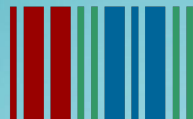


“Disputed by”

Analyzing disputes and collaboration in Wikidata

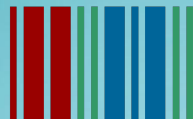
Diego Sáez Trumper
Senior Research Scientist
WMF

Lydia Pintscher
PM Wikidata
WMDE



WIKIDATA
DATA QUALITY
DAYS 2022

Statement disputed by (P1310)



WIKIDATA
DATA QUALITY
DAYS 2022

Jesus (Q302)

Jewish preacher and religious leader, central figure of Christianity

Jesus of Nazareth | Christ | Yehoshua | The Messiah | God the Son | Son of God | Jesus Christ of Nazareth | Jesus Christ | Yeshua ben-Yosef

Statements

instance of	 deity	
	statement is subject of	Christology
	statement disputed by	Islam
		atheism
		Judaism
statement supported by	Christianity	
	Messianic Judaism	
	Manichaeism	
	▶ 1 reference	

Few claims contain “statement disputed by” qualifier

Just 1886 claims
containing
“disputed by”
(2021-10)

prop	label	count
P17	"country"	607
P3355	"negative therapeutic predictor"	184
P3354	"positive therapeutic predictor"	139
P131	"located in the administrative territorial ent..."	128
P31	"instance of"	102
P460	"said to be the same as"	52
P6216	"copyright status"	51
P39	"position held"	31
P170	"creator"	30
P3359	"negative prognostic predictor"	29
P569	"date of birth"	23
P40	"child"	22
P570	"date of death"	18
P50	"author"	17
P361	"part of"	16
P1196	"manner of death"	15
P509	"cause of death"	14
P22	"father"	13
P106	"occupation"	11
P1376	"capital of"	11

Statement Dispute by - Topics

Geography.Regions	1729
Culture.Biography	283
Geography.Geographical	208
Culture.Media	204
STEM.STEM*	139

	ratio
base_topic	
STEM.Earth_and_environment	0.000339
History_and_Society.Society	0.000281
History_and_Society.History	0.000279
History_and_Society.Politics_and_government	0.000190
Culture.Philosophy_and_religion	0.000186

Ceuta (Q5823)

autonomous city of Spain in North Africa


country	 Spain	
	start time	1640
	statement disputed by	Morocco
	▶ 1 reference	

KRAS G13D (Q28371015)

genetic variant

GLY13ASP | RS112445441

 edit

negative therapeutic predictor	 cetuximab	 edit
	medical condition treated	colorectal cancer
	statement disputed by	Cetuximab treatment for metastatic colorectal cancer with KRAS p.G13D mutations improves progression-free survival
	determination method	CIVIC evidence level B
	rating	CIVIC 2-star trust rating
▶ 1 reference		

Barack Obama (Q76)

44th president of the United States from 2009 to 2017

Barack Hussein Obama II | Barack Obama II | Barack Hussein Obama | Obama | Barak Obama | Barry Obama | President Obama | President Barack Obama | BHO | Barack | Barack H. Obama

country of citizenship



United States of America

statement disputed by

Barack Obama citizenship conspiracy theories

▶ 4 references

Earth (Q2)

third planet from the Sun in the Solar System

Planet Earth | the Earth |  |  | World

creator



God in Christianity

statement **disputed** by

atheism

statement supported by

creationism
Christianity

▶ 1 reference



no value

statement **disputed** by

creationism

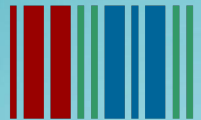
statement supported by

atheism

▼ 0 references

Reverts

Is content related with reverts?



WIKIDATA
DATA QUALITY
DAYS 2022

Reverts Predictions (wikidata-ui tag)

