Spatial Search

GRANT MCKENZIE
Doctoral Candidate, Department of Geography
University of California, Santa Barbara
Email: grant.mckenzie@geog.ucsb.edu

The term *Spatial Search* is loaded in that it caries with it multiple definitions, different meanings for various people across a range disciplines. The word S*earch* itself may reference physically searching, mentally searching or executing data searches online. The *Spatial* component of the phrase may speak to spatial objects in our natural environment, spatial relationships between entities or web documents tagged with specific geospatial locations (to name a few). In this position paper I take *Spatial Search* to mean the querying of (representations of) geospatial objects through an online service.

Spatial search has made significant advances in recent years from a number of different perspectives. Technologically speaking, spatial search has benefited from the ubiquity of location-enabled mobile devices driven by reduced cost and increased accuracy of technologies such as GPS. From a data perspective, the quantity of user-generated content has drastically enhanced existing gazetteers and given rise to new location-dependent online social networks (e.g., Foursquare). From a computational side, advances in spatial indexing, linked-data stores and geographic information retrieval have pushed the boundaries of what is possible with spatial search. Up until this point though, the majority of research has focused solely on the spatial or geospatial dimension of search ignoring additional and often highly influential dimension such as *time* and *theme*. Research should continue to pull on this thread of dimensionally-enhanced spatial search, asking how and to what degree concepts such as time and topic influence search and the way we interpret search results.

Take for example the task of "checking in," a concept made popular by the transportation industry, but more recently reflecting the act of publishing your platial location through an online social networking application. Computationally, *checking in* often involves taking a spatial location (set of geographic coordinates) obtained through a location-enable mobile device, and determining the platial location of the user controlling the device. In many ways this is an example of *spatial search* in its most basic form. Given a set of coordinates and a gazetteer of point of interests (POI), return a set of places ordered by Euclidean distance from closest to furthest away. Current *check-in* applications simply assume the user is at the closest place, returns the name of the venue, and the process is complete. While this works quite well in many cases, it makes the erroneous assumption that all places have the same probability of being visited at any time of day and day of the week. This methods implies that the only property of a place influencing an individual's decision to visit, is its spatial location.

In actuality, time plays a significant role in spatial search, and in particular this example of *checking in*. The probability of checking in to a *Night Club* at 8am on a Tuesday is much less than say a *Coffee Shop* and the opposite is true for Saturday at 1am [1]. There are specific temporal patterns and behaviors that are linked to human activities. The places that we chose to visit are often defined by the activities that the places afford. In many ways this is a logical extension of the socio-institutional affordances described by Raubal et al. [2]. The human construct of *Categories* has arisen from this need to group places by the allowed activities.

Leaving *Time* aside, additional dimensions such as *Theme* or *Topic* play an important role in both how spatial data is queried as well as how a resulting data set is interpreted. Let us explore another example specific to user-generated geo-content, but this time from a spatio-thematic standpoint. In many ways the value of *big data* lies in its variety. Large POI datasets generated through user contributed means (e.g., Yelp, Foursquare, Geonames) come with bias imposed by both the collection platform as well as the users contributing the data. By combining multiple datasets, much of the bias is reduced while increasing the breadth of content. For example, descriptive information for a specific POI such as price, noise level and wheel-chair accessibility may be present in the Yelp representation of the data, while number of check-ins, tips and rating may be contained in the Foursquare representation of the same POI. The results of a *Spatial Search* on a merged/conflated dataset would be of greater value than individual datasets alone.

Matching and conflating this data is an important, yet difficult task. In order to accomplish this task, dimensions other than the geospatial dimension must be explored. Simply matching user-generated POI between providers based solely on geographic coordinates has been shown to be less than adequate [3]. Inclusion of additional attribute dimensions is essential for increasing the accuracy of POI matching. By accessing thematic attributes of the data such as tips and reviews, POI can be compared in *topic* space based on the words, phrases and content contributed by individual users. Matching based on multiple dimensions (such as topics based on reviews and categories based on temporal check-ins) has been shown to dramatically increase the accuracy of POI matching [3].

In summary, this position paper takes the stance that *Spatial Search* is not only about *spatial.* In many cases dimensions outside of X, Y and Z can and should be employed to enhance spatial search. A discussion concerning the role that other dimensions play in searching inherently spatial data would be of considerable interest from my perspective.

References

[1] McKenzie, G., Janowicz, K., Gao, S., Yang, J. A., & Hu, Y. (2014) POI Pulse: A Multi-Granular, Semantic Signatures-Based Approach for the Interactive Visualization of Big Geosocial Data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, The University of Toronto Press

[2] Raubal, M., Miller, H. J., & Bridwell, S. (2004). User-Centred Time Geography for Location-Based Services. *Geografiska Annaler: Series B, Human Geography 86*(4): 245–265.

[3] McKenzie, G., Janowicz, K., Adams, B. (2014) A Weighted Multi-Attribute Method for Matching User-Generated Points of Interest. *Cartography and Geographic Information Science* 41(2): 125–137, Taylor & Francis.