which means "surround with a rampart or other fortification" **circumvallate**

Since 2009 Google has continuously scanned Books and published this data on the Google Books Ngram Corpus. The Dataset is based on Books from around the 1500s up to 2019 and consists of Ngrams that appear at least 40 times across the corpus. These Data is stored as plain text in the following format. The ngram Word combination followed by the Year, in this case 1978, the 335 occurrences of this ngram in that year overall and lastly the 91 distinct books in which the ngram was found.

The resulting time series can be of use for historians or linguistics. This Plot of the Relative frequency of the 1-gram Revolution in relation to famous historic events may be a concrete use case for the Google Books Ngram Corpus Data.

If we take a closer look into the Dataset as a whole, then it becomes clear that larger n-grams dominate the space requirements for every language provided by Google. Englisch works are clearly overrepresented but German and French 5-grams have also reached Terabytes of storage needs.

The Corpus has unique properties such as space requirements growing in relation to n and data based on natural language. Therefore, with existing dependencies between different ngram time series based on grammatical rules.

Time Series compression