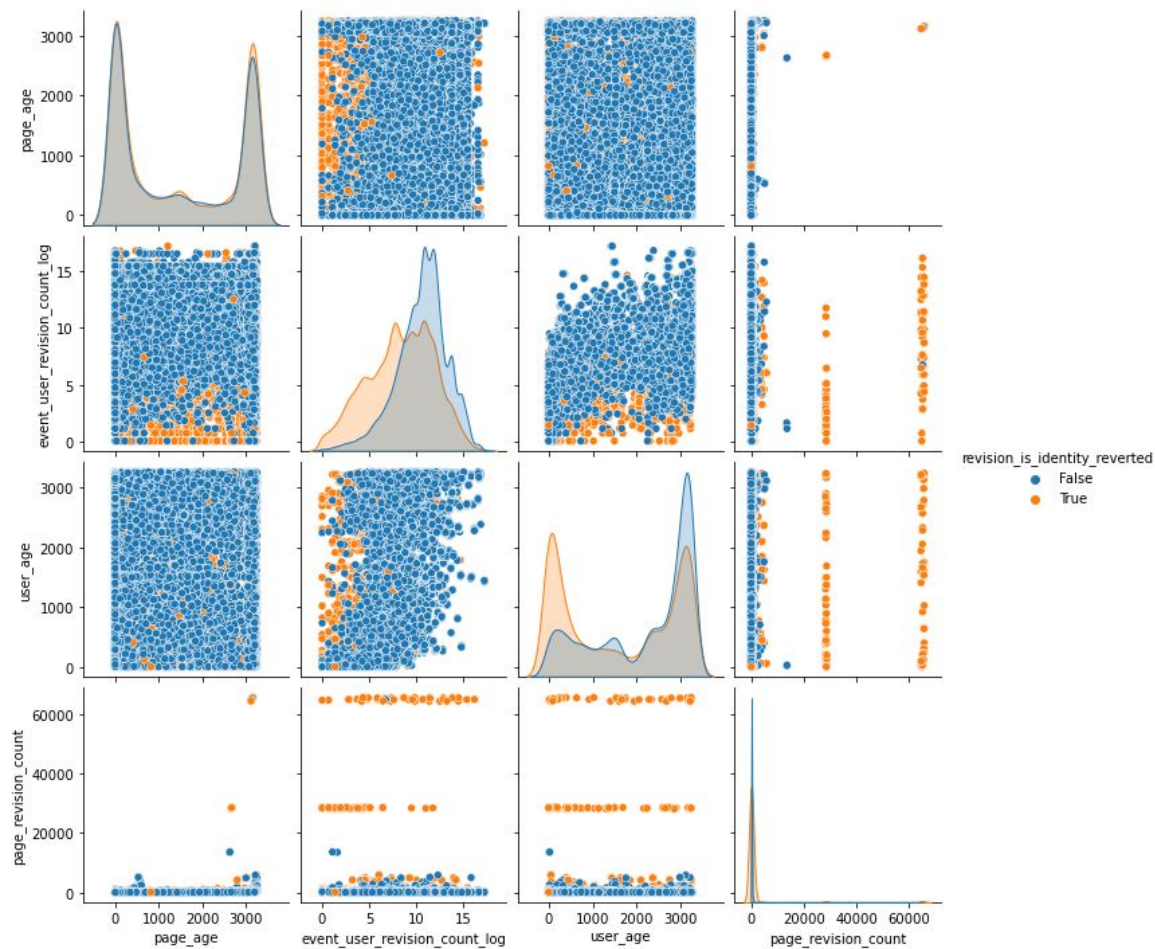
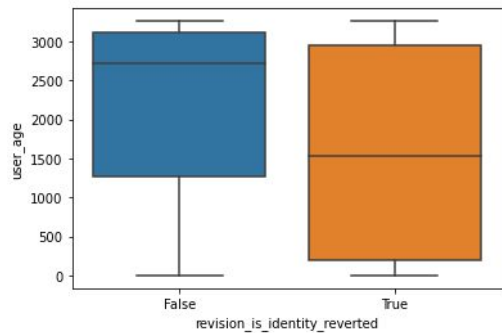
- Random Forest precision considering user & content features: 75.8% (on balanced dataset)
- RF with the top 3 features gets 70 %

```
[(0.00702043530179124, 'summaryType_wbsetclaim-create'),  
(0.00910458528202965, 'summaryType_wbremoveclaims-remove'),  
(0.009445046928611582, 'summaryType_wbsetaliases-add'),  
(0.014901230723854752, 'summaryType_wbsetdescription-add'),  
(0.01788767139909128, 'summaryType_'),  
(0.022476704328922383, 'summaryType_wbsetlabel-add'),  
(0.023233843999529716, 'summaryType_wbsetsitelink-remove'),  
(0.19391786777850195, 'page_age'),  
(0.2232950172768442, 'page_revision_count'),  
(0.27681151582195623, 'user_age')]
```

Features	Importance
Account age	0.28
Page revision count	0.22
Page age	0.19

```
[[3809 1051]  
 [1143 3077]]  
0.7583700440528635
```

Non-content features and their relation with reverted images



Co-editing items

Do people co-edit items?

