "11010101"이라는 수평 방향 시퀀스가 생성되고, 수평 영역(CV_2)에 대해서는 "11010011"이라는 수평 방향 시퀀스가 생성된 것을 볼 수 있다.

[0056] 또한, 수직 분해는 최소 영역의 수직 중심선(V)을 기준으로 구분되는 각 수직 영역(RV_1 , RV_2)에 대해 수직 방향으로 인덱스 값을 순차적으로 배열하여 인덱스 값들의 수직 방향 시퀀스를 생성하는 것일 수 있다. 도 5를 보면, 수직 영역(RV_1)에 대해서는 "10101110"이라는 수직 방향 시퀀스가 생성되고, 수직 영역(RV_2)에 대해서는 "00011111"이라는 수직 방향 시퀀스가 생성된 것을 볼 수 있다.

[0057] 다음으로, 컴퓨팅 장치(12)는 각 수평 영역의 수평 방향 시퀀스와 각 수직 영역의 수직 방향 시퀀스에 대해 인덱스 값에 따라 키-값 쌍을 생성할 수 있다. 여기서, 키(key)는 수평 방향 시퀀스 또는 수직 방향 시퀀스에서 인덱스 값 1이 있는 위치를 나타낼 수 있다. 또한, 값(value)은 수평 방향 시퀀스 또는 수직 방향 시퀀스에서 인덱스 값 1이 연속되는 횟수를 나타낼 수 있다.

[0058] 이에 의하면, 수평 영역(CV_1)의 수평 방향 시퀀스인 "11010101"에 대해서는 (1, 2), (4, 1), (6, 1), (8, 1)이라는 키-값 쌍들이 생성될 수 있다. 또한, 수평 영역(CV_2)의 수평 방향 시퀀스인 "11010011"에 대해서는 (1, 2), (4, 1), (7, 2)라는 키-값 쌍들이 생성될 수 있다. 또한, 수직 영역(RV_1)의 수직 방향 시퀀스인 "10101110"에 대해서는 (1, 1), (3, 1), (5, 3)이라는 키-값 쌍들이 생성될 수 있다. 또한, 수직 영역(RV_2)의 수직 방향 시퀀스인 "00011111"에 대해서는 (4, 5)라는 키-값 쌍이 생성될 수 있다.

- [0059] 다음으로, 컴퓨팅 장치(12)는 회소 영역의 수평 영역 및 수직 영역들 중 키-값 쌍의 개수가 가장 작은 영역(도 5에서는 Rv_2)을 선택할 수 있다.
- [0060] 다음으로, 컴퓨팅 장치(12)는 선택된 영역(수평 영역 또는 수직 영역)을 하나 이상의 기 설정된 압축 알고리즘을 통해 압축할 수 있다. 압축 알고리즘이 복수 개인 경우, 각 압축 알고리즘에 따라 압축된 데이터 량이 가장 적은 것을 사용할 수 있다.
- [0061] 개시되는 실시예에 의하면, 그래프 데이터에 대해 인접 행렬을 생성하고 인접 행렬을 재정렬 한 후, 회소 영역에 대해서는 수평 분해 및 수직 분해를 수행하여 압축함으로써, 그래프 데이터를 효율적으로 압축하여 저장하는 데이터 양을 축소시킬 수 있게 된다.
- [0063] 도 6은 본 발명의 일 실시예에 따른 그래프 압축 장치의 구성을 나타낸 블록도이다. 도 6을 참조하면, 그래프 압축 장치(100)는 행렬 처리 모듈(102), 영역 분류 모듈(104), 및 압축 처리 모듈(106)을 포함할 수 있다.
- [0064] 행렬 처리 모듈(102)은 그래프 데이터를 획득하여 인접 행렬을 생성할 수 있다. 또한, 행렬 처리 모듈(102)은 인접 행렬을 재정렬하여 재정렬 된 행렬을 생성할 수 있다.
- [0065] 영역 분류 모듈(104)은 재정렬 된 행렬을 기 설정된 크기의 블록으로 구분할 수 있다. 또한, 영역 분류 모듈(104)은 재정렬된 행렬에서 각 블록 별로 해당 영역을 회소 영역 또는 밀집 영역으로 구분할 수 있다.
- [0066] 압축 처리 모듈(106)은 재정렬된 행렬에서 밀집 영역으로 구분된 영역은 기 설정된 제1 압축 처리 과정을 통해 압축할 수 있다. 또한, 압축 처리 모듈(106)은 재정렬된 행렬에서 회소 영역으로 구분된 영역은 기 설정된 제2 압축 처리 과정을 통해 압축할 수 있다.
- [0067] 본 명세서에서 모듈이라 함은, 본 발명의 기술적 사상을 수행하기 위한 하드웨어 및 상기 하드웨어를 구동하기 위한 소프트웨어의 기능적, 구조적 결합을 의미할 수 있다. 예컨대, 상기 "모듈"은 소정의 코드와 상기 소정의 코드가 수행되기 위한 하드웨어 리소스의 논리적인 단위를 의미할 수 있으며, 반드시 물리적으로 연결된 코드를 의미하거나, 한 종류의 하드웨어를 의미하는 것은 아니다.
- [0069] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

