

Citation Location Needed

In-Source Locators and Wikipedia's Growing Verification Cost

doi.org/10.1145/3805696.3846057

Mario Morvan

JCDL 2026

Background

- **Verifiability** on Wikipedia:

“people can check that facts or claims correspond to reliable sources”

- In practice : only few editors do check sources (0.3% of clicks in 2020, [Piccardi et al])
 - **Automated verification** progressing but : imperfect, costful, ≠ human verification
 - Unknown rate of **claims unsupported** by their sources
- ...but likely substantial : 40% on WAFER [Petroni et al], 44% on WICE [Kamoi et al]
- Other axes of citation studies :
citation need, span, verification, reliability, source reliability and accessibility

Wikipedia citation infrastructure

- Citation = anything within <ref> tag
- References can be named for reuse in an article
- Templated VS untemplated citations
- Citation templates → structure fields for various source types
- Citation groups on English Wikipedia : CS1, CS2, short cites...

In-source locators : concept

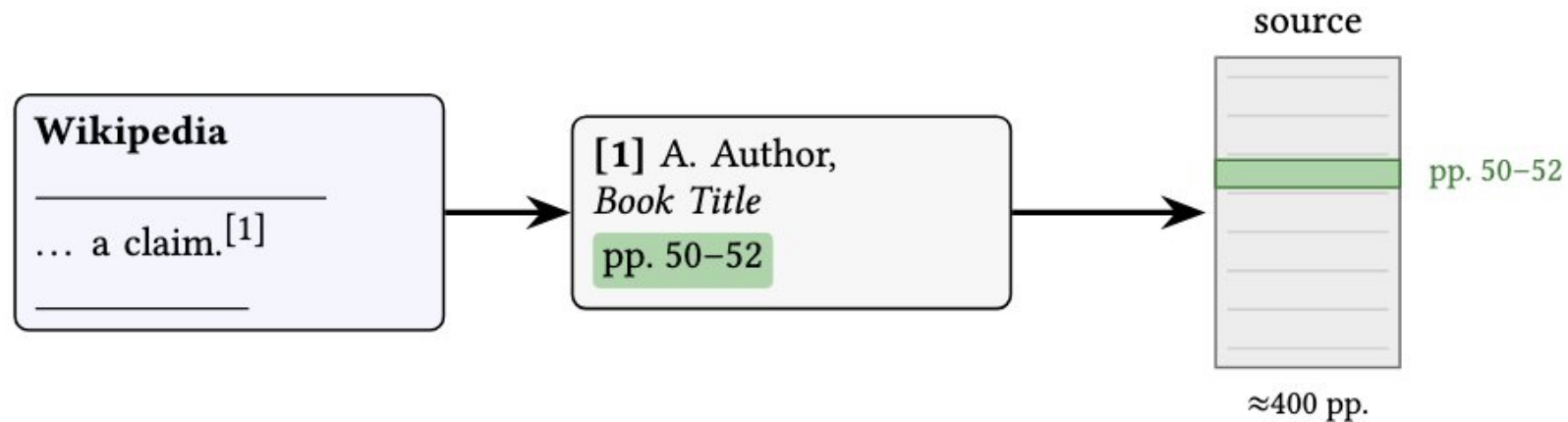


Figure 1: An *in-source locator* (here pp. 50–52) points a citation to the specific passage that supports a claim, narrowing verification from the whole source to a smaller part.

