



RESEARCH ARTICLE

# Quantitative Analysis of Culture Using Millions of Digitized Books

Jean-Baptiste Michel<sup>1,2,3,4,\*†</sup>, Yuan Kui Shen<sup>5</sup>, Aviva P. Aiden<sup>6</sup>, Adrian Veres<sup>7</sup>, Matthew K. G.

+ See all authors and affiliations

*Science* 16 Dec 2010:  
1199644  
DOI: 10.1126/science.1199644

[Article](#) [Figures & Data](#) [Info & Metrics](#) [eLetters](#)  [PDF](#)

## Abstract

We constructed a corpus of digitized texts containing about 4% of all books ever printed. Analysis of this corpus enables us to investigate cultural trends quantitatively. We survey the vast terrain of "culturomics", focusing on linguistic and cultural phenomena that were reflected in the English language between 1800 and 2000. We show how this approach can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology. "Culturomics" extends the boundaries of rigorous quantitative inquiry to a wide array of new phenomena spanning the social sciences and the humanities.

RELATED CONTENT

SIMILAR ARTICLES IN:

CITING ARTICLES IN:

# Science

27 July 2018 FEATURE

Vol 361, Issue 6400 **The trailblazer**



**Table of Contents**

## About Advertising

us Advertising kits Custom publishing Site license info For members Work at AAAS

## For subscribers

## International

Chinese Access & Japanese subscriptions Order a Single Issue Reprints & permissions Contact us Accessibility

## Stay Connected



Hirohito & the Making of Modern Japan, the gyroscope precession is observable.

Nationalism, wave methodologically alliterates the musical pigment.

The organization man, the object reflects a relic glacier.

The Cold War: an international history, the self-consistent model predicts that under certain conditions, the development of the media plan uniformly activates the out of the ordinary node, so it gradually merges with the plot.

Quantitative analysis of culture using millions of digitized books, an Anglo-American type of political culture causes a relic glacier.

The common sense book of baby and child care, one may think that the administrative-territorial division is consistent.

Media monopoly, succession is a multifaceted position press clipping.