Câu 1: (4 điểm)

a) So sánh kết quả các phương pháp tìm tập phổ biến về: kết quả, không gian lưu trữ và thời gian khai

thác.

b) Cho cơ sở dữ liệu giao tác như sau:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mã giao dịch | A | B | C | D | E | F |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| 2 | 1 | 0 | 1 | 1 | 1 | 0 |
| 3 | 0 | 0 | 1 | 1 | 0 | 0 |
| 4 | 0 | 0 | 0 | 1 | 1 | 0 |
| 5 | 1 | 1 | 1 | 0 | 1 | 0 |

1. Tìm tất cả các tập phổ biến (FIs) có trong CSDL với minSup = 40% theo phương pháp FP-Tree

**Giải:**

minSup = 40% = 40 \* 5 / 100 = 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Danh Mục | A | B | C | D | E | F |
| σ | 2 | 1 | 4 | 3 | 4 | 1 |

Sắp xếp theo σ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Danh Mục | C | E | D | A | B | F |
| σ | 4 | 4 | 3 | 2 | 1 | 1 |

Xây dựng cây như sau:

C:4

E:1

D:1

D:1

E:3

D:1

A:1

A:1

{}

Chiếu trên A:2, ta có

* CSDL cục bộ: {CED:1, CE:1}
* Cây cục bộ:

{}

C:2

E:2

* Các tập con là: {∅, E:2, C:2, EC:2}
  + Các tập phổ biến là: {A:2, AE:2, AC:2, AEC:2}

Chiếu trên D:3, ta có

* CSDL cục bộ: {CE:1, C:1, E:1}
* Cây cục bộ

{}

C:2

* Các tập con là: {∅, C:2}
  + Các tập phổ biến là: {D:3, DC:2}

Chiếu trên E:4, ta có

* CSDL cục bộ: {C:3}
* Cây cục bộ:

{}

C:3

* Các tập con là: {∅, C:3}
  + Các tập phổ biến là: {E:4, EC:3}

Chiếu trên C:4, ta được {0}

* Tập phổ biến là: {C:4}

ii) Tìm tất cả các tập phổ biến đóng theo thuận toán CHARM với minSup = 40%