

# 추천시스템 구현 프로젝트

## Recommender System

## 1. 추천시스템 이해

- 1) 추천시스템 개념
- 2) 추천시스템 방법론
- 3) 기업에서의 추천시스템
- 4) 추천시스템 Trend

## 2. 추천시스템 구현

- 1) 과거 추천시스템
- 2) 콘텐츠 기반의 추천시스템
- 3) 협업필터링 기반 추천시스템
- 4) 딥러닝 기반의 추천시스템



# 추천시스템 구현 - 과거의 추천시스템

- "연관 분석"에 의한 추천
  - 룰 기반의 모델로서 상품과 상품 사이에 어떤 연관이 있는지 찾아내는 알고리즘
  - 장바구니 분석





# 추천시스템 구현 - 과거의 추천시스템

## ■ 연관 분석 평가 지표

- 지지도 - 전체 거래 중 상품 A, B를 동시에 구매하는 거래의 비율
- 신뢰도 - 상품 A가 포함된 거래 중 상품 B를 포함하는 거래의 비율
- 향상도 - 상품 B를 구매한 고객 대비 상품 A를 구매한 후 상품 B를 구매하는 고객에 대한 확률

✓ support (지지도)

*For the rule  $A \rightarrow B$ ,  
 $support(A) = P(A)$*

✓ lift (향상도)

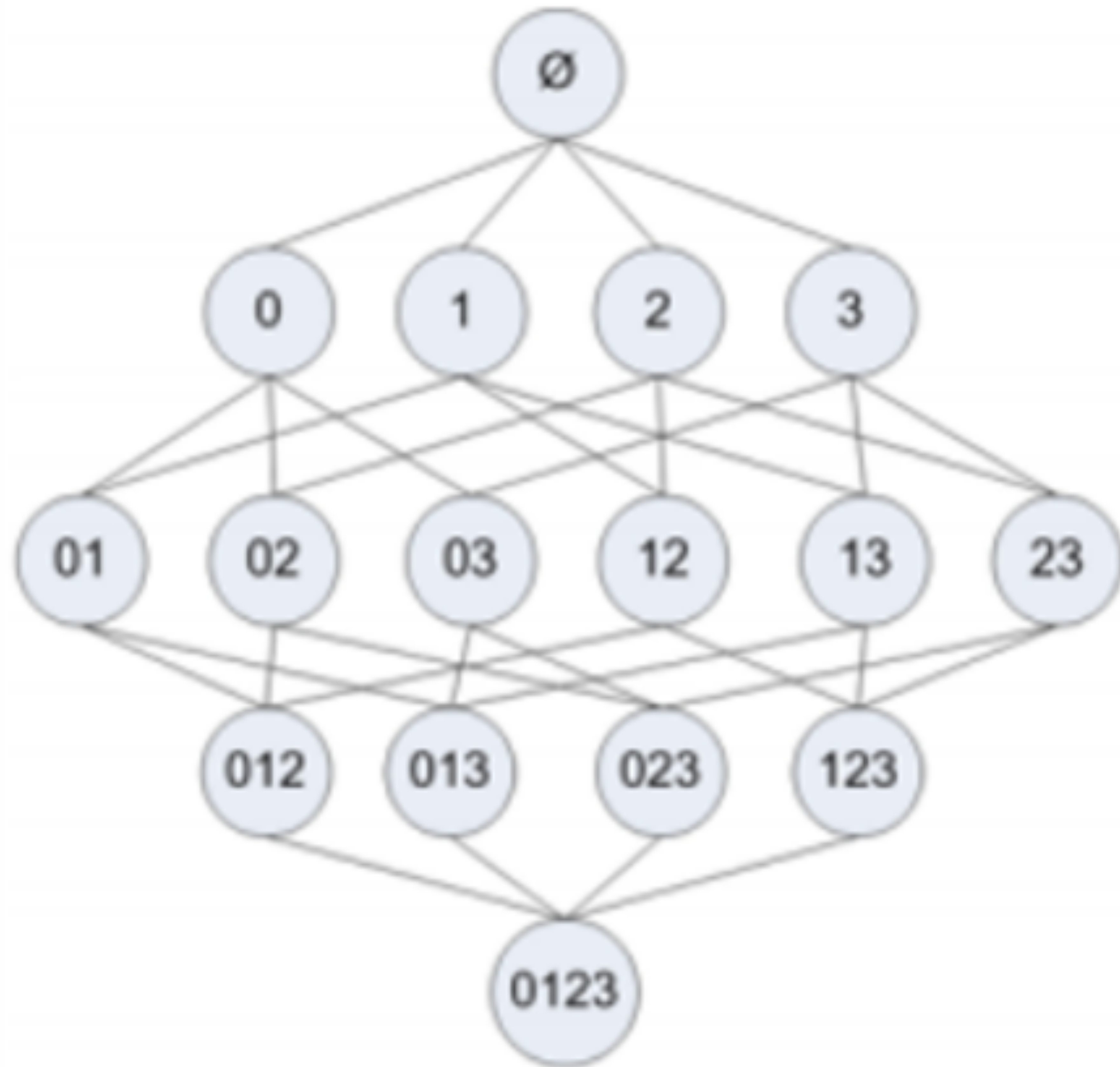
$$lift(A \rightarrow B) = \frac{P(A, B)}{P(A) \cdot P(B)}$$

✓ confidence (신뢰도)

