

## < Programming Assignment #1 >

See announcement in our LMS (learning.hanyang.ac.kr)

- Due Date, submission e-mail address, etc

### 1. Environment

- OS: Windows, Mac OS, or Linux
- Languages: Java or Python (any version is ok)

### 2. Goal: to find **association rules** using the **Apriori** algorithm

### 3. Requirements

The program must meet the following requirements:

- Execution file name: apriori.py (or, apriori.exe, apriori.etc...)
- Execute the program with three arguments: **minimum support (%)**, input file name, output file name

■ Example:

```
C:\#>apriori.exe 5 input.txt output.txt
```

- **Minimum support = 5%**, input file name = 'input.txt', output file name = 'output.txt'

- Input file format (.txt)

[item\_id]\t[item\_id]\n

[item\_id]\t[item\_id]\t[item\_id]\t[item\_id]\t[item\_id]\n

[item\_id]\t[item\_id]\t[item\_id]\t[item\_id]\n

■ Each line: transaction

■ item\_id is a numerical value

■ There is no duplication of items in each transaction

■ Example:

18	2	4	5	1	
1	11	15	2	7	16
2	1	16			
15	7	6	11	18	9
11	2	13	4		

- Output file format (.txt)

[item\_set]\t[associative\_item\_set]\t[support(%)]\t[confidence(%)]\n

[item\_set]\t[associative\_item\_set]\t[support(%)]\t[confidence(%)]\n

■ [item\_set]\t[associative\_item\_set]: association rules with minimum support

- [item\_set] → [associative\_item\_set]
- Use braces to represent item sets: {[item\_id],[item\_id],...} (Important!!)
  - e.g., {0}    {3,1}

- *Support*: probability that a transaction contains  $[item\_set] \cup [associative\_item\_set]$
- *Confidence*: conditional probability that a transaction having  $[item\_set]$  also contains  $[associative\_item\_set]$
- The order of output is unimportant.
- The value of support and confidence should be rounded to **two decimal places**.
  - e.g., 24.631 should become 24.63.
- Result example:

					⋮
					⋮
					⋮
{12,16}	{13}	5.20	38.81		
{13,16}	{12}	5.20	37.68		
{1}	{3,8,16}		9.40	31.54	
{3}	{1,8,16}		9.40	31.33	
{1,3}	{8,16}	9.40	87.04		
{8}	{1,3,16}		9.40	20.80	
{1,8}	{3,16}	9.40	61.04		
{3,8}	{1,16}	9.40	36.43		
{1,3,8}	{16}	9.40	97.92		
{16}	{1,3,8}	9.40	22.17		
{1,16}	{3,8}	9.40	58.02		
{3,16}	{1,8}	9.40	37.30		
{1,3,16}	{8}	9.40	97.92		
{8,16}	{1,3}	9.40	31.13		
{1,8,16}	{3}	9.40	81.03		
{3,8,16}	{1}	9.40	39.17		

- **Note: Please make sure to match the output format!**  
**If the format is not correct, you can't get any score.**

#### 4. Submission

- Please submit a single .zip file to TA's email address (noticed in our LMS)
  - Guileline
    - The file format of report must be \*.pdf. or \*.doc or \*.hwp
    - Content
      - ✓ Instructions for compiling and running your source codes on other person's computer (e.g. screenshot) (*Important!!*)
      - ✓ Java or Python version, and any other specification that our TA must know for running your code
  - Program files
    - An executable file (.exe or .py)
    - All source files
      - ✓ MakeFile if you use Linux

#### 5. Penalty

- Late submission
  - 1 week delay: 20%
  - 2 weeks delay: 50%
  - Delay more than 2 weeks: 100%
- Requirements unsatisfied
  - Penalty up to 100% will be given depending on how the requirements are well-satisfied