Toward a Systematic Assessment of Map Reproduction Success Using OCR

ACM REP '25 proceedings

Link to

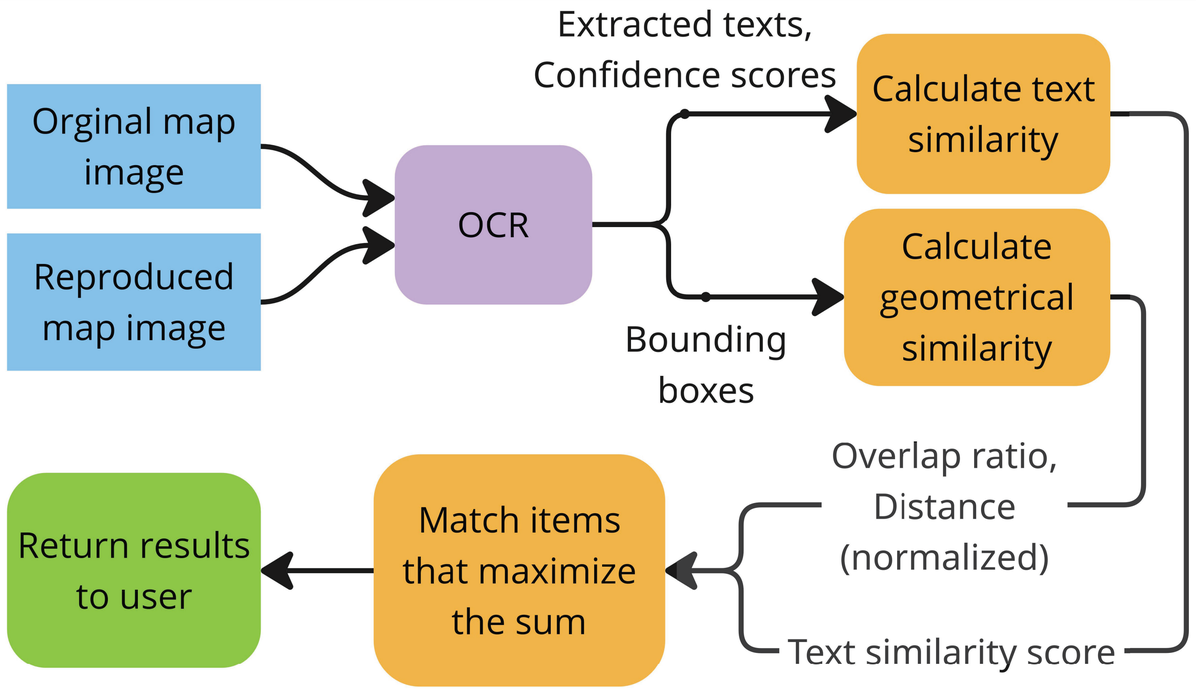
Yohannes Abrha Mulaw, Eftychia Koukouraki, Christian Kray
Institute for Geoinformatics - University of Münster, Germany

Motivation and Problem Statement

- Maps are key in communicating spatial research findings.
- The assessment of reproduced maps typically relies on a manual, side-by-side comparison with the original. This method is subjective and time-consuming.
- There are currently no standardized metrics for determining reproduction success.
- Textual and spatial discrepancies in reproduced maps affect the meaning and interpretation.

Suggested Approach

- Uses Optical Character Recognition (OCR) to extract text from map images.
- Matches and compares text based on:
  - Text similarity (Levenshtein distance).
  - Geometrical similarity (bounding box overlap ratio and centroid distance).
- Supports content and layout comparison and does not rely on pixel-level discrepancies.
- Helps detect missing or misplaced labels, mislabeled data, and legend differences.
- Web-based tool with intuitive UI (see below).