$$confidence(A \rightarrow B) = \frac{P(A, B)}{P(A)}$$

# 추천시스템 구현 - 과거의 추천시스템

- 연관 분석 알고리즘 : 가능한 모든 경우의 수를 탐색해서 지지도, 신뢰도, 향상도가 높은 규칙을 찾아내는 방식



상품이 4개일 때, 전체 경우의 수

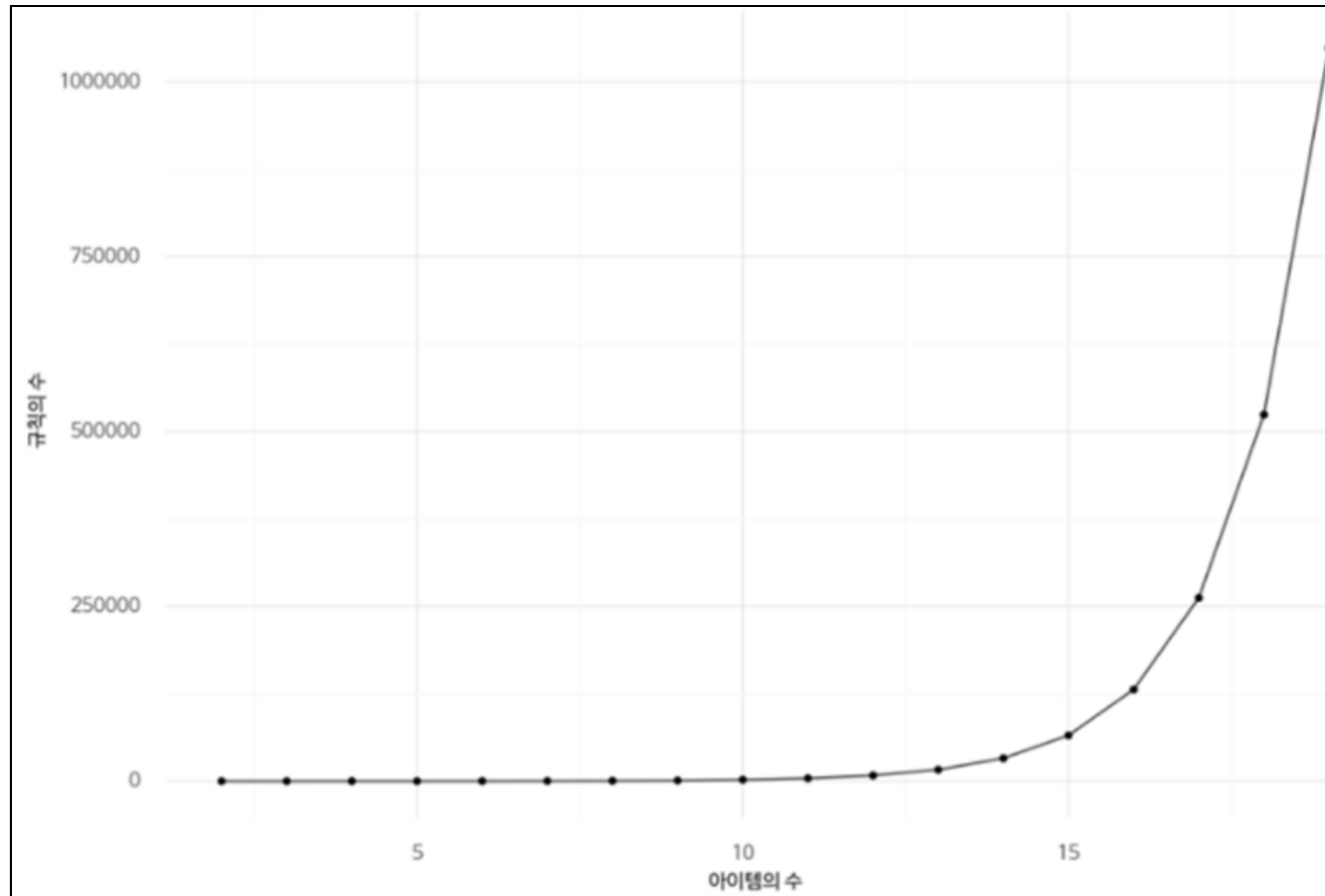
- $4C1 : 4$
- $4C2 : 6$
- $4C3 : 4$
- $4C4 : 1$

전체 경우의 수 :  $4 + 6 + 4 + 1 = 15$



# 추천시스템 구현 - 과거의 추천시스템

- 연관 분석 알고리즘 : 가능한 모든 경우의 수를 탐색해서 지지도, 신뢰도, 향상도가 높은 규칙을 찾아내는 방식



\* 아이템 수에 따른 규칙의 수

# 추천시스템 구현 - 과거의 추천시스템

- Apriori 알고리즘 : "빈번한 아이템 셋은 하위 아이템 셋 또한 빈번할 것이다."  
"빈번하지 않은 아이템 셋은 하위 아이템 셋도 빈번하지 않다."



## 아이디어

- {2, 3}의 지지도 > {0, 2, 3}, {1, 2, 3}의 지지도
- $P(\text{item 2, item 3}) > P(\text{item 0, item 2, item 3})$
  - $P(\text{item 2, item 3}) > P(\text{item 1, item 2, item 3})$



# 추천시스템 구현 - 과거의 추천시스템

- Apriori 알고리즘 : "빈번한 아이탬 셋은 하위 아이탬 셋 또한 빈번할 것이다."  
"빈번하지 않은 아이탬 셋은 하위 아이탬 셋도 빈번하지 않다."



- k개의 아이탬을 가지고 단일항목집단 생성
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

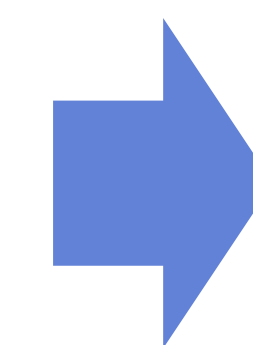


# 추천시스템 구현 - 과거의 추천시스템

- Apriori 알고리즘(예)

상품	우유, 기저귀, 쥬스, 양상추, 맥주
----	----------------------

거래 번호	상품 목록
0	우유, 기저귀, 쥬스
1	양상추, 기저귀, 맥주
2	우유, 양상추, 기저귀, 맥주
3	양상추, 맥주



거래 번호	우유	양상추	기저귀	쥬스	맥주
0	1	0	1	1	0
1	0	1	1	0	1
2	1	1	1	0	1
3	0	1	0	0	1



# 추천시스템 구현 - 과거의 추천시스템

## ▪ Apriori 알고리즘(예)

- **k개의 아이템을 가지고 단일항목집단 생성**
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목 집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

우유, 기저귀, 주스, 양상추, 맥주



# 추천시스템 구현 - 과거의 추천시스템

## ▪ Apriori 알고리즘(예)

- k개의 아이টে을 가지고 단일항목집단 생성
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목 집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

최소 지지도 : 0.5 이상

$P(\text{우유}) : 0.5$

$P(\text{양상추}) : 0.75$

$P(\text{기저귀}) : 0.75$

~~$P(\text{쥬스}) : 0.25$~~

$P(\text{맥주}) : 0.75$



# 추천시스템 구현 - 과거의 추천시스템

## ▪ Apriori 알고리즘(예)

- K개의 아이টে을 가지고 단일항목집단 생성
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목 집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

{우유, 양상추, 기저귀, 맥주}



2개 항목집단 생성

{우유, 양상추}

{우유, 기저귀}

{우유, 맥주}

{양상추, 기저귀}

{양상추, 맥주}

{기저귀, 맥주}



# 추천시스템 구현 - 과거의 추천시스템

## ▪ Apriori 알고리즘(예)

- k개의 아이টে을 가지고 단일항목집단 생성
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목 집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 k개의 k-item frequent set을 생성할 때까지 반복

최소 지지도 : 0.5 이상

~~$P(\text{우유}, \text{양상추}) : 0.25$~~

$P(\text{우유}, \text{기저귀}) : 0.5$

~~$P(\text{우유}, \text{맥주}) : 0.25$~~

$P(\text{양상추}, \text{기저귀}) : 0.5$

$P(\text{양상추}, \text{맥주}) : 0.75$

$P(\text{기저귀}, \text{맥주}) : 0.5$



# 추천시스템 구현 - 과거의 추천시스템

## ▪ Apriori 알고리즘(예)

- K개의 아이টে을 가지고 단일항목집단 생성
- 단일항목집단에서 최소 지지도 이상의 항목만 선택
- 앞 단계에서 선택된 항목만을 대상으로 2개 항목 집단 생성
- 2개 항목집단에서 최소 지지도 혹은 신뢰도 이상의 항목만 선택
- 위의 과정을 K개의 k-item frequent set을 생성할 때까지 반복



