

Overview of an ESTC query 1700-1799 garden

Mikko Tolonen, Leo Lahti, Eetu Makela et al.

2016-10-07

Technical notes

Query IDs (n=41309) from file: garden.txt

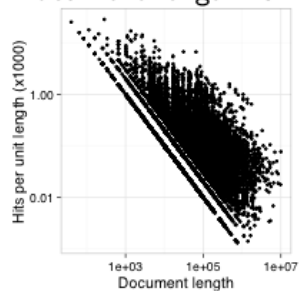
IDs found from complete ESTC: n=41294 (99.96 %)

IDs found after the subset: n=40139 (97.17 %)

Duplicated IDs in ESTC: n = 36 (0%). Very few, each occurrence is considered separately.

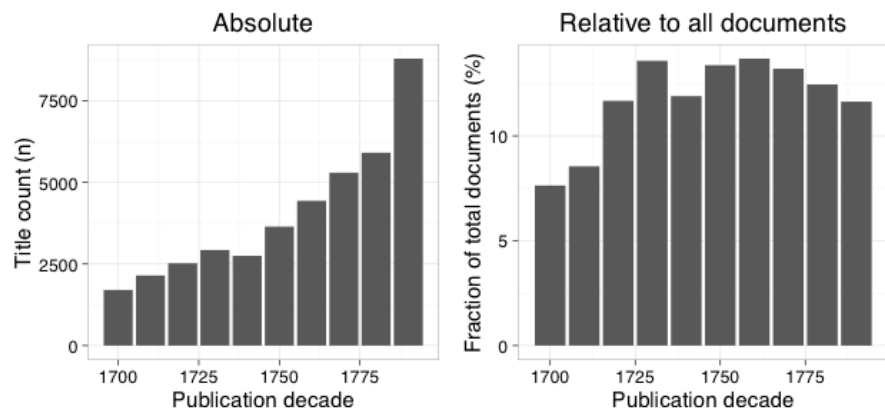
Duplicated IDs in the query: 10159 (24.6%). According to a single check, these are often series with same ID over many issues. The total sum of query hits and lengths is used.

Document length vs. normalized hit frequency:

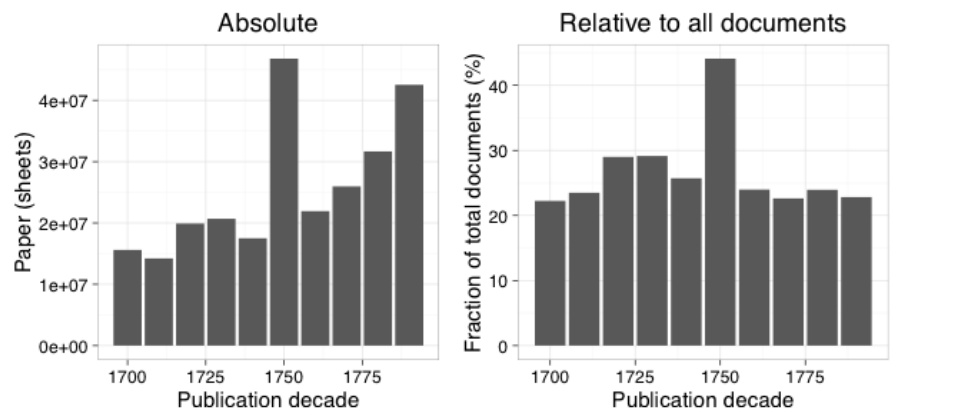


Navigation icons: back, forward, search, and other controls.

