

Data set for the CWTS Leiden Ranking 2017 website

Nees Jan van Eck and Ludo Waltman

Centre for Science and Technology Studies, Leiden University, The Netherlands

{ecknjpvan, waltmanlr}@cwts.leidenuniv.nl

This document describes a data set of visits to the website of the 2017 edition of the CWTS Leiden Ranking (www.leidenranking.com/ranking/2017) in the period between May 17, 2017 and February 28, 2018. The analysis presented in our research article *Analyzing the activities of visitors of the Leiden Ranking website*¹ is based on this data set. The data set is licensed under a CC BY 4.0 license.

The data set is made available in five plain text files. The files have an UTF-8 encoding. Each file consists of a number of columns. Columns are separated from each other by a tab. Each file also includes a header line.

We first consider the file `LR2017_website.txt`. Each line in this file represents an action taken by a visitor of the website. We refer to an action as a view. Each view has a unique identifier. Furthermore, each view is part of a session, and each session also has a unique identifier. Within a session, views are ordered by their identifiers. The view with the lowest identifier represents the first action taken within a session, the view with the second lowest identifier represents the second action taken within a session, and so on. In addition to view and session identifiers, each line in the file `LR2017_website.txt` also contains the page on the Leiden Ranking website that was visited (i.e., list view page, chart view page, map view page, or university page), the date at which the visit took place, and the country from which the visit took place, as determined by the visitor's IP address.

We now consider the other four files. The files `LR2017_website_list.txt`, `LR2017_website_chart.txt`, `LR2017_website_map.txt`, and `LR2017_website_university.txt` provide data on views in which,

¹ Van Eck, N.J., & Waltman, L. (2018). *Analyzing the activities of visitors of the Leiden Ranking website*. arXiv:1804.03869.

respectively, the list view page, the chart view page, the map view page, and the university page was visited. Each line in these files corresponds with a view, and each view has an identifier that corresponds with an identifier of a view in the file `LR2017_website.txt`. In addition to an identifier of a view, each line in the four files mentioned above also contains the settings that were used in that view, for instance the time period, the field, and the indicators that were selected. In the files `LR2017_website_list.txt` and `LR2017_website_chart.txt`, the settings also include the continent or country that was selected. If no continent or country was selected in a view, the continent or country column is empty. In the file `LR2017_website_university.txt`, each line also contains the name and the country of the university that was selected.