Greg Janée is director of the newly-launched Data Curation Program at UC Santa Barbara, which offers consultation services to campus researchers and he is working now to establish a curated research data storage facility at UCSB. He also works at the California Digital Library, where he is principal developer of the EZID persistent identifier service. His previous work was on the Alexandria Digital Library (ADL) and related efforts. In addition to serving as principal developer of the ADL system, his research areas included query languages and frameworks, query rewriting, metadata mapping, spatial ranking, search result visualization, and provenance analysis. He is co-author of the ADL Gazetteer Protocol.

Statement

In my view, spatial search (specifically, geographic search) is hampered by mismatches and bifurcations. The problems are well known:

- Spatial extent can be described precisely: think of the complex, projection—and datum-aware geometries that can be expressed in GML. But many spatial search user interfaces and search engines operate only on bounding boxes, which suffer from precision problems.
- Reconciling fundamentally different ways of addressing space—coordinate-based descriptions and placenames—remains difficult.
- There are multiple ways any given spatial extent can be described due to differences of scale and simply differences of choice. Polygon, box, or simply a point? Which placename(s)? And in practice, there is little uniformity in how spatial resources get cataloged.
- Spatial and non-spatial search are poorly integrated, if at all. A resource lacking a spatial extent typically won’t show up at all in a spatial search engine, and conversely spatial extents typically provide no input to a non-spatial search engine.

All these problems make for a spatial search experience that is fragile in the face of the limitations of real-world metadata, which is often incomplete, incorrect, misplaced, inappropriate, or simply missing. Contrast this with web search, which is much more resilient and works quite effectively despite having to contend with malformed HTML, spelling errors, broken links, and the like. Might the advances that enabled the web search we enjoy today provide a path to better spatial search? Web search has improved over the last decade, not by getting better at finding words, but by collapsing file formats and data types into a single search space, by recognizing relationships between documents, and by employing sophisticated ranking algorithms. Are there analogies for spatial search?