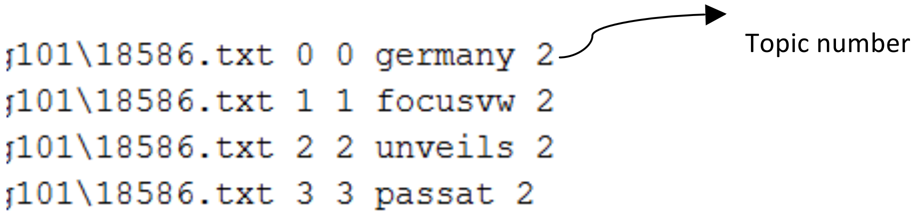
After implementing the LDA model, we obtain ".gz" file, which includes the every word assign to different topic.



```
ʝ101\18586.txt 0 0 germany 2
ʝ101\18586.txt 1 1 focusvw 2
ʝ101\18586.txt 2 2 unveils 2
ʝ101\18586.txt 3 3 passat 2
```

We use this file to create transactional databases for pattern mining.

In the Association rule mining.zip, you can see the introduction of the format of the transactional database in Using ARMSystem.txt. You can refer to the format of mushroom-bin.data and mushroom-bin.names. For each topic, we need one pair: .data and .names.

Document	Z_1	Z_2	Z_3
	words	words	words
d_1	w_1,w_2,w_3,w_2,w_1	w_1,w_9,w_8	w_7,w_{10},w_{10}
d_2	$w_2,w_{4\,,}w_4$	w_7, w_8,w_1,w_8, w_8	w_1,w_{11},w_{12}
d_3	w_2,w_1,w_7,w_5	w_7,w_1,w_3,w_2	w_4,w_7,w_{10},w_{11}
d_4	w_2,w_7,w_6	w_9,w_8,w_1	w_1,w_{11},w_{10}

Suppose we have 5 documents, d1, d2, d3, d4, d5. Words assignments are listed in the figure. Then the three transactional databases for 3 topics are like below:

Transactional datasets

transaction	topic document transaction
1	$\{w_1, w_2, w_3\}$
2	$\{w_2, w_4\}$
3	$\{w_1, w_2,w_5,w_7\}$
4	$\{w_2, w_6, w_7\}$

transaction	topic document transaction
1	$\{w_1, w_8, w_9\}$
2	$\{w_1, w_7, w_8\}$
3	$\{w_1, w_2,w_3,w_7\}$
4	$\{w_1, w_8, w_9\}$

transaction	topic document transaction
1	$\{w_7, w_{10}\}$
2	$\{w_1, w_{11}, w_{12}\}$
3	$\{w_4, w_7, w_{10}, w_{11}\}$
4	$\{w_1, w_{11}, w_{10}\}$

$\mathcal{T}_1$ $\mathcal{T}_2$ $\mathcal{T}_3$

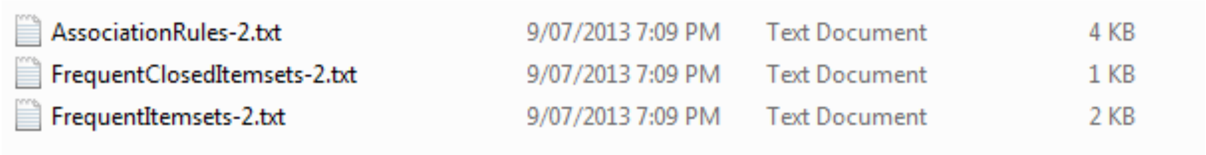
Note that when one word occurs in one document many times, we just count it for once. In each transactional database, a set of words are without any duplicates. After pattern mining, we get patterns in each topic. For example, the table below list the patterns in topic2. The threshold is 0.5.

Patterns	*supp*
$\{w_8\}$,$\{w_1$,$w_8\}$	0.75
$\{w_9\}$,$\{w_8$,$w_9\}$,$\{w_1, w_9\}$,$\{w_1, w_8, w_9\}$,$\{w_1, w_7\}$	0.5

Enter in “…\PS-System\Build”, command to call ARM is

```
java -Xms64M -Xmx1024M ARM “../[inputfilename.data]" 0.2 0.5 T F F F 1 2 2 "../../outFiles"
```

Notes: 0.2 is the minimum threshold, outFiles should be created in advance.



AssociationRules-2.txt	9/07/2013 7:09 PM	Text Document	4 KB
FrequentClosedItemsets-2.txt	9/07/2013 7:09 PM	Text Document	1 KB
FrequentItemsets-2.txt	9/07/2013 7:09 PM	Text Document	2 KB

In outFiles, frequent pattern and closed pattern are obtained for the particular topic.