

Content Integrity via Wikimedia Enterprise

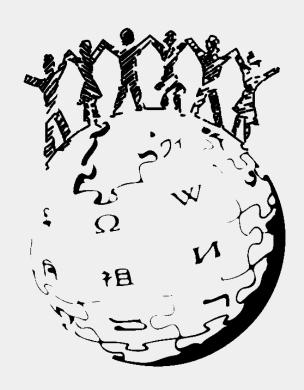
by Francisco Navas, Product Manager | November 2023 for WikiCon



Written by everyone and anyone

The largest, most complete, cross-lingual, dataset of *verifiable* facts and histories.

Editors in projects in more than 300 of their native languages





Our content and data are used as the backbone for training LLMs and act as information retrieval or identifiers in **Knowledge Graphs**

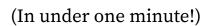


01

How did Google know Queen Elizabeth II had passed?



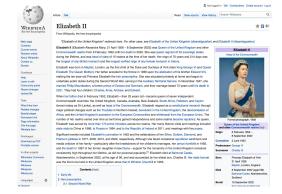


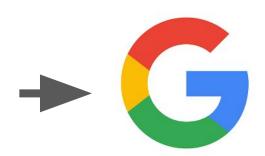




by CHRIS TOMLINSON | 14 Sep 2022 | 33

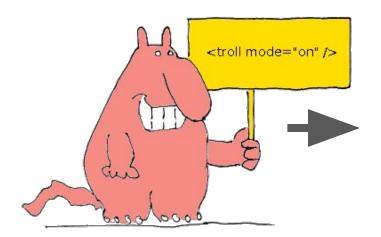


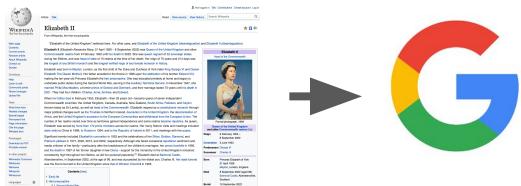




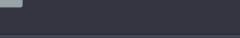
(In under one minute!)







(In under one minute!)



\$

As of my last knowledge update in January 2022, Queen Elizabeth II was alive. However, I do not have access to real-time information. To find out the current status of Queen Elizabeth II, I recommend checking the latest news or an official source.



Using Wikimedia content in

a third-party environment

carries some

challenges

- 1) Speed
- 2) Usability
- 3) Machine Readability
- 4) Content Integrity



"Trustworthiness" is a key pain point

"Who's using Wikipedia and for what purpose? Indirectly we care about the community, more editors for you means more content for you means more content for us."

"How do you distinguish between assertions you're making that editors have vetted versus those added by a user a few minutes ago? Internal systems that could distinguish that would be super valuable to us."



We seek to mitigate these challenges to support the proliferation of free and open-source knowledge, engendering trust in the content and the work of the Movement

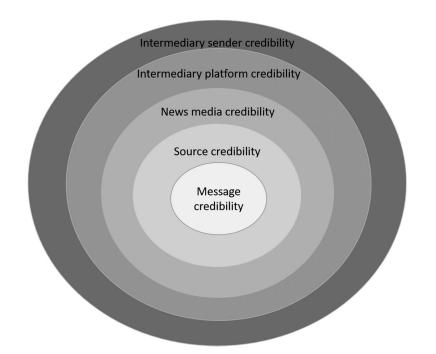


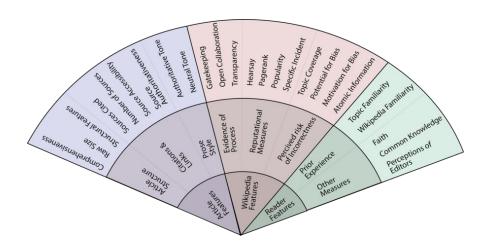
02

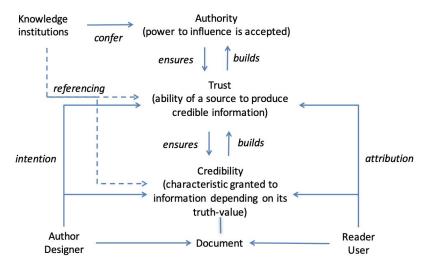
Research



What is credibility?









