

## David Anderson

The MITRE Corporation

[dca@mitre.org](mailto:dca@mitre.org)

781-271-2594



### Bio & Current Work

BS Computer Science and Engineering, MIT, 1990. Spent ten years in the commercial software development business, i.e. the dot-com bubble. Worked on a full-text retrieval system, starting before the web. Also worked on scalable parallelizing middleware for business intelligence applications. My interest in digital gazetteers dates to 2001, with the development of a system to extract place-names from Colombian news websites and plot them on a map of Colombia using the NGA GeoNames dataset. Management of place-name ambiguity has been an enduring challenge.

My interest is in developing new solutions to real problems, based on combining an understanding of what's possible with a sense of the challenges faced by my customers. I have three immediate problem areas:

- Supporting the process of updating the NGA GeoNames dataset by identifying matches and new entries in new supplement geospatial / names datasets. I suspect this will generalize to matching entries across disparate gazetteers. This may contribute to solving the larger problem of federating many small gazetteers.
- Improving the geo-referencing and typing of ambiguous or vague textual references. This is the beginning of a 3-year research effort with Inderjeet Mani to develop a spatial markup language and foster a community of interest analogous to that which has been effective in addressing temporal references with TimeML. We also want to reason over these spatial and temporal attributes of entities and events in text, and hopefully apply these models across languages.
- I'm interested in extracting text names and references to feature types that accompany map coordinates in text. This can initially provide a dense gist of the map dot, and potentially constrain search for matching entries in gazetteers or other spatial datasets in the face of imprecise spatial positioning.

“Practical Joining of Geographic Feature Datasets” accepted for presentation to the 7<sup>th</sup> International Conference on Military Geology and Geography (ICMGG'07). Presents a method for aligning gazetteer entries in different datasets that predicts ground-truth join keys with 96% accuracy.