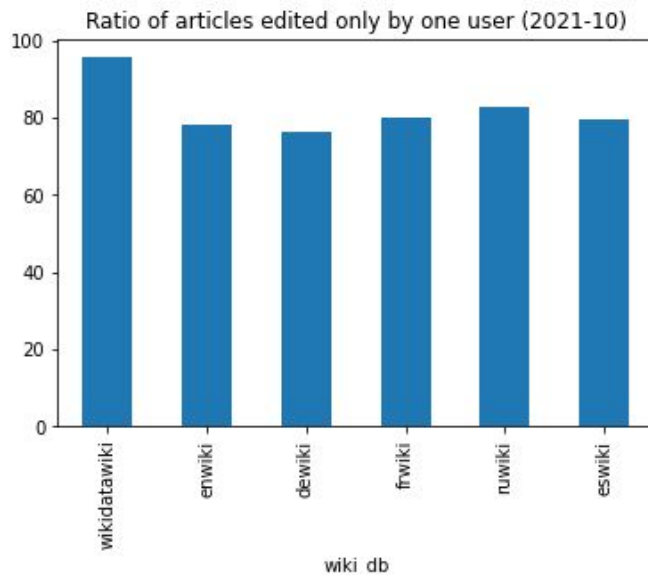


WIKIDATA

DATA QUALITY
DAYS 2022

Co-edited items in 2021-10 (wikidata-ui)

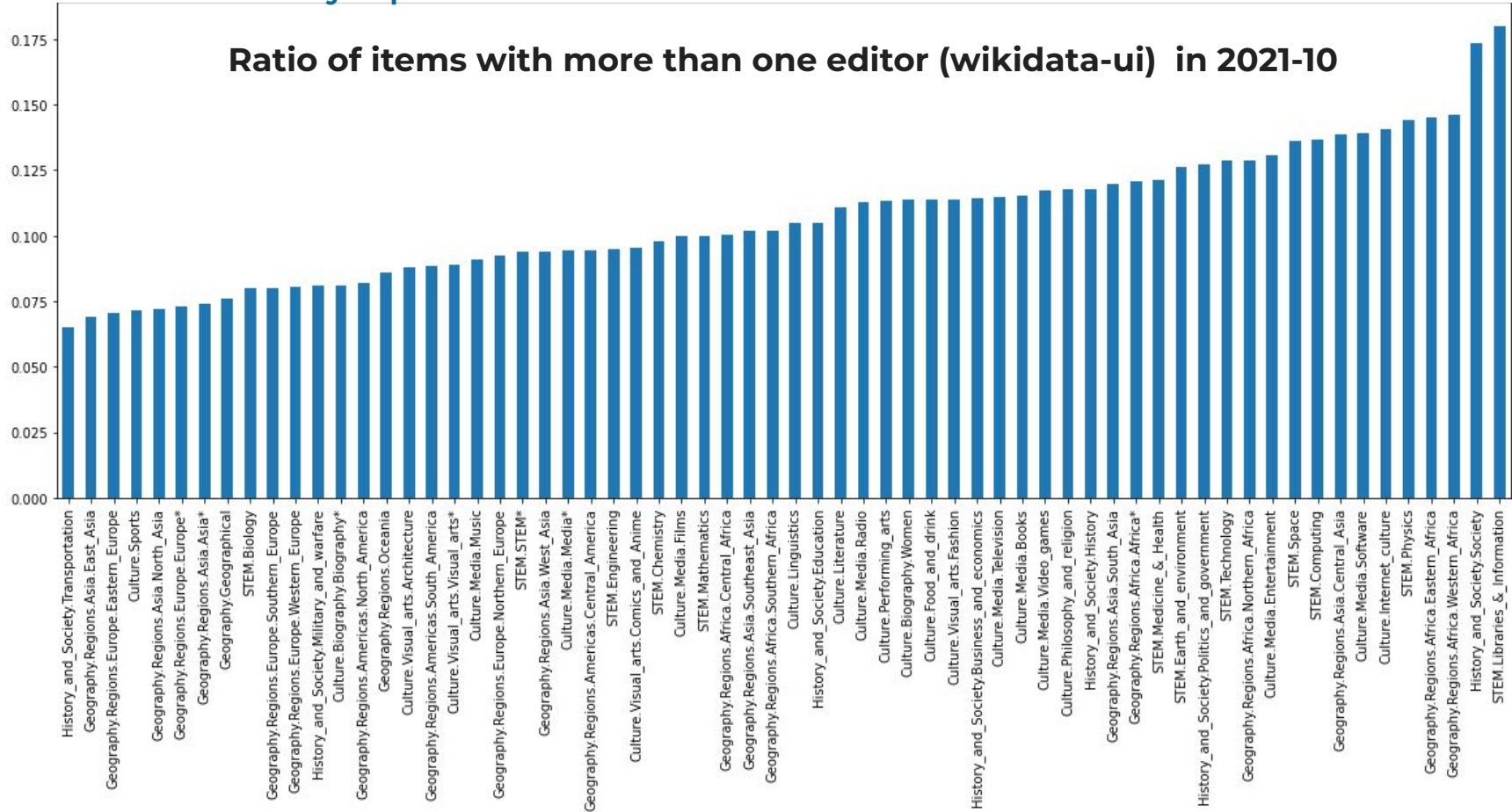
Wikidata: The **95%** of the items were edited just by one user.