# 추천시스템 구현 - 과거의 추천시스템

- Apriori 알고리즘 : "빈번한 아이템 셋은 하위 아이템 셋 또한 빈번할 것이다."  
"빈번하지 않은 아이템 셋은 하위 아이템 셋도 빈번하지 않다."
- 장점
  - 원리가 간단하여 사용자가 쉽게 이해할 수 있고 의미를 파악할 수 있음
  - 유의한 연관성을 갖는 구매패턴을 찾아줌
- 단점
  - 데이터가 클 경우(item이 많은 경우)에 속도가 느리고 연산량이 많음
  - 실제 사용시에 많은 연관상품들이 나타나는 단점이 있음



# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth : FP Tree 구조를 사용, Apriori의 속도 측면의 단점을 개선한 알고리즘
- 알고리즘

- 모든 거래를 확인하여, 각 아이템마다의 지지도(support)를 계산하고 최소 지지도 이상의 아이템만 선택
- 모든 거래에서 빈도가 높은 아이템 순서대로 순서를 정렬
- 부모 노드를 중심으로 거래를 자식노드로 추가해주면서 tree를 생성
- 새로운 아이템이 나올 경우에는 부모노드부터 시작하고, 그렇지 않으면 기존의 노드에서 확장
- 지지도가 낮은 순서부터 시작하여 조건부 패턴을 생성, 모든 아이템에 대해 반복
- 조건부 패턴을 기반으로 패턴 생성

# 추천시스템 구현 - 과거의 추천시스템

## ■ FP-Growth 알고리즘(예)

1. 모든 거래를 확인하여, 각 아이템마다의 지지도(support)를 계산하고 최소 지지도 이상의 아이템만 선택
2. 모든 거래에서 빈도가 높은 아이템 순서대로 순서를 정렬

거래번호	아이템
0	우유, 기저귀, 주스
1	양상추, 기저귀, 맥주
2	우유, 양상추, 기저귀, 맥주
3	양상추, 맥주



거래번호	정렬된 아이템
0	기저귀, 우유
1	양상추, 기저귀, 맥주
2	양상추, 기저귀, 맥주, 우유
3	양상추, 맥주

주스가 삭제되고, 빈도가 높은 {양상추, 기저귀, 맥주} -> {우유} 순서대로 정렬

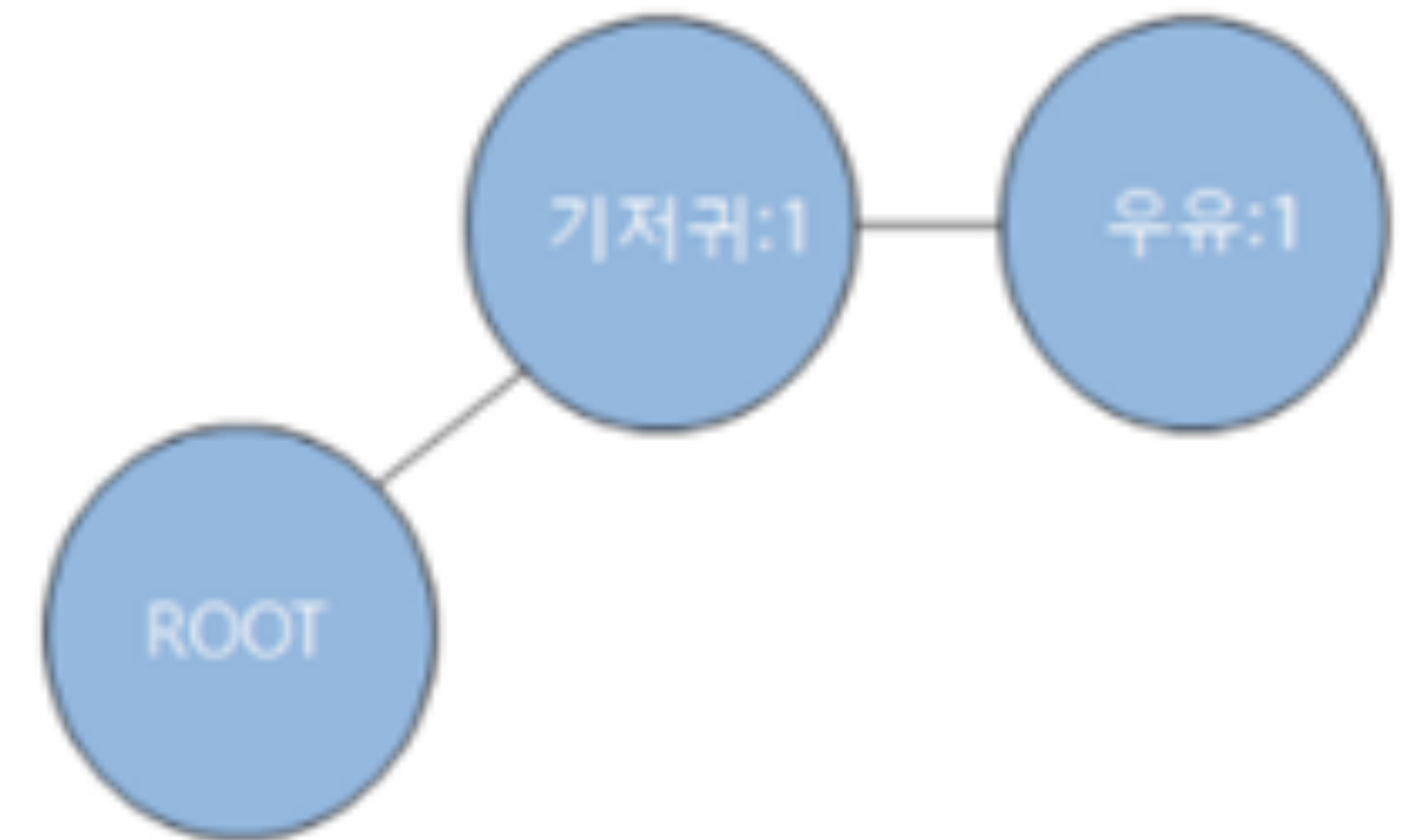


# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 알고리즘(예)