Questions

- *What use of located citations on English Wikipedia?*

6 random samples of 20,000 enwiki articles from 2014 to 2026

precise parsing of references and locators

- Resolves parameters across templates and named references

wikiloc parser is released on Github and PyPi

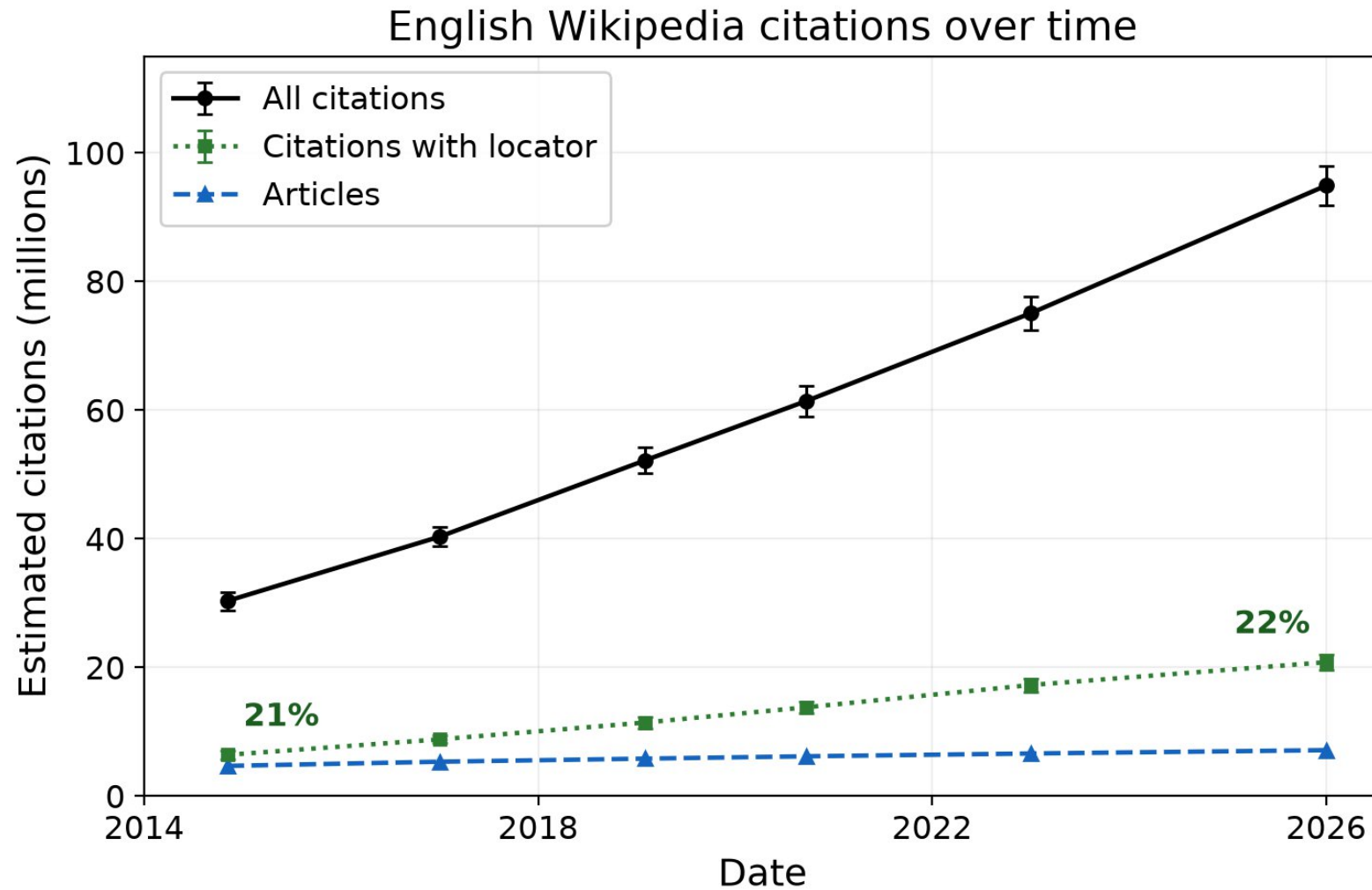
- *What associated verification cost?*

Retrieved located & total number of pages for books and articles

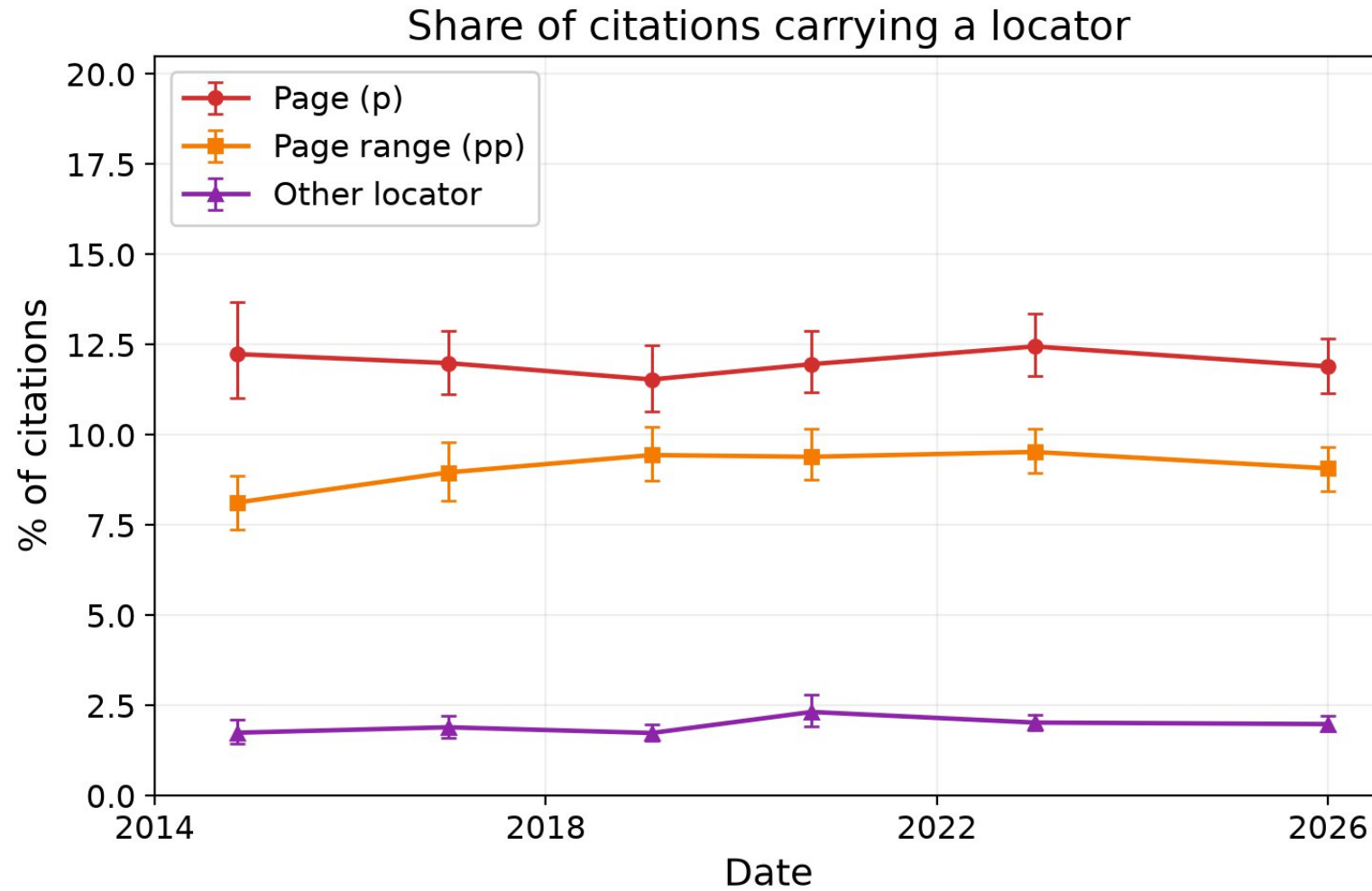
2026 locator coverage

Template	# cit.	% of citations with locator			
		any	page	pages	other
<i>All</i>	266,728	21.9	11.9	9.1	2.0 (chapter, quote, loc)
cite web	124,471	2.0	0.8	0.4	0.8 (quote)
<i>untemplated</i>	48,705	22.5	16.8	5.8	–
cite news	35,894	16.8	12.0	3.4	1.7 (quote)
cite book	17,590	71.7	34.5	34.7	13.5 (chapter, quote, at)
cite journal	12,893	85.6	9.8	76.0	1.3 (quote, at)
sfn	9,861	88.1	68.4	18.0	2.4 (loc)
cite magazine	3,788	34.4	26.2	7.6	2.2 (quote)
citation	2,653	42.3	13.3	25.2	8.1 (chapter, quote, at)
harv	1,892	91.3	65.5	22.3	8.0 (loc)
rp	873	99.7	80.3	19.4	–
sfnp	817	74.7	55.8	14.2	5.1 (loc)
cite av media	680	11.8	1.3	0.3	10.3 (time, chapter, at)
cite certification	627	0.0	0.0	0.0	–
cite report	594	25.4	13.5	10.6	6.9 (chapter, quote)
cite encyclopedia	507	29.4	8.3	19.5	8.7 (chapter, quote)