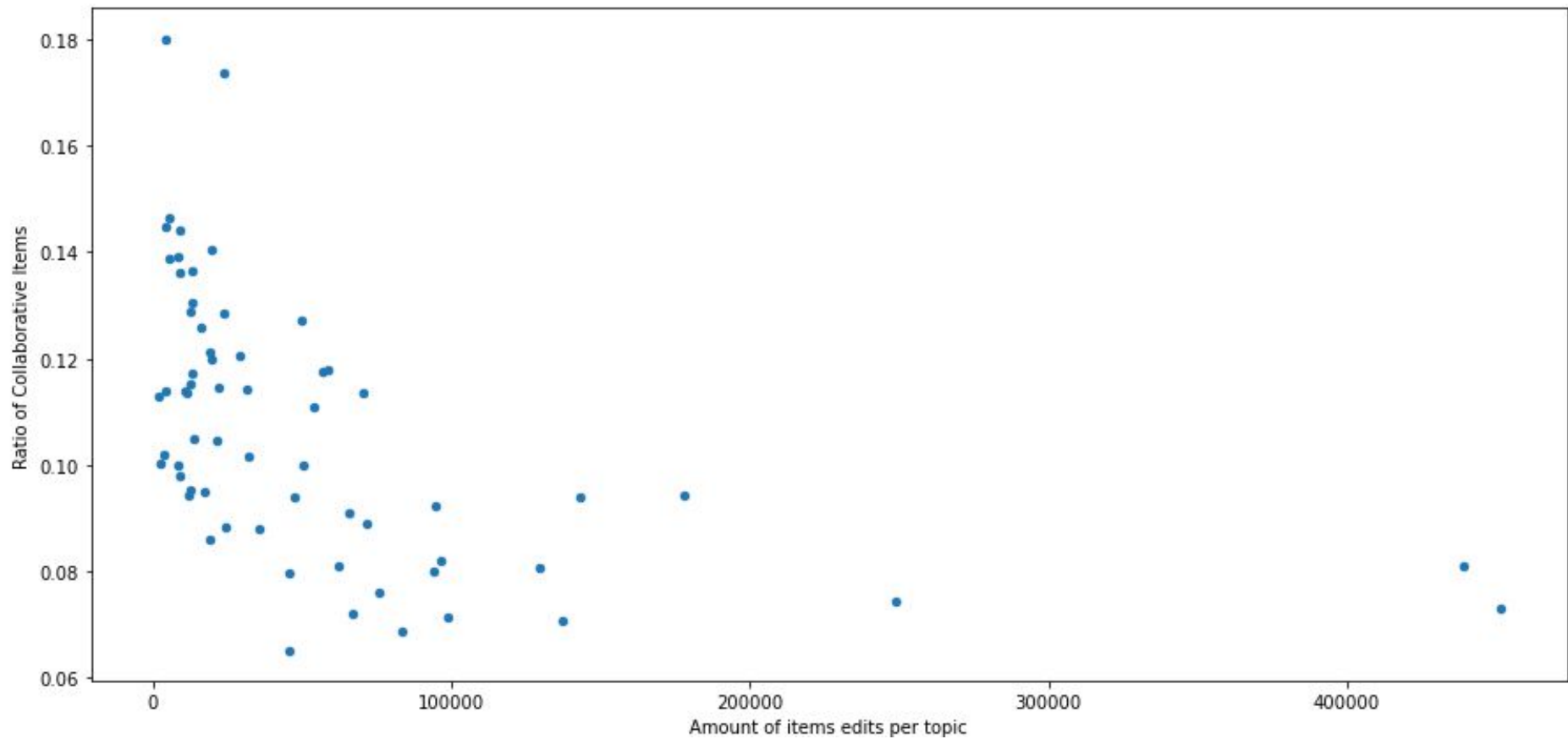
of editors per item
In Wikidata

num_users	num_items	ratio
1	440806	0.947167
2	21159	0.045465
3	2392	0.005140
4	566	0.001216
5	192	0.000413
6	89	0.000191
7	46	0.000099
8	29	0.000062
9	32	0.000069
10	10	0.000021

Co-edited items by topic