Breaking News on Wiki

"Early editors of breaking news articles are unlikely to contribute at similar stages to other breaking news articles. In contrast, historical and nonbreaking articles see a more gradual attrition among articles sharing editors,"

— Joe Sutherland, T&S

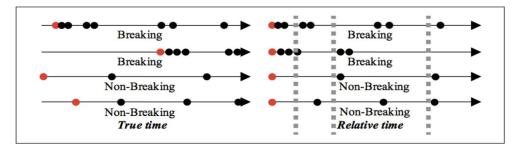
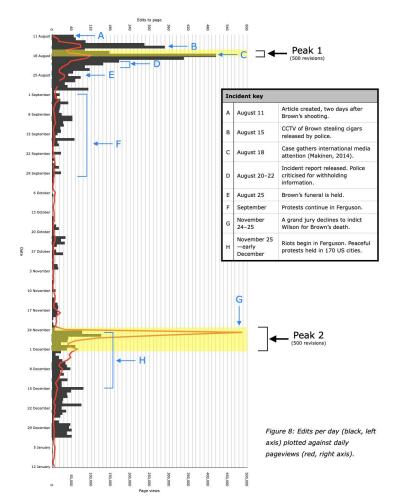


Figure 2. Schematic representation of edit history in true time, and its transformation and alignment into relative time, for two breaking articles and two nonbreaking articles Red dots reflect the first edits made to an article by an editor, black dots reflect subsequent revisions to that article. Examples of time thresholds are given by vertical gray dotted lines.

— Brian Keegan, CU Boulder

4.1 IDENTIFYING SPIKES AND PEAKS





What we need to answer to signal credibility per revision

How do Wikipedia projects existing content integrity practices function?

How can we impact reuser understanding and trust of Wikipedia content?

How can our signals mirror community efforts in content integrity?

How can our signals be as useful to Wikipedia communities as they are to third party reusers?

How do we communicate all this programmatically?



Key learnings

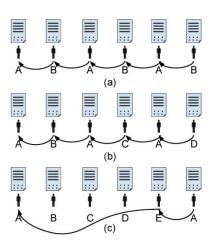


Figure 3: Types of Edit Wars: (a) Single-revision/Alternating users Edit wars (b) Multi-user Edit Wars (c) Multiple-revisions Edit Wars

- 1) Higher contact equals higher quality, even when the short-term result is vandalism
- Context, semantic and non-semantic, are key
- We must take care in how we make this information publicly accessible and usable



Wikipedia has 3 Atomic Units

- 1. Editors
- 2. Edits
- 3. Citations/References

The same must drive signals ...

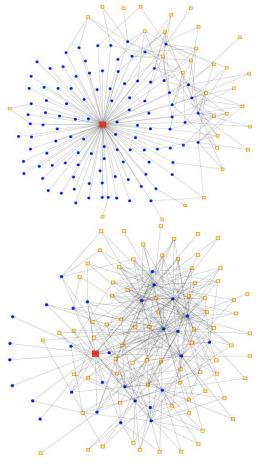


Figure 1: Two sample edit networks from Wikipedia. The red squares are the key articles, the yellow squares are other articles and the blue circles are editors. The top article is "Family History" a Start Class article, and the bottom article is "Eardwulf of Northumbria" a Featured Article.



03

Credibility Signals



Credibility Signals support practices of data validation.

They are not declarative.

Editors use signals on the page to make integrity decisions already — Enterprise will make them accessible in a structured manner to content reusers to use in the same way

Out intent is to work with interested customers to ensure we're packaging these signals so they they can be consumed in the best and most efficient way possible



Credibility signals provide community-sourced contextual information that can help our reusers make sense of when a particular edit is to their standard, in real-time.



Examples of Credibility Signals

- they've been editing, how many edits they have, access or permission levels
- 2) Semantic: Writing style, naive vs cynical vandalism
- Non-semantic: Sourcing, controversy, and edit warring



Indicators of Reliability

- Machine learning based credibility scores
- Articles flagged as needing citations
- Auto Patrolling, Flagged Revisions, etc.
- Time of live revision
- Locked article status, changes
- Editor experience
- Editors watching data

Other

- Frequency of revisions
- Redirects and constantly moving articles
- Feed/List of perennial sources
- Feed/List of conspiracy sources
- Pageviews

Indicators of Controversy

- Recently reported vandalism or sockpuppet activity
- "Oversighted" revisions
- Events on specific time horizons
- Discussion Talk page data



O4 What can you do with Credibility Signals?



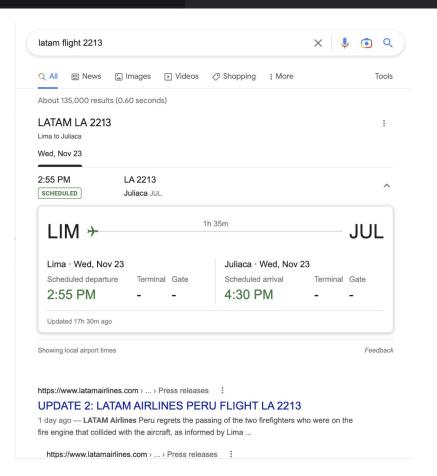
Combining credibility signals creates new types of functionality, like our "Breaking News" signal