3. 부모 노드를 중심으로 거래를 자식노드로 추가해주면서 tree를 생성

거래번호	정렬된 아이템
0	기저귀, 우유
1	양상추, 기저귀, 맥주
2	양상추, 기저귀, 맥주, 우유
3	양상추, 맥주



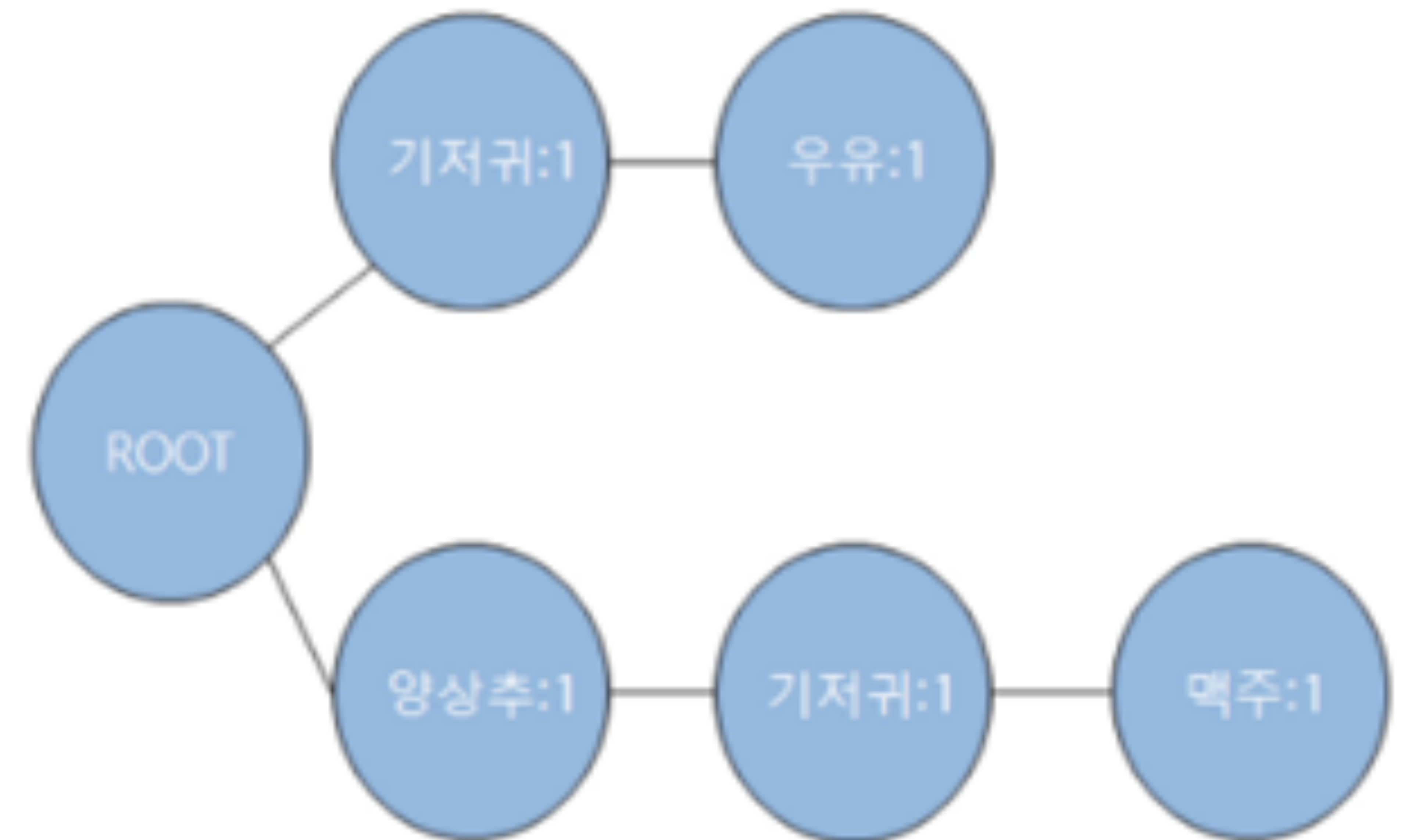


# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 알고리즘(예)

4. 새로운 아이템이 나올 경우에는 부모노드부터 시작하고, 그렇지 않으면 기존의 노드에서 확장

거래번호	정렬된 아이템
0	기저귀, 우유
1	양상추, 기저귀, 맥주
2	양상추, 기저귀, 맥주, 우유
3	양상추, 맥주



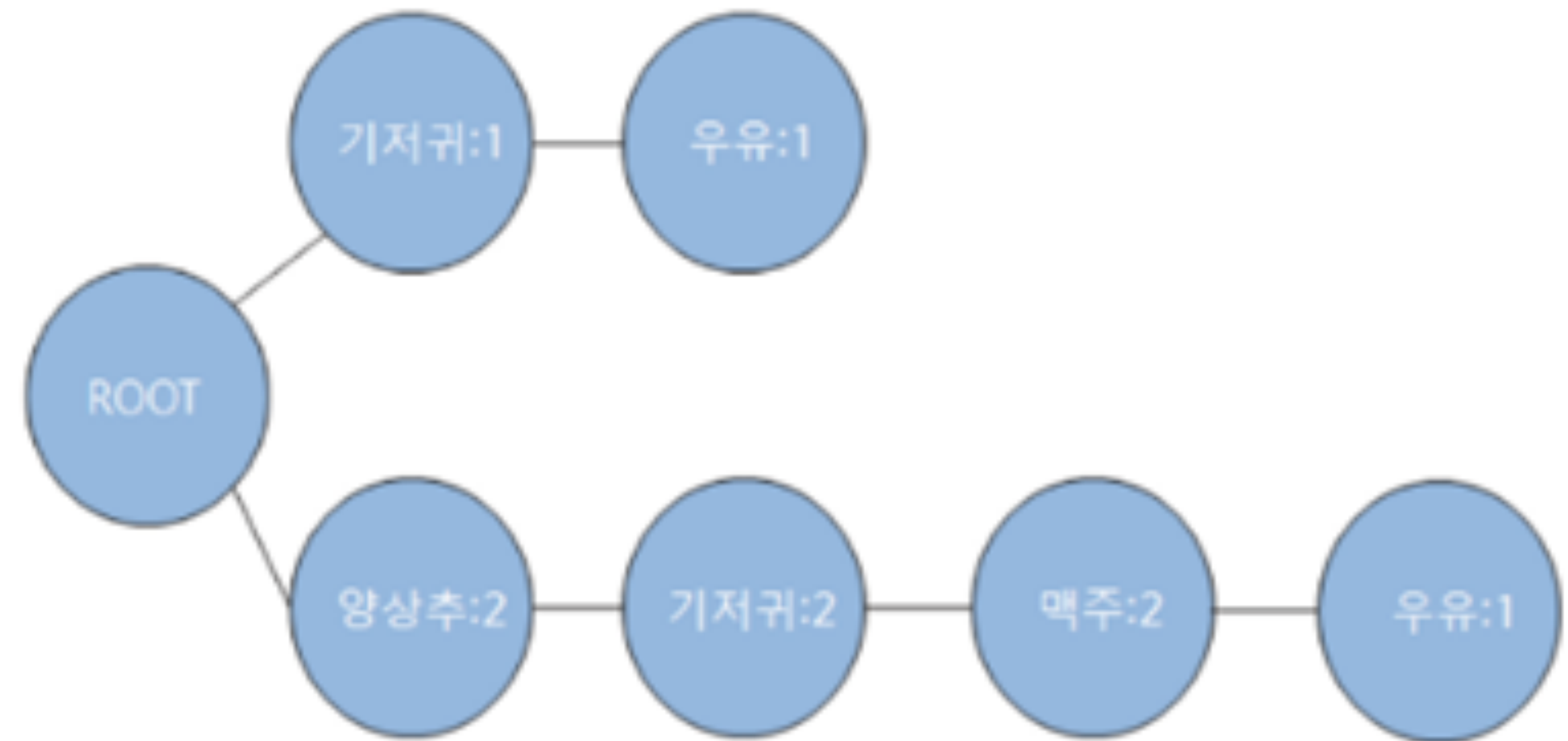


# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 알고리즘(예)