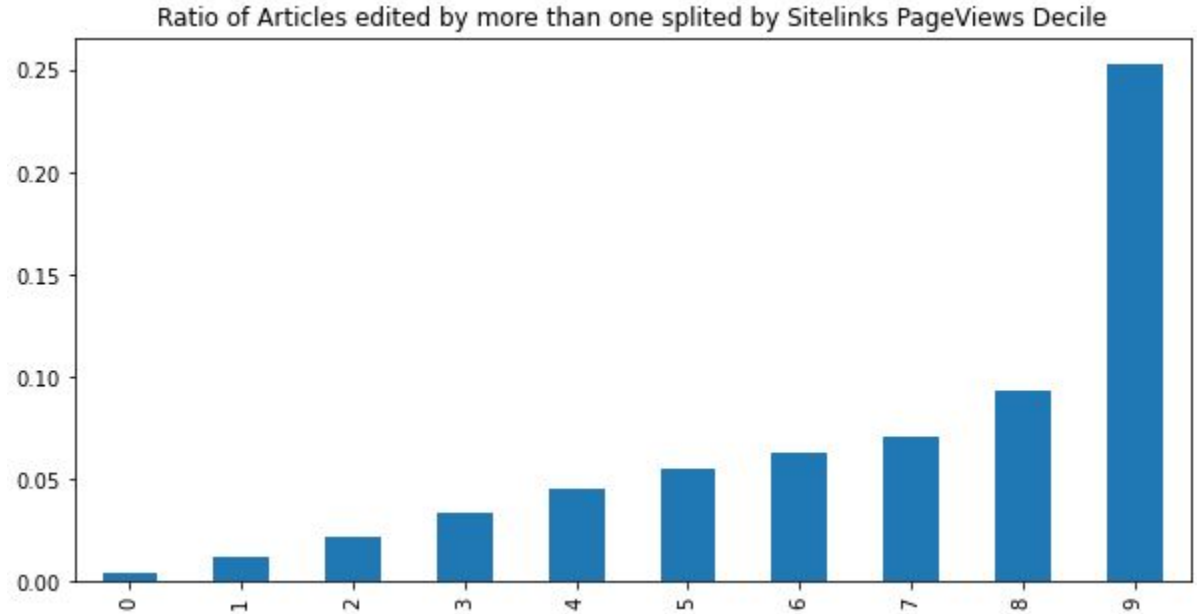
Relation between collaborative-items and total amount of edits per topic



Relation between items “popularity” and amount of editors

We compute the popularity of an item summing all the pageviews on it's sitelinks.

More popular items tends to have more editors.