Name ψ	Unique Editors ↓	Unique Editors Within 1st Hour ψ	Anonymous Editors Within 1st Hour ψ	Editors Ratio Within 1st Hour ψ	Number Of Edits ψ	Project ψ	Indications ψ	Date Created ↓	Date Modified ψ	Date Namespace Moved ψ
2022_missile_explosion_in_Poland	131	12		0.4166666666 666667	1250	English Wiki (enwiki)	Template:Cite news Template:Current event	2022-11-15T18:43:58Z	2022-11-17T09:00:33Z	
Donald_Trump_2024_presidential_campa	36	12		0.083333333 3333333	326	English Wiki (enwiki)	Template:Cite news Category:Current events	2020-11-12T05:33:06Z	2022-11-17T07:00:33Z	2022-11-16T02:25:42Z
Sam_Trabucco					46	English Wiki (enwiki)	Template:Cite news Template:Current person	2022-11-15T18:53:40Z	2022-11-17T07:05:31Z	
Donald_Trump_2024_presidential_campa					24	English Wiki (enwiki)	Template:Cite news Category:Current events	2020-11-15T18:19:33Z	2022-11-16T02:24:19Z	2022-11-16T02:24:23Z
Jorge_Barrosa						English Wiki (enwiki)	Template:In use Template:Cite news	2022-11-15T23:56:40Z	2022-11-15T23:56:48Z	
Juan_Brito_(infielder)						English Wiki (enwiki)	Template:In use Template:Cite news	2022-11-16T00:53:08Z	2022-11-16T00:53:20Z	
José_Cruz_(pitcher)					12	English Wiki (enwiki)	Template:In use Template:Cite news	2022-11-15T23:42:34Z	2022-11-15T23:44:14Z	
Colin_Selby						English Wiki (enwiki)	Template:In use Template:Cite news	2022-11-15T23:24:39Z	2022-11-15T23:24:48Z	
Keaton_Winn						English Wiki (enwiki)	Template:In use Template:Cite news	2022-11-15T23:41:29Z	2022-11-15T23:41:58Z	
2022_Russian_missile_strike_on_Poland						English Wiki (enwiki)	Template:Cite news Template:Current event	2022-11-15T18:43:58Z	2022-11-15T21:02:30Z	
2022年導彈擊中波蘭事件					201	Chinese Wiki (zhwiki)	Template:Cite news	2022-11-15T20:41:19Z	2022-11-17T00:26:08Z	

WIKIMEDIA





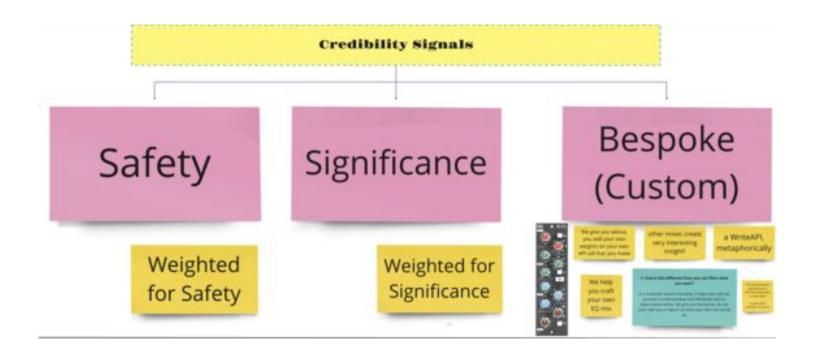
w en.wikipedia.org

LATAM Perú Flight 2213 - Wikipedia

LATAM Perú Flight 2213 (LP2213/LPE2213) was a scheduled domestic passenger flight in Peru from Lima to Juliaca. On 18 November 2022, the Airbus A320neo operated by LATAM Chile was taking off from Jorge Chávez International Airport when it collided with ...



The Future





Content Integrity Facts

760 Total Revisions

Event Type

New Article Revision

Safety	Score .22	Signif.	Score .91
Visibility	0.01	Breaking News	true
Stability	0.9	Edit type	сору
Collaboration	0.333	Topic	music
'Revertrisk' (ML)	0.21	Editor (exp.)	0.9
Controversy	0.1	Trending	true
Citations	0.77	Art. Quality (ML)	0.4



critiques?

questions?

Thank you

ideas?

concerns?

say hello!!!

thoughts?

Reach out to me via email

fnavas@wikimedia.org



Aknowledgements

Sam Walton Claudia Lo **Peter Pelberg** Olga Vasileva **Alex Stinson Alex Hollender Costanza Sciubba Caniglia** Joseph Seddon **Mikhail Popov Diego Saez-Trumper** Isaac Johnson Camelia Boban **Denis Barthel Ed Erhart** Joe Sutherland

Silvia Gutiérrez de la Torre Ryan Brounly Liam Wyatt Luca de Alfaro Jake Orlowitz Brian Keegan Pablo Aragón Julia Stoyanovich Hal Triedman