4. 새로운 아이템이 나올 경우에는 부모노드부터 시작하고, 그렇지 않으면 기존의 노드에서 확장

거래번호	정렬된 아이템
0	기저귀, 우유
1	양상추, 기저귀, 맥주
2	양상추, 기저귀, 맥주, 우유
3	양상추, 맥주



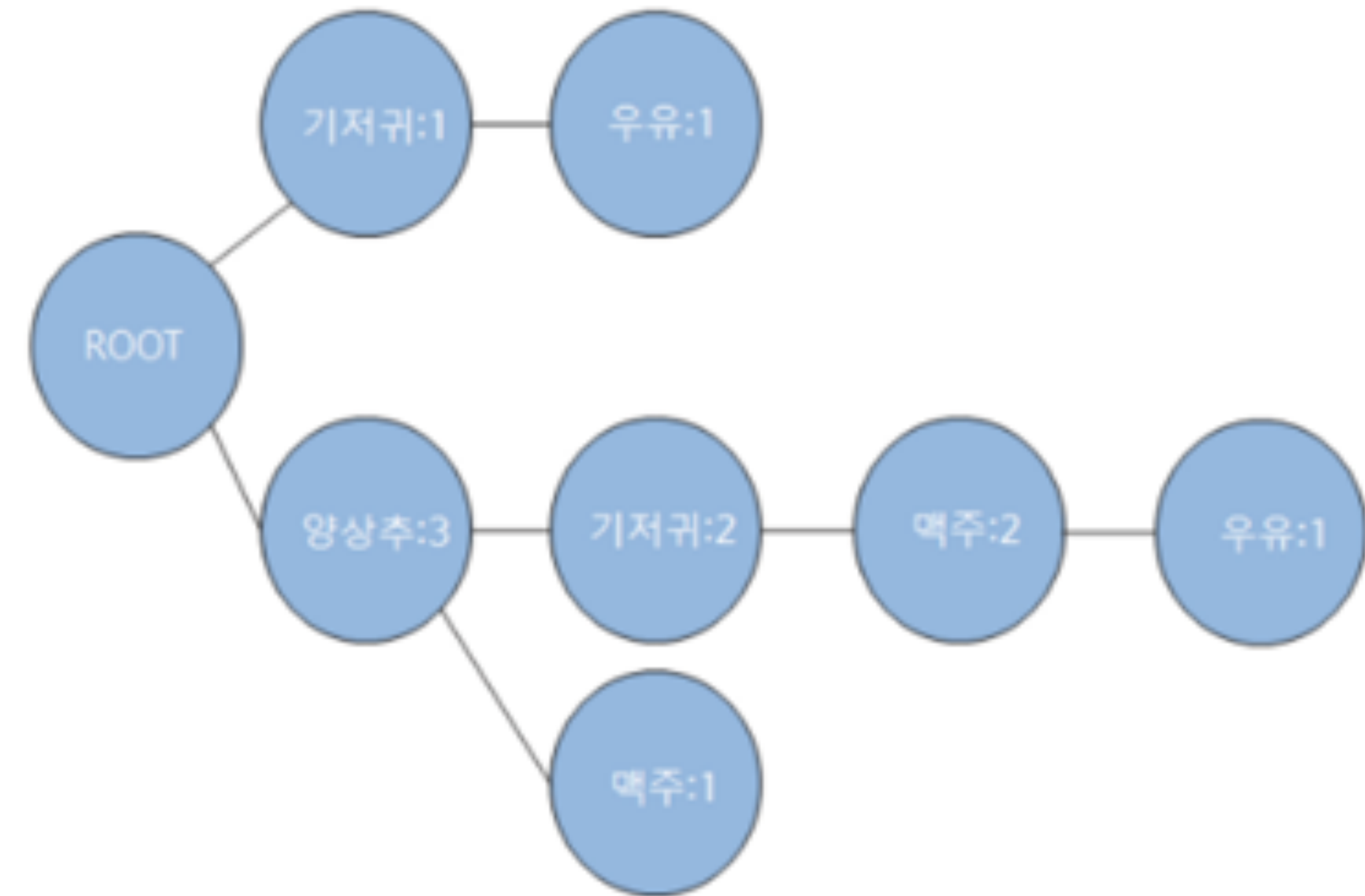


# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 알고리즘(예)

