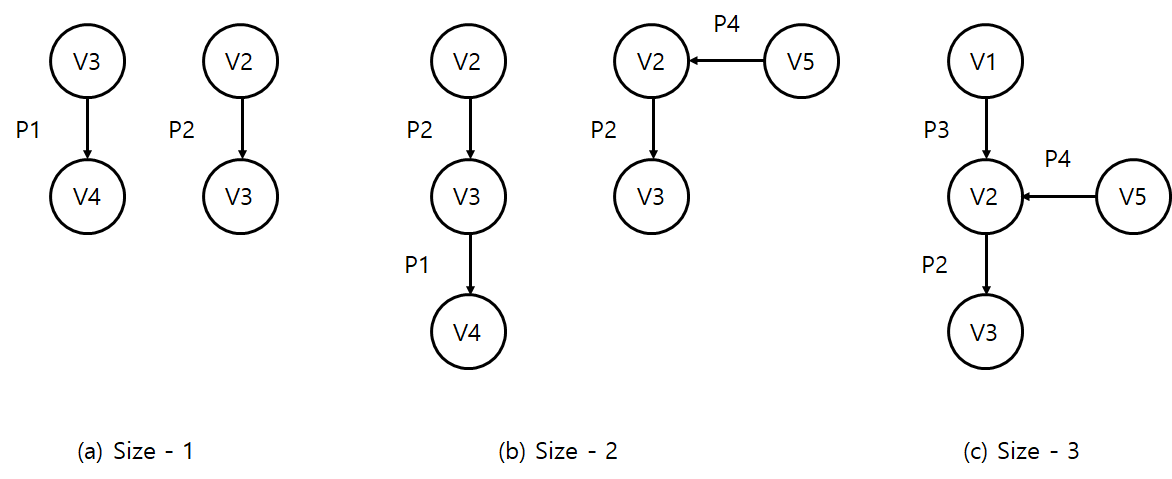
**RG-index**

RG-index의 물리적 구조는 다음과 같다.



[Fgure 1 : k-Neighborhood subgraph]

RG-index는 트리플 필터링을 위한 필터데이터에 직접적인 접근을 제공하고, RDF 데이터베이스 안에 서브그래프 패턴을 위한 하나의 노드리스트의 집합을 유지한다.

하나의 노드리스트는 그래프 패턴의 모든 노드를 구성하고, query 노드와 매칭 되는 ID들을 포함한다. RG-index는 RDF 데이터베이스의 그래프 패턴을 인덱스한다.

꼭지점이 모두 변수이고 가장자리 레이블이 모두 한정되어있는, 즉 변수가 아닌 그래프 패턴만 고려된다. 그래프 패턴을 다음과 같이 정의된다.

(1) Graph pattern

그래프 패턴은 연결 그래프로 vertex은 모두 변수이고 edge의 레이블은 모두 URI이다. 그래프 패턴은 트리플 패턴이 조건을 만족시키는 SPARQL 쿼리 gp로 볼 수 있다.

DRW00002ab444d1

(2) Vertex list

그래프 패턴의 vertex 리스트는 다음과 같이 형식적으로 정의된다.

그래프 패턴 DRW00002ab444d3와 vertex DRW00002ab444d5가 주어졌을 때, vertex 리스트 DRW00002ab444d7는 DRW00002ab444d9라 할 수 있고 v 사영시킨 집합 G를 뜻한다. G에 대한 모든 vertex 리스트의 집합은 DRW00002ab444db로 나타낸 수 있다.

(3) RG-index

쿼리 그래프로 그래프 패턴을 다루고, vertex리스트를 정의하기 위해 DRW00002ab444dd를 사용한다. RG-index의 정의는 다음과 같다.

최대 길이가 maxL를 가진 RDF 데이터베이스 D에서 RG-index는 DRW00002ab444df의 집합으로 정의할 수 있다. 여기서 G는 크기가 maxL보다 작거나 같은 D의 그래프 패턴이다.