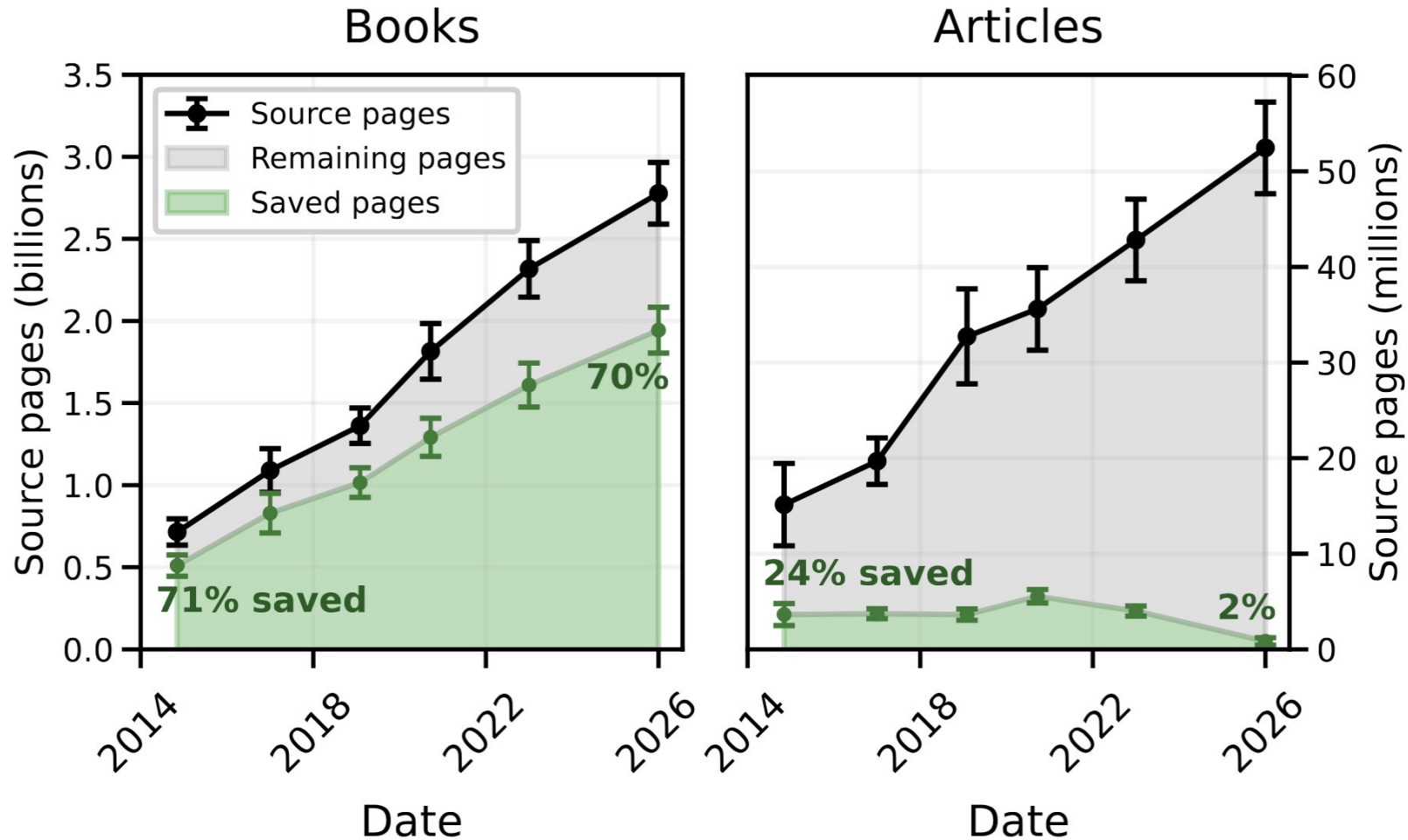
Locator evolution



Locator evolution



Verification cost evolution



Discussion and conclusion

Locators are **underused** and could make verification **more accessible**

- **Less time** spent looking for relevant information
- Short fragments could be **linked** on Wikipedia

Tools : <https://github.com/mrmvn/wikiloc/tree/main/tools>

- Audit article citations when it comes to locators
- Suggest edits for some standard misuse of locator parameters

Further work :

- Combine with **automated verification**
- Extend parser to support **other languages**
- Extend cost study to **other citation templates**.

Relevance **beyond Wikipedia**

Thank you !

- References :

- Wikipedia:Verifiability, <https://en.wikipedia.org/wiki/WP:Verifiability>
- Wikipedia:Citing_sources, https://en.wikipedia.org/wiki/Wikipedia:Citing_sources
- Tiziano Piccardi, Miriam Redi, Giovanni Colavizza, and Robert West. 2020. *Quantifying Engagement with Citations on Wikipedia*. In Proceedings of The Web Conference 2020 (New York, NY, USA) (WWW '20). 2365–2376. [doi.org/10.1145/ 3366423.3380300](https://doi.org/10.1145/3366423.3380300)
- Ryo Kamoi, Tanya Goyal, Juan Rodriguez, and Greg Durrett. 2023. *WiCE: RealWorld Entailment for Claims in Wikipedia*. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (Singapore). 7561–7583. [doi.org/10.18653/v1/2023.emnlp- main.470](https://doi.org/10.18653/v1/2023.emnlp-main.470)
- Fabio Petroni, Samuel Broscheit, Aleksandra Piktus, Patrick Lewis, Gautier Izacard, Lucas Hosseini, Jane Dwivedi-Yu, Maria Lomeli, Timo Schick, Michele Bevilacqua, Pierre-Emmanuel Mazaré, Armand Joulin, Edouard Grave, and Sebastian Riedel. 2023. *Improving Wikipedia verifiability with AI*. Nature Machine Intelligence 5, 10 (Oct. 2023), 1142–1148. doi.org/10.1038/s42256-023-00726-1