4. 새로운 아이템이 나올 경우에는 부모노드부터 시작하고, 그렇지 않으면 기존의 노드에서 확

거래번호	정렬된 아이템
0	기저귀, 우유
1	양상추, 기저귀, 맥주
2	양상추, 기저귀, 맥주, 우유
3	양상추, 맥주





# 추천시스템 구현 - 과거의 추천시스템

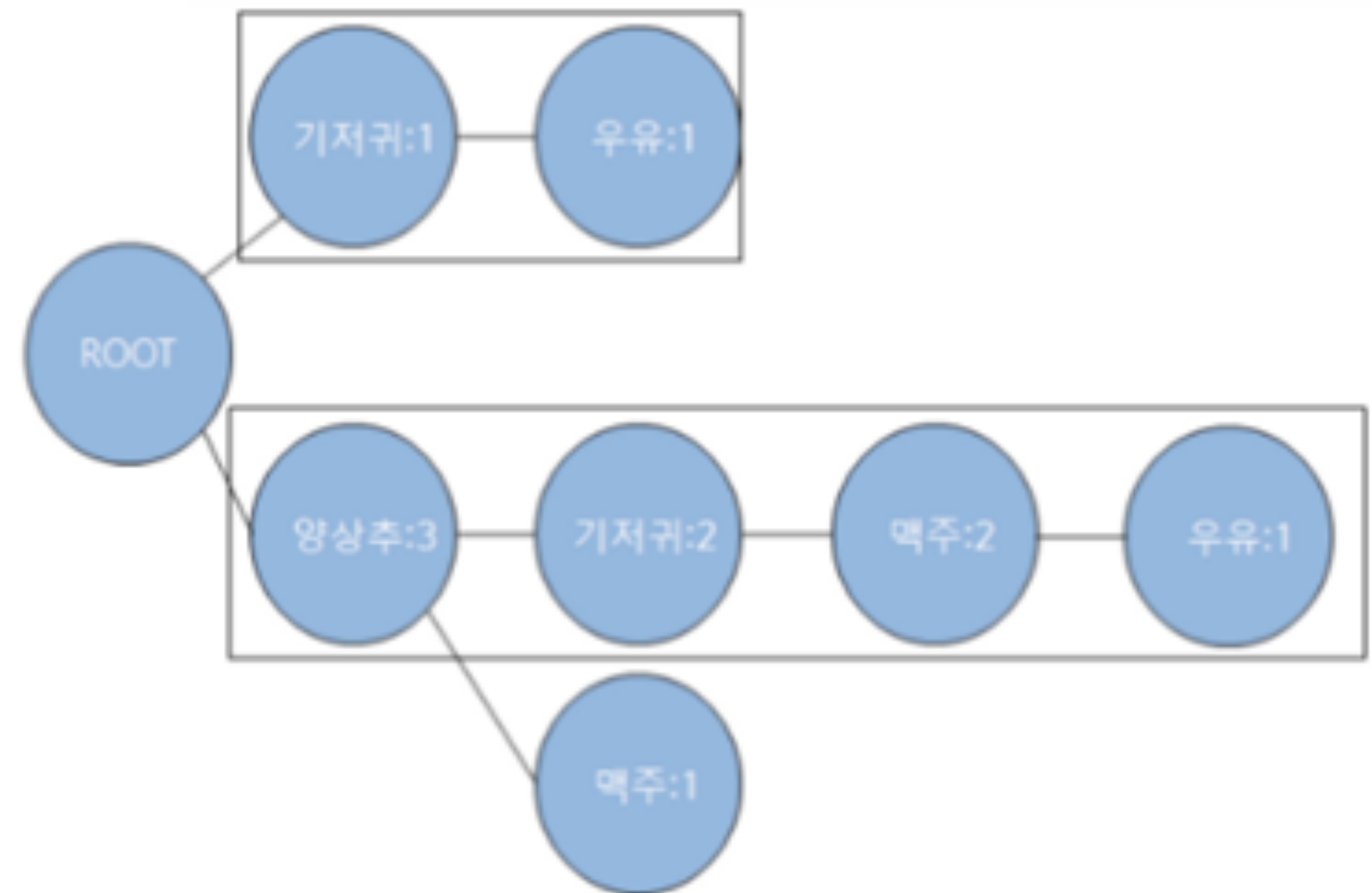
## ■ FP-Growth 알고리즘(예)

5. 지지도가 낮은 순서부터 시작하여 조건부 패턴을 생성, 모든 아이템에 대해 반복

\* 우유

- {양상추, 기저귀, 맥주} : 2
- {기저귀} : 1

아이템	지지도	Conditional Pattern bases
기저귀	0.75	
양상추	0.75	
맥주	0.75	
우유	0.5	{양상추, 기저귀, 맥주} : 2 {기저귀} : 1





# 추천시스템 구현 - 과거의 추천시스템

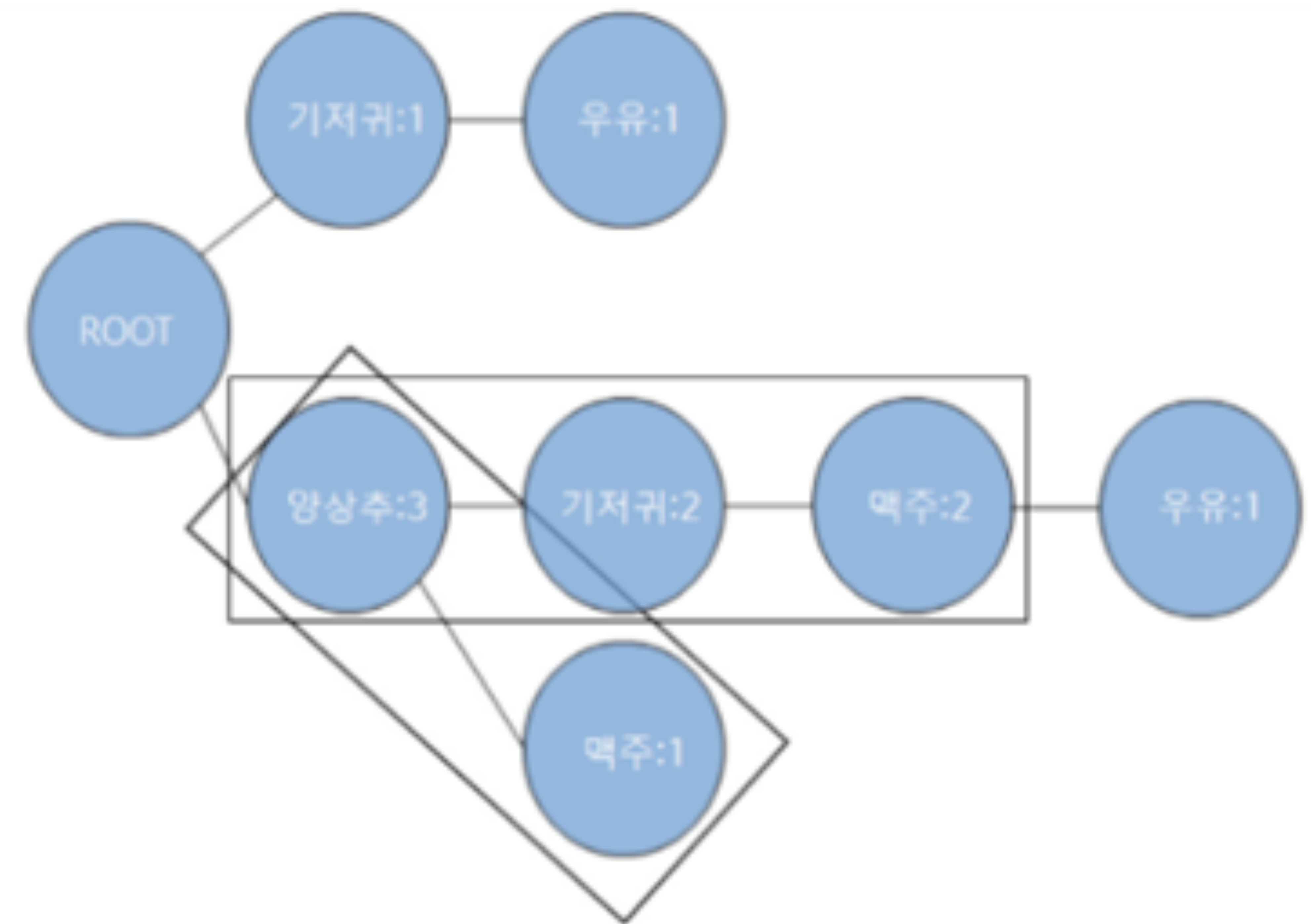
## ■ FP-Growth 알고리즘(예)