이 DFS subscription 및 DFS edge 표현을 사용하면 그래프 패턴을 DFS edge의 시퀀스인 DFS 코드에 매핑 할 수 있다. DFS 코드에서 그래프 패턴의 edge에 대한 DFS edge는 다음과 같이 순서가 지정된다.

- forward edge가 발견되면 순서대로 정렬된다.

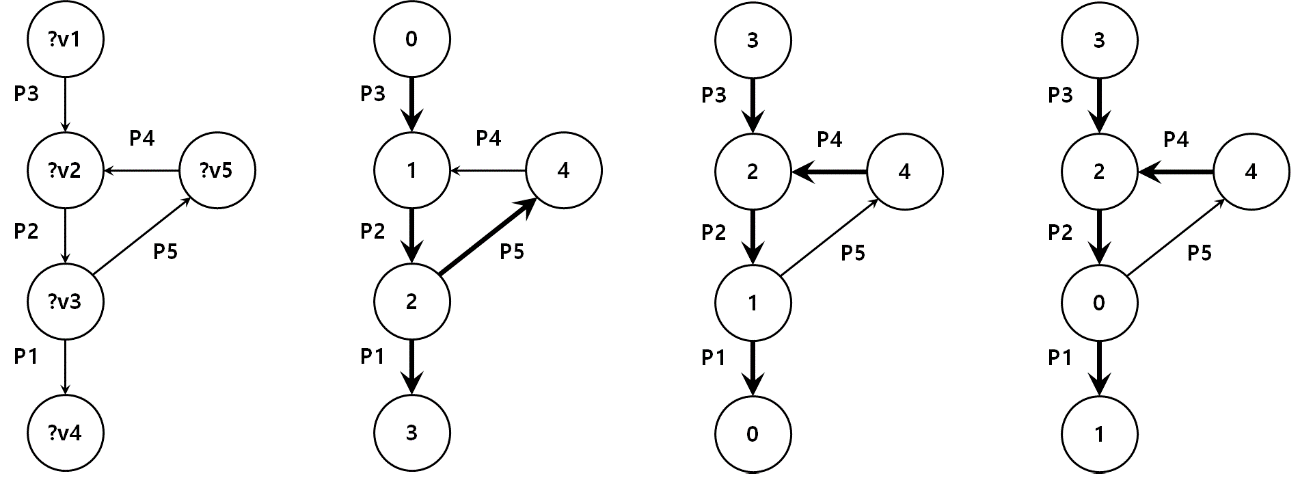
- 정점 v가 주어지면 모든 forward edge는 v를 가리키는 forward edge 뒤에 나타

나야한다.

- 같은 vertex로부터 backward edges 사이, 즉 (vi, vj), (vi, vk)에서, 만약 j < k라면

(vi, vj)는 (vi, vk) 전에 나타나야 한다.

gSpan은 DFS 코드 중 사전 식 순서를 정의한다. 두 개의 주어진 DFS edge의 경우 순서는 먼저 두 개의 인덱스에 의해 결정된 다음 edge 레이블에 의해 결정되고 마지막으로 방향에 따라 결정된다.



(a) Graph Pattern (b) DFS Tree (c) DFS Tree (d) DFS Tree

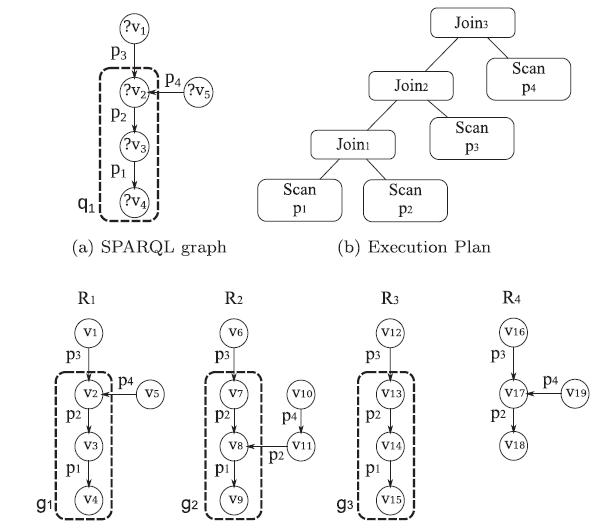
[Figure 2 : DFS subscriptions]

Figure 2는 그래프패턴(a)와 세 개의 subscriptions(b~d)를 보여준다. 각 vertex는 각 subscription으로 처리된다. Forward egde는 두껍게 표시하고 backward edge는 얇게로 표시했다.

