PAPER DIGEST

CERTIFICATE OF RECOGNITION

MOST INFLUENTIAL EMNLP 2021 PAPER

This is to certify that the paper

Documenting Large Webtext Corpora: A Case Study on The Colossal Clean Crawled Corpus

Jesse Dodge, Maarten Sap, Ana Marasovic, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, Matt Gardner

presented at the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021

is recognized among the Most Influential EMNLP 2021 Papers





Paper Digest
New York, New York

Edition 2026-03
Issued 2026-03 - Impact Factor 7

Scan to verify



Verify this certificate at paperdigest.org

Certificate ID 596610525d5251cf - Ranking constructed from citations in research papers and granted patents.