5. 지지도가 낮은 순서부터 시작하여 조건부 패턴을 생성, 모든 아이템에 대해 반복

\* 맥주

- {양상추, 기저귀} : 2
- {양상추} : 1

아이템	지지도	Conditional Pattern bases
기저귀	0.75	
양상추	0.75	
맥주	0.75	{양상추, 기저귀} : 2 {양상추} : 1
우유	0.5	{양상추, 기저귀, 맥주} : 2 {기저귀} : 1





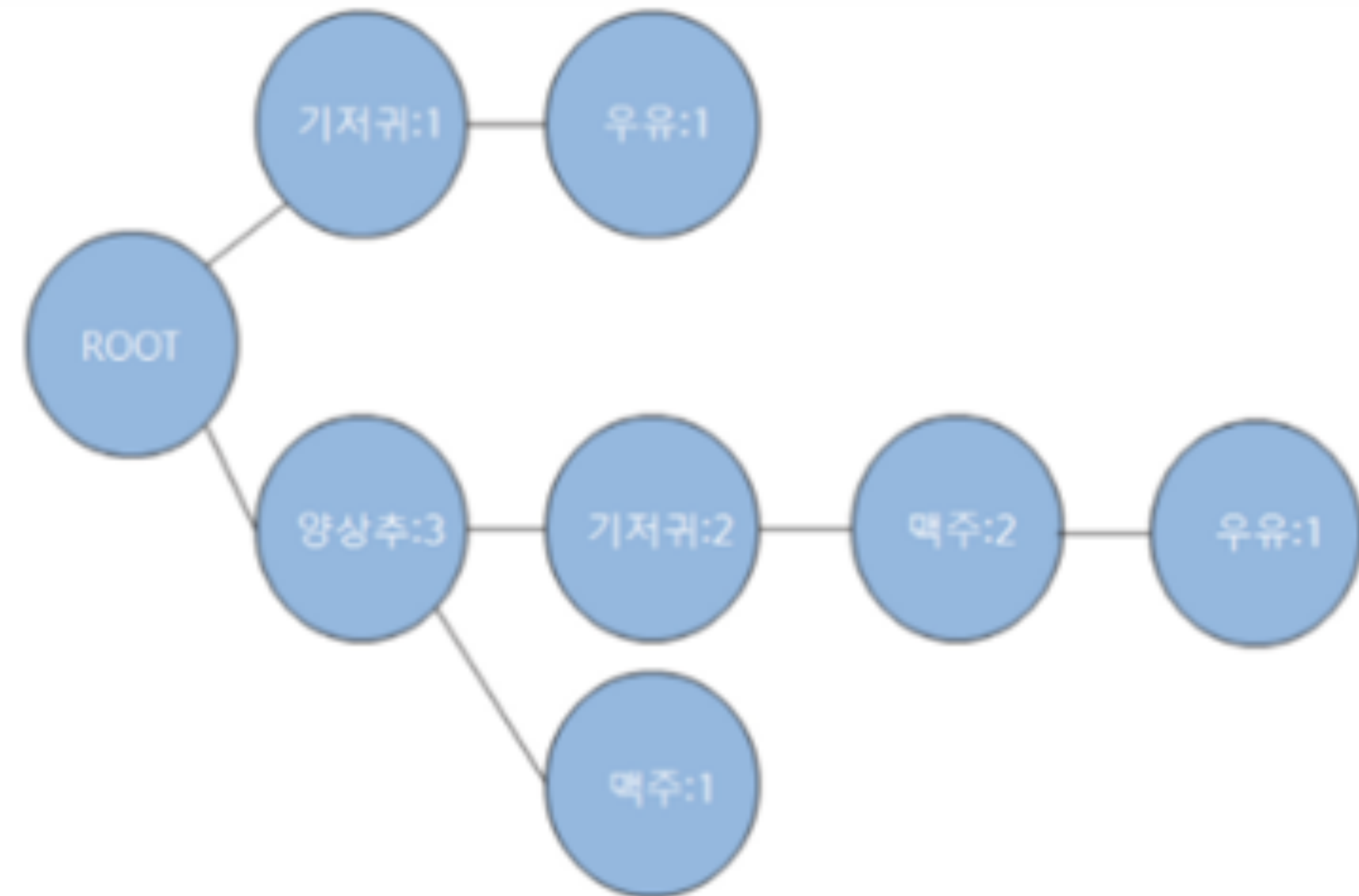
# 추천시스템 구현 - 과거의 추천시스템

## ■ FP-Growth 알고리즘(예)

5. 지지도가 낮은 순서부터 시작하여 조건부 패턴을 생성, 모든 아이템에 대해 반복

\* 양상추

아이템	지지도	Conditional Pattern bases
기저귀	0.75	
양상추	0.75	{}
맥주	0.75	{양상추, 기저귀} : 2 {양상추} : 1
우유	0.5	{양상추, 기저귀, 맥주} : 2 {기저귀} : 1





# 추천시스템 구현 - 과거의 추천시스템

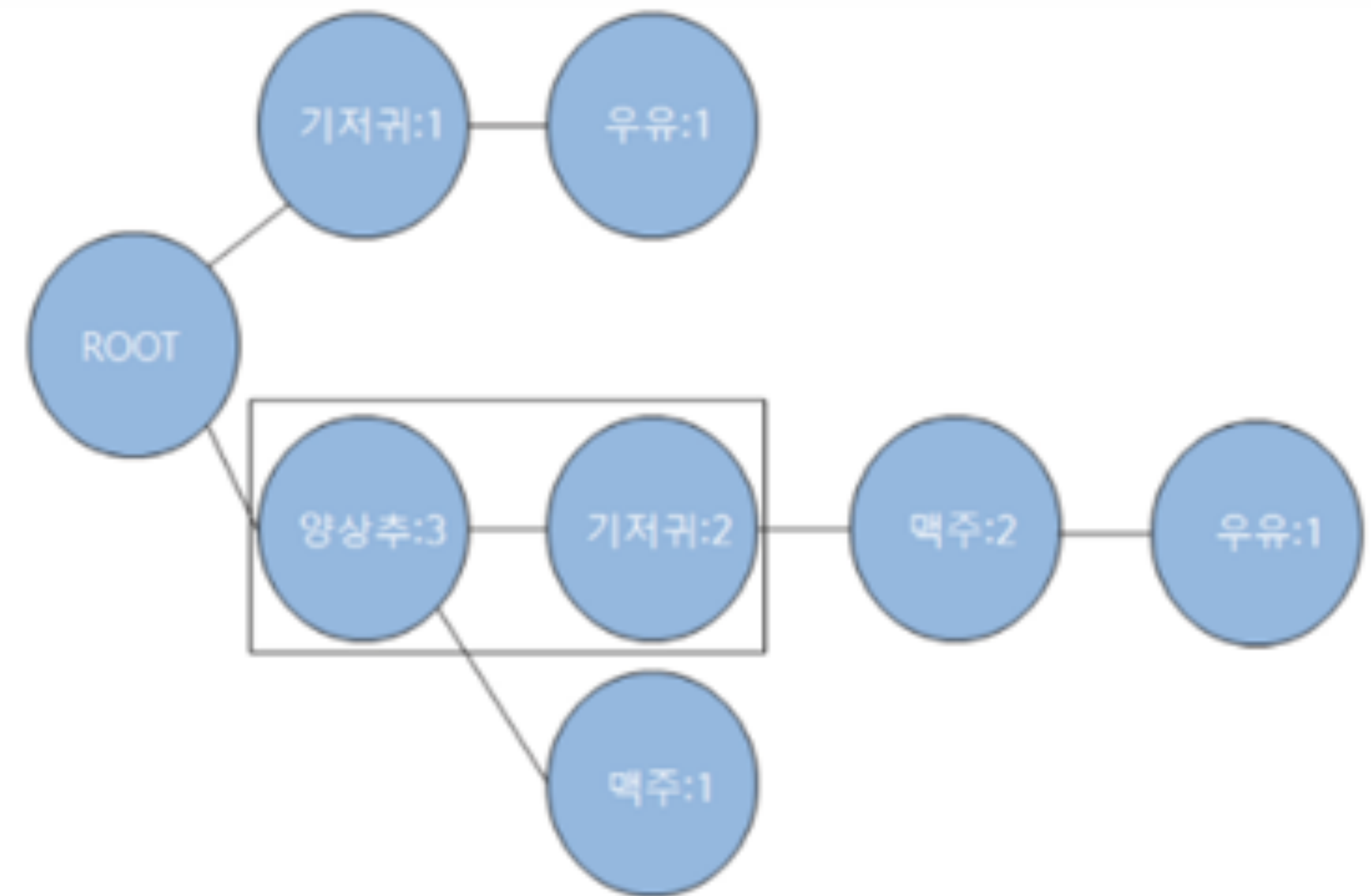
## ■ FP-Growth 알고리즘(예)