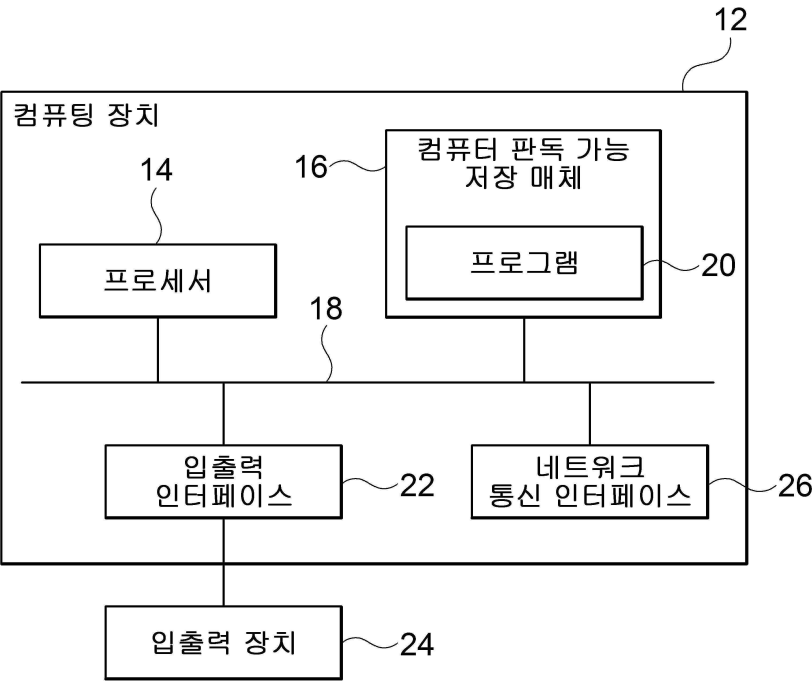
부호의 설명

- [0071] 10 : 컴퓨팅 환경
- 12 : 컴퓨팅 장치
- 14 : 프로세서
- 16 : 컴퓨터 판독 가능 저장 매체
- 18 : 통신 버스
- 20 : 프로그램
- 22 : 입출력 인터페이스
- 24 : 입출력 장치
- 26 : 네트워크 통신 인터페이스
- 100 : 그래프 압축 장치
- 102 : 행렬 처리 모듈
- 104 : 영역 분류 모듈
- 106 : 압축 처리 모듈

