

Geospatial Grounding of the Web Graph

Dirk Ahlers, NTNU

Abstract

The pages and directed hyperlinks of the Web make up the Web graph, one of the largest graph structures commonly analyzed [KR+00]. We think of this graph as largely immaterial, but we can generate links towards the real world to ground it and allow additional spatial exploration [BA08].

Geoparsing and geocoding allows to georeference Web pages, images, or social media snippets and users to pinpoint them on a map. The georeferencing can even be extended to the links between them. As shown in Figure 1, this adds the geospatial dimension to form a bipartite graph of the Web and the geospatial world [Ah11]. For nonconsecutive pages with higher link distance between them, spatial links can still be constructed with a cumulative or emerging spatial distance. The principle also holds for pages with multiple locations, but the resulting graph then gains complexity. Additional complexity is introduced by media that can only be georeferenced to a rough area such as a city or region, as distance measure have to account for the resulting vagueness.

Spatial links allow additional analysis such as distance of links, footprints of pages, outreach, similarity of regions, ranking of pages depending on local or global impact, etc. Possible applications are improved Geospatial Search, added visualization support, or improved understanding of social media. We will present further technical details of the grounded Web graph and explore analyses and applications drawing from it.

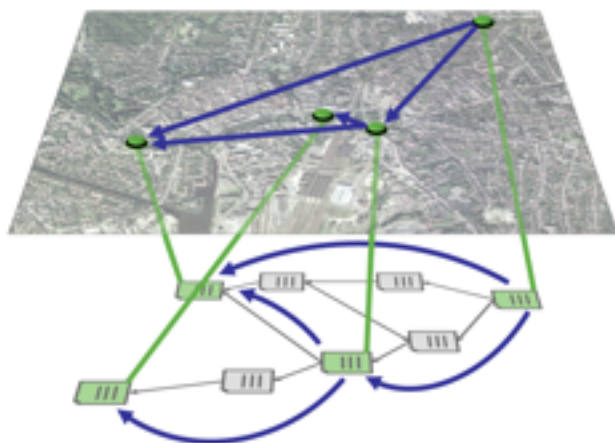


Figure 1. Exemplary selection of the Web graph with normal (grey) and georeferenced pages (green) and links between them. Links between georeferenced pages gain a geospatial component (blue) and can be put on the map.

References

- [Ah11] Dirk Ahlers. Geographically Focused Web Information Retrieval. OIWIR, 2011.
- [BA08] Susanne Boll, Dirk Ahlers. A Web more Geospatial: Insights into the Location Inside. In WebEvolve2008 at WWW2008, 2002.
- [KR+00] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tompkins, and Eli Upfal. The Web as a graph. In PODS '00. ACM, 2000.