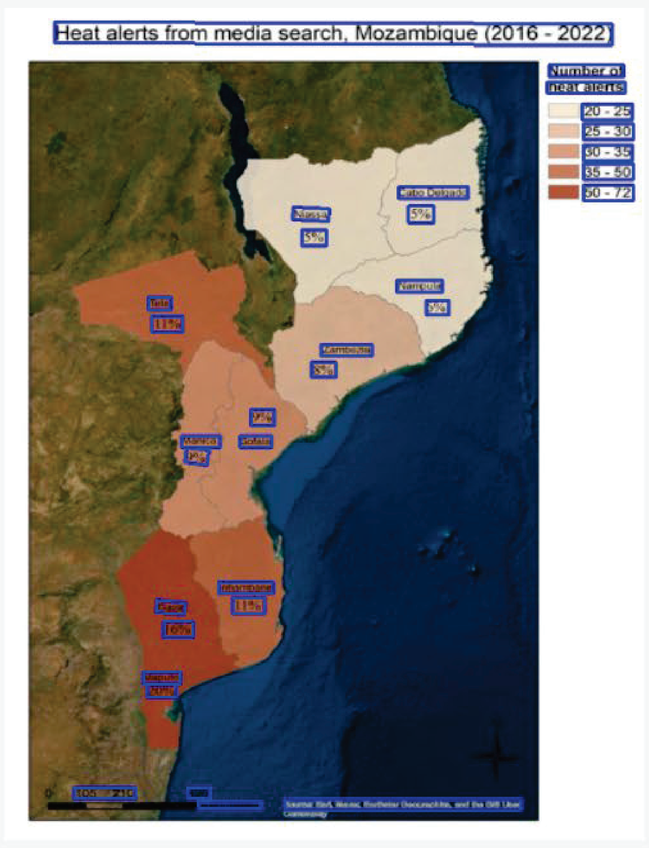


Reproducibility Assessment Results

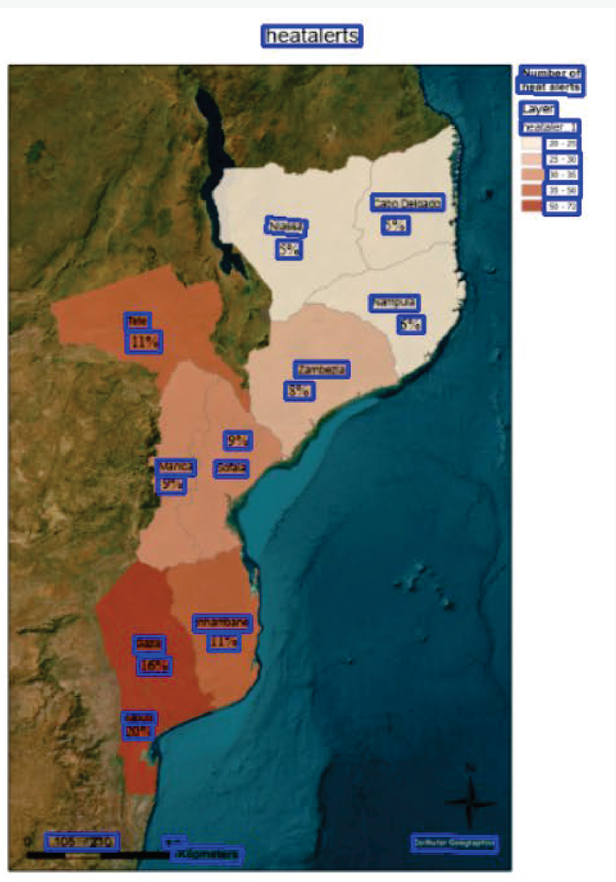
Original Text	Reproduced Text	Text Similarity Score	BB Overlap Ratio	Distance b/n Bounding Boxes	Final Score	Match Status
6%	6%	100	0.6616	0.0026	100	Matched!
11%	11%	100	0.6709	0.0047	100	Matched!
Zambezia	Zambezia	100	0.6373	0.0032	100	Matched!
8%	8%	100	0.6802	0.0024	100	Matched!
%6	No Similar text Found	0	0.0000	0.0000	0	Not Matched
Manica	Manica	100	0.6339	0.0064	100	Matched!
Sofala	Sofala	100	0.7145	0.0034	100	Matched!

‹ Page 3 ›

Download Excel



Original Map



Reproduced Map

Contribution

- Automates part of the map comparison process.
- Makes reproducibility checks easier and more systematic.
- Enables deeper exploration of differences in scientific maps.

Outlook

- Expand the assessment to include other visual variables, such as colors, symbols, and shapes.
- Add more languages beyond English.
- Use more sophisticated methods to quantify map content differences.

References

[1] Koukouraki, E. and Kray, C. (2023). Map Reproducibility in Geoscientific Publications: An Exploratory Study. *In: Beecham, R.; Long, J. A.; Smith, D.; Zhao, Q.; Wise, S. (eds.), 12th International Conference on Geographic Information Science (GIScience 2023).* Dagstuhl, Germany: Dagstuhl Publishing.
[2] Koukouraki, E. and Kray, C. (2024). A systematic approach for assessing the importance of visual differences in reproduced maps. *In: Cartography and Geographic Information Science.*
[3]Koukouraki, E. (2023). Reproducibility review of: Extreme heat alerts and impacts across Mozambique 2016 - 2022: gathering evidence from media articles. Publisher: OSF.
[4]Pereira Marghidan, C., van Aalst, M., Blanford, J., Maure, G. and Marrufo, T. (2023). Extreme heat alerts and impacts across Mozambique 2016 – 2022: gathering evidence from media articles. *In: AGILE: GIScience Series (AGILE 2023).* Publisher: Copernicus GmbH

Check out the app code!

eftychia.koukouraki@uni-muenster.de

Universität Münster

ifgi