Mapping Location-Based Social Topics in Today's Evolving Cities

Andrew Jenkins

Abstract Today's modern cities present diverse landscapes of socially trending topics that often have short life spans and evolve at various urban scales. Social topics can occur unbeknown to those living in the periphery or unplugged from social channels. These topics (i.e. sports, politics, nightlife, etc.) often appear as single snapshots in time and lack assertive descriptions that often preclude entry in authoritative databases, and overlooked by local news media. The proliferations in location-based services, social media, and Web 2.0 technologies present new opportunities to explore these social topics with greater spatial fidelity. However, these advancements have drawbacks highlighted by current literature calling for new processes for harvesting and manipulating this data for scientific research and real-world applications. The geospatial community currently lacks a standardized framework for exploring, analyzing, and mapping these "geosocial" datasets. This project introduces a (novel) framework to harvest geolocated information for venues from diverse location-based social networking (LBSN) services, and analyze this information in order to support data integration, information aggregation, and knowledge discovery. By doing so, we introduce a new approach to conflation, focusing on the social media presence of these enterprises, and linking their various manifestations into a single common reference social object. Accordingly, the thesis hypothesis can be formulated as: the various references to venues in diverse LBSN services can be aggregated into individual common reference social objects through the analysis of their geolocation information and accompanying content. Data integration issues and concordance between location-based social networking sites like Twitter, Yelp, and Foursquare along with volunteered geographic information from OpenStreetMap are addressed at the venue level to account for positional differences. A topic modeling algorithm is applied to find latent terms within short messages, which are clustered into time groups and visualized using temporal map tools. This work is significant in attempting to understand the combined benefits of integrating volunteered geographic information, social media, and location-based services to model information flow, data convergence, or conducting ambient geospatial analysis.