

Association Rules: Exercises/Assignment

We use the R package `arules` to perform market-basket analysis using the apriori algorithm. The following commands will install and load `arules`, assuming you have permission to write to the folder `D:/`

```
install.packages("arules", lib="D:/", repos="http://cran.ms.unimelb.edu.au")
library("arules", lib="D:/")
```

If you have administrator privileges then the `lib` argument may be omitted, and R will use the default location to store the package.

Exercises

Familiarise yourself with `arules` by working through the examples from the *DATA MINING Desktop Survival Guide* by Graham Williams. You might also like to read the *Introduction to arules* by Hahsler et al. (Note that Example 5.2 in Hahsler et al uses essentially the same data as Williams' survey dataset example.)

Assignment

Load the Epub dataset described in Example 5.1 of Hahsler et al and remove the itemsets of size 50 or larger.

1. Which document is accessed most often, and what is its support?
2. What are the support, confidence and lift of the following rule
`doc_11d and doc_364 => doc_ec`
(Hint: see `?sets`)
3. Find all rules will support of at least 0.001. Of those rules with maximal confidence, which has the highest lift?