

Is Small More Interesting? Examining Countries' GeoNames Places linked to Wikipedia

Dirk Ahlers
NTNU – Norwegian University of Science and Technology
dirk.ahlers@idi.ntnu.no

ABSTRACT

Following up on previous analyses [1], we examine the geospatial and thematic data in GeoNames [2]. It is the largest freely available gazetteer – a geographical thesaurus – with a worldwide coverage. One measure of interestingness for a country can be its number of populated places represented as pages in Wikipedia. In GeoNames, on average only 20% of populated places, i.e., cities, towns, villages, etc. have a link to Wikipedia. This clearly shows that not all places are created equal in terms of interestingness.

Of the multitude of interesting lessons, we pick one that matches the small nature of this paper. For simplicity, we define a 'small' country as one which has a population that is within the first quartile of all global population numbers (1.76×10^5 inhabitants). For all 63 identified small countries, we compare populated places with links to Wikipedia to the remaining countries. We hypothesise that small countries may show a significant higher percentage than large ones in their linkage. First, there are not that many populated places and they carry a higher significance or interest. Second, for larger countries, more very small places may have been added to GeoNames. Of course this argument is about relative coverage; absolute coverage of small countries is less than 1% of populated places or Wikipedia links on a world scale.

This finding can inform reasoning about biases in information density and coverage on a world scale, for example for small or developing countries.

BODY

Places in small countries are three times as likely to be linked to Wikipedia with 41% of coverage compared to 13% for larger countries.

REFERENCES

- [1] D. Ahlers. Assessment of the Accuracy of GeoNames Gazetteer Data. In *Proceedings of the 7th Workshop on Geographic Information Retrieval, GIR '13*. ACM, 2013.
- [2] GeoNames.org. GeoNames Dataset, <http://download.geonames.org/export/dump/>.

Volume 4 of Tiny Transactions on Computer Science

This content is released under the Creative Commons Attribution-NonCommercial ShareAlike License. Permission to make digital or hard copies of all or part of this work is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
CC BY-NC-SA 3.0: <http://creativecommons.org/licenses/by-nc-sa/3.0/>.