2 Hands On: Web Usage Mining

Read the file **log.csv**, containing information on the pages visited by a set of users, into a data frame in R.

2.1 Simple Recommendation Strategies

Most Visited Pages

1. Recommend the 3 most visited pages. For that purpose:

(a) inspect how many times each page was visited.

(b) sort the pages by decreasing number of visits.

(c) obtain the top 3 pages for recommendation.

Using Clustering Results

2. Suppose we want to form two clusters of users, according to the pages they have visited. For that purpose:

(a) start by transforming the log access data into a matrix that has on each row a user and for each user the information on his visits to each page; this can be obtained with the `table()` function

(b) use the function `dist()` to obtain a distance matrix with the Euclidean distance between the users;

(c) check for alternatives in the help page of `dist()`;

(d) use the function `hclust()` with the distance matrix to obtain an agglomerative clustering model of this data;

(e) visualize the obtained dendogram with function `plot()`;

(f) visualize the dendogram again, but now with option hang=-0.1.

(g) use the function `cutree()` to ”cut” the hierarchical clustering in just two clusters; inspect the cluster membership of each user;

(h) use the function `rect.hclust()` to draw the previous solution in the dendogram.

3. Recommend the top 2 pages for users of cluster 1. For that purpose:

(a) inspect what were the pages visited by users in cluster 1.

(b) inspect how many times each of these pages was visited.

(c) sort the pages by decreasing order of visits.

(d) obtain the top 2 pages for recommendation.

4. Recommend the top 2 pages for users of cluster 2.

5. Using the same clustering results, recommend the top 3 pages for user u2. From those top pages you should remove the pages that the user has already visited.