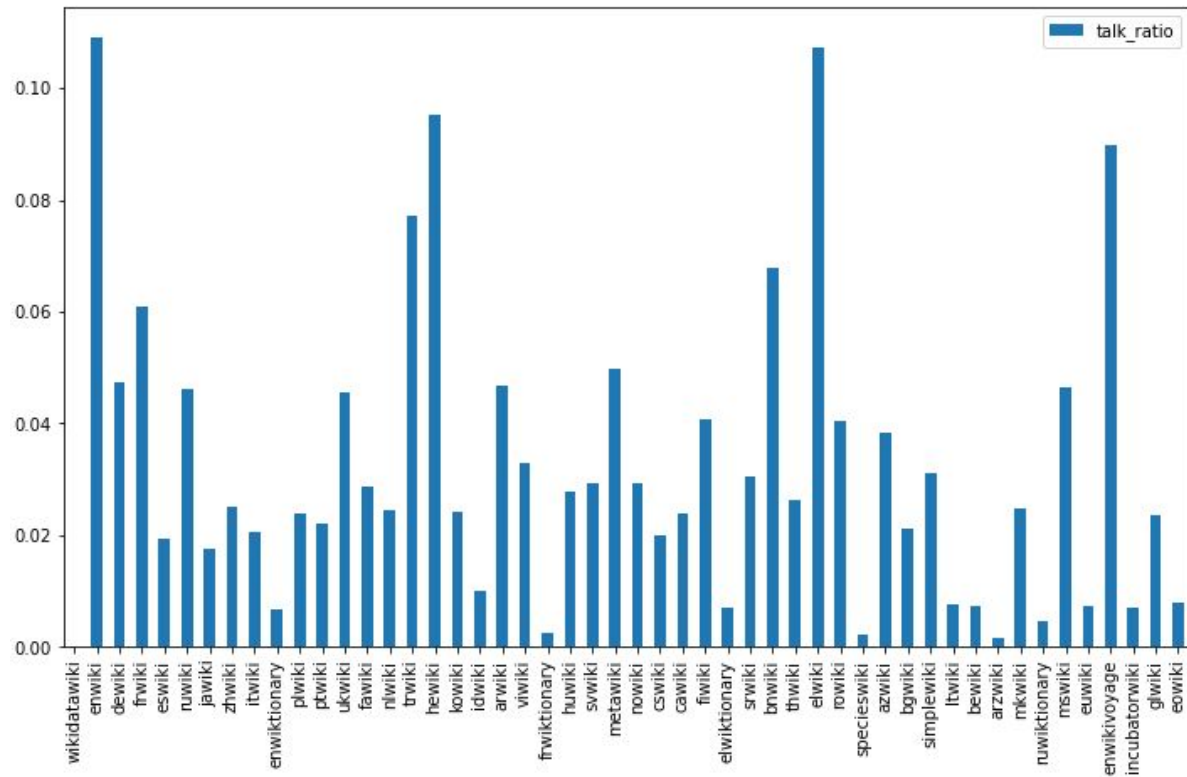
Items with more editors (2021-10, wikidata-ui)

Most of the items on this list are related with events (nobel prize, death, terror attack)

label	reason for edits
Abdulrazak Gurnah	nobel prize
Halyna Hutchins	death
Squid Game	popular tv show
Pandora Papers	media story
Meta Platforms	company rename
Dmitry Muratov	nobel prize
Ardem Patapoutian	nobel prize
Syukuro Manabe	nobel prize
Benjamin List	nobel prize
Klaus Hasselmann	nobel prize
David Julius	nobel prize
Alexander Schallenberg	became new chancellor of austria
David Amess	death
David MacMillan	nobel prize
Kongsberg attack	terror attack
Maria Ressa	nobel prize
Giorgio Parisi	nobel prize
Fumio Kishida	became prime minister of japan
Frances Haugen	blew the whistle on facebook
HoYeon Jung	stars in squid game
October 2021 Sudanese coup d'état	coup happened
Colin Powell	death
2021 Facebook outage	facebook went down
squid game	game the tv show is related to
Lars Vilks	death
ITA Airways	airline privatisation

Usage of talk pages

There almost no usage of talk pages in Wikidata. During 2021-10 just 1331 edits



Next steps

- A new model for automatic:
 - edit quality assessing
 - Item quality evaluation

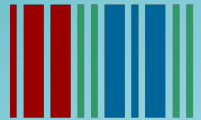


Thanks for your attention!

Diego Sáez-Trumper
User:Diego (WMF)

Lydia Pintscher
User:Lydia Pintscher (WMDE)

More info:
[[m:Research:Identifying_Controversial_Content_in_Wikidata]]



WIKIDATA
DATA QUALITY
DAYS 2022