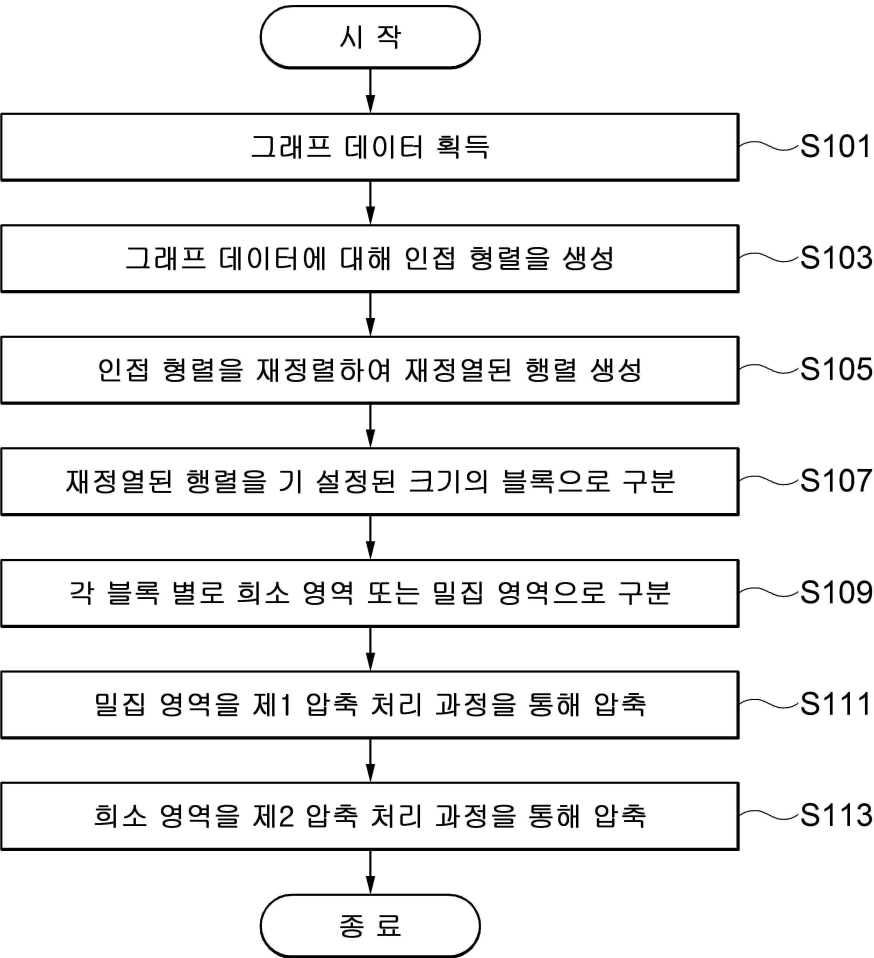
도면

도면1

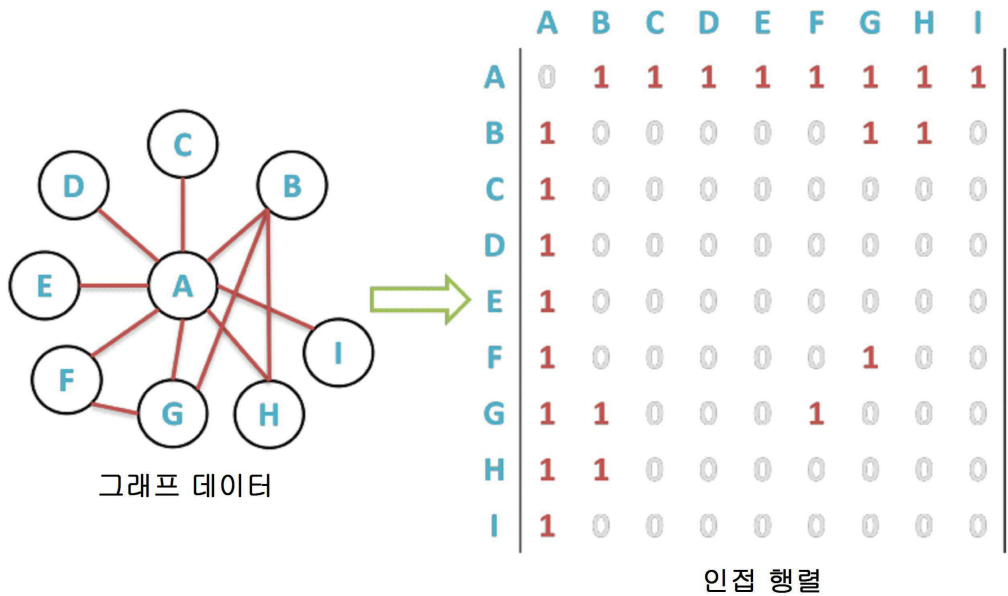
10



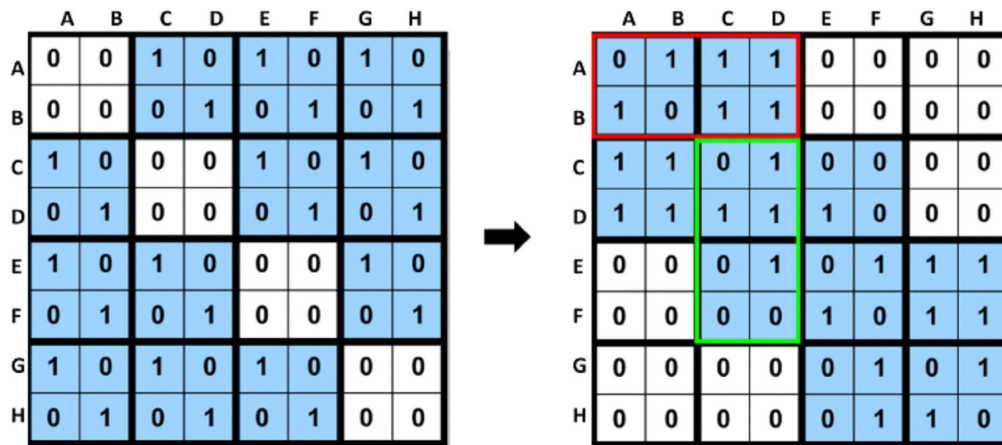
도면2



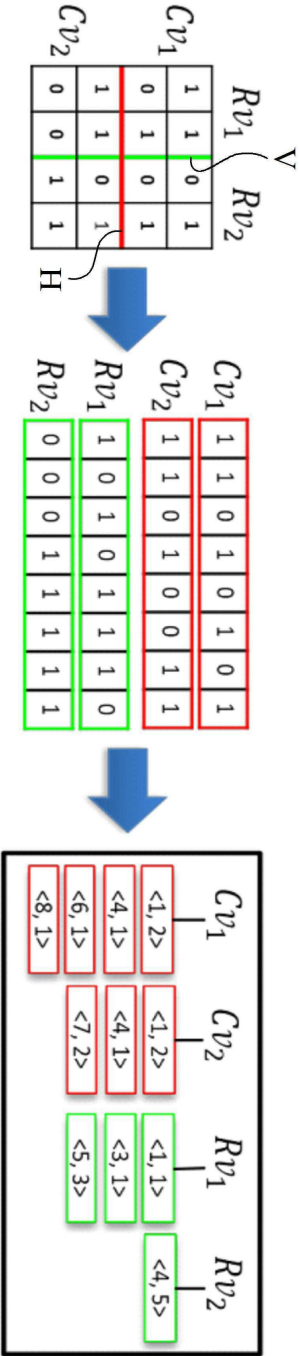
도면3



도면4



도면5



도면6

100