5. 지지도가 낮은 순서부터 시작하여 조건부 패턴을 생성, 모든 아이템에 대해 반복

\* 기저귀

- {양상추} : 2

아이템	지지도	Conditional Pattern bases
기저귀	0.75	{양상추} : 2
양상추	0.75	{}
맥주	0.75	{양상추, 기저귀} : 2 {양상추} : 1
우유	0.5	{양상추, 기저귀, 맥주} : 2 {기저귀} : 1





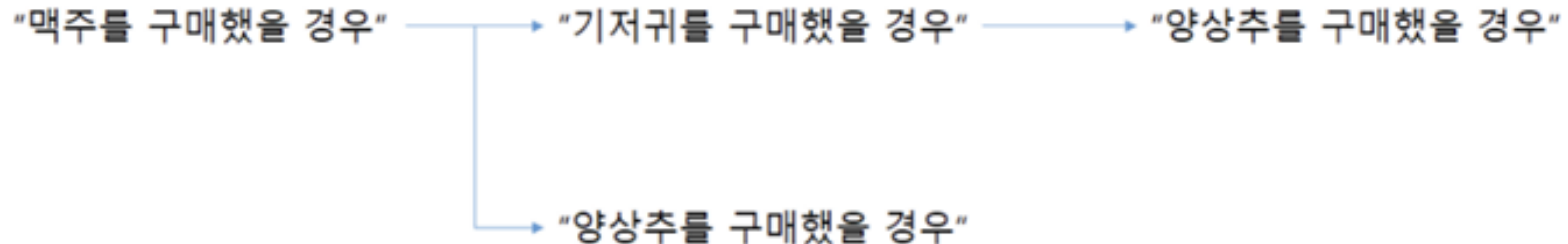
# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 알고리즘(예)

- 6. 조건부 패턴을 기반으로 패턴 생성

Conditional Pattern bases를 기반으로 패턴 생성

아이템	지지도	Conditional Pattern bases
기저귀	0.75	{양상추} : 2
양상추	0.75	{}
맥주	0.75	{양상추, 기저귀} : 2 {양상추} : 1
우유	0.5	{양상추, 기저귀, 맥주} : 2 {기저귀} : 1



# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth : FP Tree 구조를 사용, Apriori의 속도 측면의 단점을 개선한 알고리즘
  - 장점
    - Apriori 알고리즘보다 빠르고 2번의 탐색만 필요
    - 후보 item sets을 생성할 필요없이 진행 가느
  - 단점
    - 대용량의 데이터 셋에서 메모리를 효율적으로 사용하지 않음
    - Apriori 알고리즘에 비해서 설계하기 어려움
    - 지지도의 계산이 FP-Tree가 만들어지고 나서야 가능함



# 추천시스템 구현 - 과거의 추천시스템

- Apriori 코드

```
import mlxtend          #mlxtend 모듈은 데이터의 전처리부터 분류, 군집 등의 모델링 기능 제공
import numpy as np
import pandas as pd

data = np.array([
    ['우유', '기저귀', '쥬스'],
    ['양상추', '기저귀', '맥주'],
    ['우유', '양상추', '기저귀', '맥주'],
    ['양상추', '맥주']
])
```

# 추천시스템 구현 - 과거의 추천시스템

- Apriori 코드

```
from mlxtend.preprocessing import TransactionEncoder    #연관성분석을 위한 데이터셋으로 변환
te = TransactionEncoder()
te_ary = te.fit(data).transform(data)    #2차원 list or array형태를 받아서 matrix로 변환
df = pd.DataFrame(te_ary, columns=te.columns_)    #데이터프레임으로 변경
df
```

	기저귀	맥주	양상추	우유	쥬스
0	True	False	False	True	True
1	True	True	True	False	False
2	True	True	True	True	False
3	False	True	True	False	False



# 추천시스템 구현 - 과거의 추천시스템

- Apriori 코드

```
from mlxtend.frequent_patterns import apriori
frequent_itemsets = apriori(df, min_support=0.5,
                             use_colnames=True)
frequent_itemsets
```

	support	itemsets
0	0.75	(기저귀)
1	0.75	(맥주)
2	0.75	(양상추)
3	0.50	(우유)
4	0.50	(기저귀, 맥주)
5	0.50	(기저귀, 양상추)
6	0.50	(기저귀, 우유)
7	0.75	(맥주, 양상추)
8	0.50	(기저귀, 맥주, 양상추)

# 추천시스템 구현 - 과거의 추천시스템

- FP-Growth 코드

```
from mlxtend.frequent_patterns import fpgrowth
frequent_itemsets = fpgrowth(df, min_support=0.5,
                             use_colnames=True)

frequent_itemsets
```

	support	itemsets
0	0.75	(기저귀)
1	0.50	(우유)
2	0.75	(양상추)
3	0.75	(맥주)
4	0.50	(기저귀, 맥주)
5	0.50	(기저귀, 양상추)
6	0.50	(기저귀, 맥주, 양상추)
7	0.50	(기저귀, 우유)
8	0.75	(맥주, 양상추)



감/사/합/니/다