

Twitter data serialization methods and benchmarks

Kia Teymourian
Boston University
kiat@bu.edu

Saeed Fathollahzadeh
Free University of
Bozen-Bolzano
s.fathollahzadeh@gmail.com

Chris Jermaine
Rice University
cmj4@rice.edu

Table 1: Frequency of some Tweet Objects (for 1 million tweets)

Object Name	Parent Object	Frequency	AVG Length(byte)
tweet	root object	1,000,000	6599.305602
users	tweet	1,000,000	1453.798722
coordinates	tweet	1586	57.87705
place	tweet	11974	392.8976115
quoted status	tweet	177537	3993.947594
retweeted status	tweet	598517	4775.996797
entities	tweet	1,000,000	336.02128
extended entities	tweet	51479	1133.00505
hashtags	entities	398720	44.0340615
media	entities	51481	762.556516
urls	entities	417176	187.6829563
user mentions	entities	1063216	126.845482
symbols	entities	7793	36.55832
sizes	media	7793	36.55832
media sizes	sizes	51485	194.91743
thumb	media sizes	51485	37.989
large	media sizes	51485	37.524
medium	media sizes	51485	37.412
small	media sizes	51485	36.993

ABSTRACT

abstract

PVLDB Reference Format:

Ben Trovato, G. K. M. Tobin, Lars Thørväld, Lawrence P. Leipuner, Sean Fogarty, Charles Palmer, John Smith, Julius P. Kumquat, and Ahmet Sacan. Twitter data serialization methods and benchmarks. *PVLDB*, 12(xxx): xxxx-yyyy, 2019.
DOI: <https://doi.org/10.14778/xxxxxxx.xxxxxxx>

1. INTRODUCTION

Introduction

2. OVERVIEW OF DATASET

2.1 References

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.

Proceedings of the VLDB Endowment, Vol. 12, No. xxx
ISSN 2150-8097.

DOI: <https://doi.org/10.14778/xxxxxxx.xxxxxxx>

Generated by bibtex from your .bib file. Run latex, then bibtex, then latex twice (to resolve references).

APPENDIX

You can use an appendix for optional proofs or details of your evaluation which are not absolutely necessary to the core understanding of your paper.

A. FINAL THOUGHTS ON GOOD LAYOUT

Please use readable font sizes in the figures and graphs. Avoid tempering with the correct border values, and the spacing (and format) of both text and captions of the PVLDB format (e.g. captions are bold).

At the end, please check for an overall pleasant layout, e.g. by ensuring a readable and logical positioning of any floating figures and tables. Please also check for any line overflows, which are only allowed in extraordinary circumstances (such as wide formulas or URLs where a line wrap would be counterintuitive).

Use the **balance** package together with a **\balance** command at the end of your document to ensure that the last page has balanced (i.e. same length) columns.

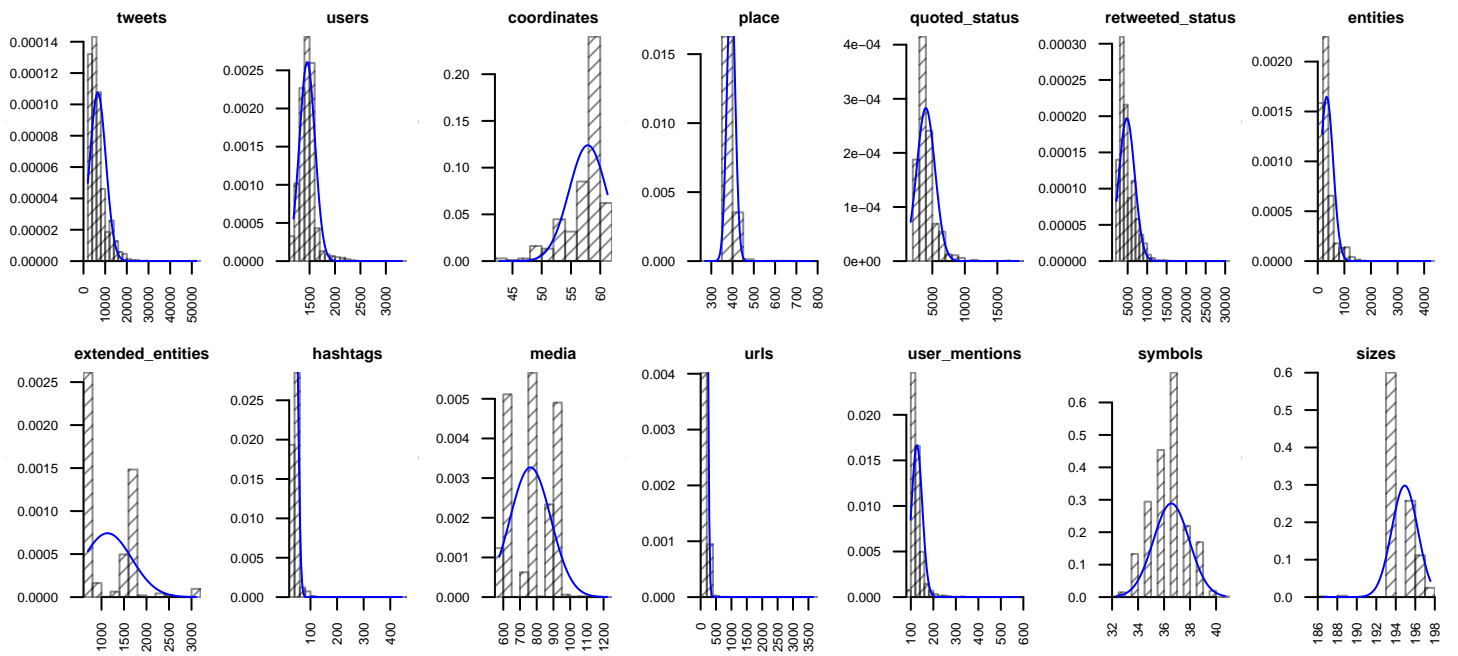


Figure 1: histograms with Standard Deviation