DISCUSSION PUBLISHERS
Research Center · Open Access Edition

Text Algorithms in Economics

Elliott Ash, Stephen E. Hansen

Discussion Research ID	DPR-2026-1113
Year	2023
Field	Social Sciences
DOI	10.1146/annurev-economics-082222-074352
Author ORCID	0000-0002-6817-7529
Licence	CC-BY
Original source	Annual Review of Economics
Permanent link	https://discussionpublishers.com/research/2d7d8d27-36d6-4d3f-9e7a-09dbe08caadc
Publisher	Discussion Publishers discussionpublishers.com

Republished by Discussion Publishers under the original CC-BY licence. Copyright remains with the original authors; please cite the original work.

Cite as: Elliott Ash, Stephen E. Hansen (2023). Text Algorithms in Economics. Discussion Publishers, DPR-2026-1113.

Abstract

This article provides an overview of the methods used for algorithmic text analysis in economics, with a focus on three key contributions. First, we introduce methods for representing documents as high-dimensional count vectors over vocabulary terms, for representing words as vectors, and for representing word sequences as embedding vectors. Second, we define four core empirical tasks that encompass most text-as-data research in economics and enumerate the various approaches that have been taken so far to accomplish these tasks. Finally, we flag limitations in the current literature, with a focus on the challenge of validating algorithmic output.

Keywords: text-as-data, document representation, word embeddings

Full text

The complete text of this work is hosted by its original publisher.
Read it via the Discussion Publishers page:
https://discussionpublishers.com/research/2d7d8d27-36d6-4d3f-9e7a-09dbe08caadc

Original record: https://doi.org/10.1146/annurev-economics-082222-074352
DOI: https://doi.org/10.1146/annurev-economics-082222-074352