Table 2는 세 가지 subscription에 대한 DFS 코드를 보여준다. DFS 코드 중 순서는 (d) <(c) <(d)이고, (d)는 그래프 패턴의 최소 DFS 코드이다.

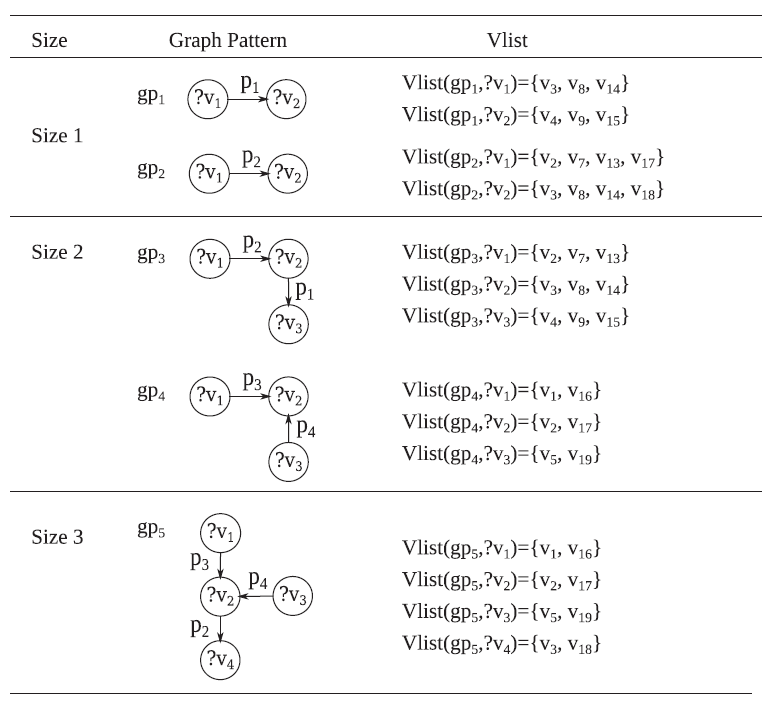
[Table 2 : Figure 9에 대한 그래프 패턴의 DFS 코드]

|  |  |  |  |
| --- | --- | --- | --- |
| Edge | (b) | (c) | (d) |
| 1  2  3  4  5 | <0, 1, p3, DRW00002ab444e1>  <1, 2, p2, DRW00002ab444e3>  <1, 4, p4, DRW00002ab444e5>  <2, 3, p1, DRW00002ab444e7>  <2, 4, p5, DRW00002ab444e9> | <0, 1, p1, DRW00002ab444eb>  <1, 2, p2, DRW00002ab444ed>  <1, 4, p5, DRW00002ab444ef>  <2, 3, p3, DRW00002ab444f1>  <2, 4, p4, DRW00002ab444f3> | <0, 1, p1, DRW00002ab444f5>  <1, 2, p2, DRW00002ab444f7>  <1, 4, p5, DRW00002ab444f9>  <2, 3, p3, DRW00002ab444fb>  <2, 4, p4, DRW00002ab444fd> |



[Figure 3] RDF graph and SPARQL query graph

그림 4는 그림의 RDF그래프에 Figure 3.c에 예시된 RG-index 예시를 보여 준다. 3c 이 RG-index는 RDF그래프에 5개의 그래프 패턴을 나타내고 있으며 14Kbists를 가지고 있다. RG-index를 사용하여 쿼리 그래프의 각 정점에 대한 후보 정점을 구할 수 있습니다. 후보자 정점 세트는 관련된 Vlist를 교차하여 획득한다.



[Figure 4] RG-index(maxL=3)