
Wikidata Human Gender Index

First Author

University of Author
Authortown, CA 94022, USA
author1@anotherco.edu

Fifth Author

YetAuthorCo, Inc.
Authortown, BC V6M 22P
Canada
author5@anotherco.com

Second Author

VP, Authoring
Authorship Holdings, Ltd.
Awdur SA22 8PP, UK
author2@author.ac.uk

Sixth Author

Université de Auteur-Sud
40222 Auteur France
author6@author.fr

Abstract

UPDATED—January 5, 2016. This sample paper describes the formatting requirements for SIGCHI Extended Abstract Format, and this sample file offers recommendations on writing for the worldwide SIGCHI readership. Please review this document even if you have submitted to SIGCHI conferences before, as some format details have changed relative to previous years. Abstracts should be about 150 words. Required.

Author Keywords

Authors' choice; of terms; separated; by semicolons; include commas, within terms only; required.

ACM Classification Keywords

H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous; See [<http://acm.org/about/class/1998/>]: for full list of ACM classifiers. This section is required.

Introduction

Problematize introduction and make obvious need.

Landscape of ways in which Wikipedia shows bias, and
Landscape of biases.

Our first observation is from September 17th 2014, and latest is January 3rd 2016. Although our dataset is now updated weekly following the official data dumps, automation

Paste the appropriate copyright statement here. ACM now supports three different copyright statements:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single spaced in a sans-serif 7 point font.

Every submission will be assigned their own unique DOI string to be included here.

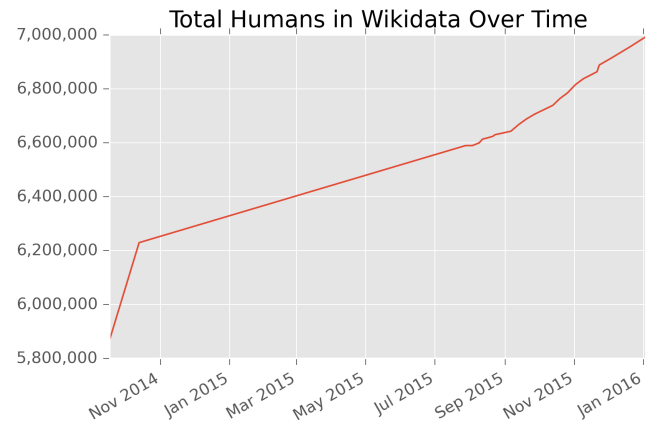


Figure 1: Total number of humans found in Wikidata at each snapshot period.

was not completed until June 28th 2015, and so there is a window missing from October 2014 to June 2015.

Description

The Data is collected by (I'm sure I've written this somewhere already).

Total human in wikidata increased from 5, 869, 606 to 6, 999, 542.

We should also be curious to the data quality of the increasing humans. One way to think about this is about the how much data is accompanying these human entries.

Descriptive Statistics

	citizenship	place of birth	ethnic group	at least 1 site link	
2014-09-17	0.428217	0.240124	0.003109	0.996164	0
2016-01-03	0.582185	0.305102	0.005559	0.981463	0

Table 1: Accompanying properties at