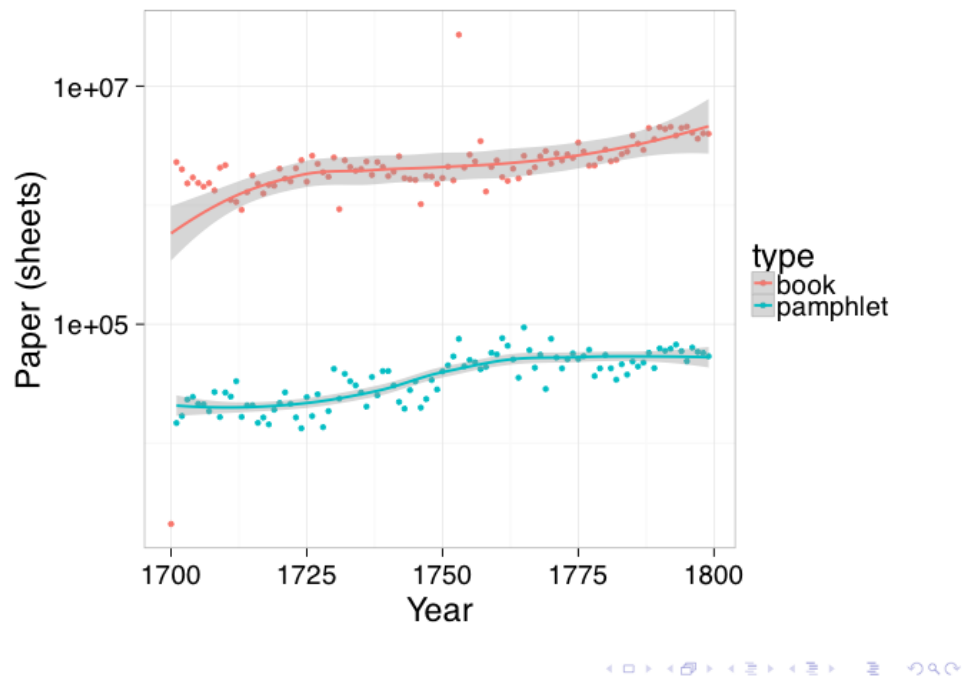
Timeline for title count



Timeline for paper consumption

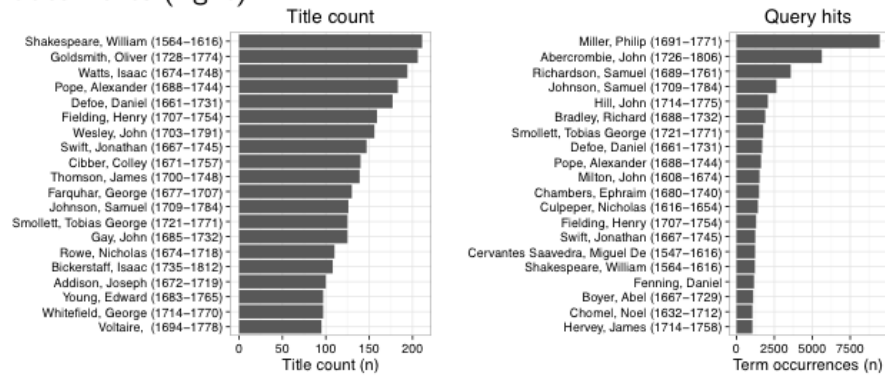


Paper consumption: books vs. pamphlets

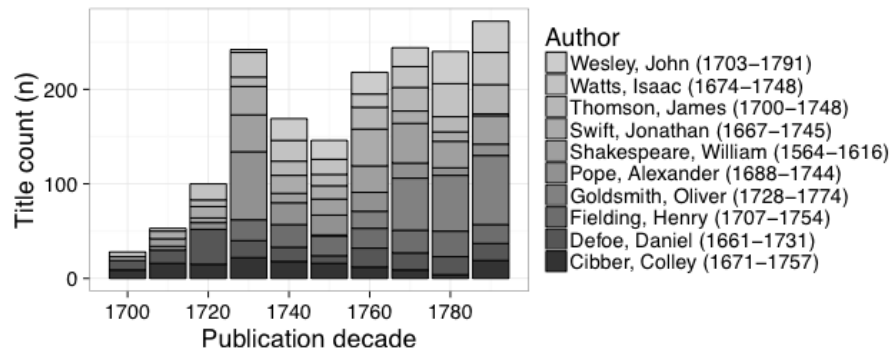


Top authors

The top authors can be defined by title count (regardless of number of hits in each title; left) or by the total sum of query hits across all documents (right).

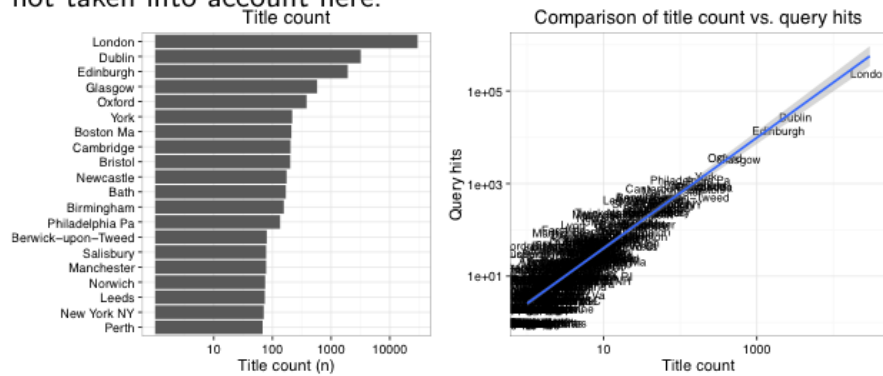


Title count timeline for top 10 authors



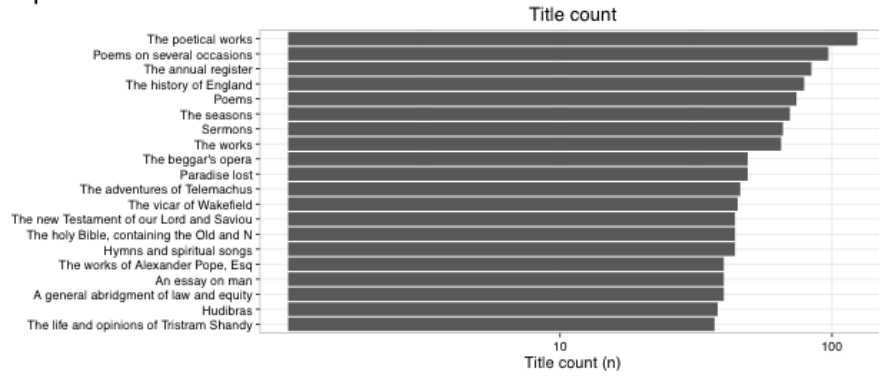
Top places

The top places defined by title count (left); Relation between titlecount and query hits (right). Note that document lengths are not taken into account here.



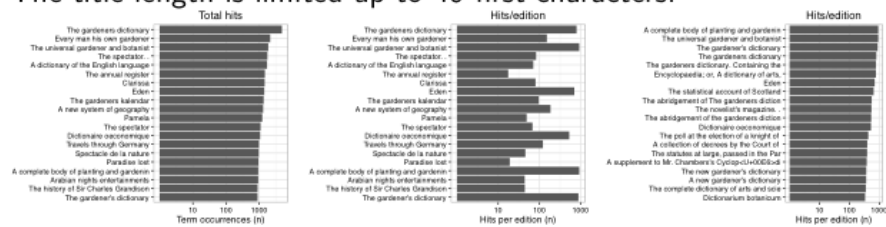
Top titles by title count

Up to 40 first characters for each title are shown.



Top hits/title

The title length is limited up to 40 first characters:



Top publishers

The top publishers can be defined by title count (regardless of number of hits in each title; left) or by the total sum of query hits across all documents (right).

