Title: Supporting Advanced Interactive Search Using Inverted Index

Speaker: Mr Zheng Yuxin
Date/Time: 24 November 2015, Tuesday, 03:30 PM to 05:00 PM
Venue: MR1, COM1-03-19
Supervisor: Dr Anthony Tung, Associate Professor, School of Computing

Abstract:

With the proliferation of geographic applications, a large amount of geographic data have been created, such as the geo-textual data and geotagged images. Meanwhile, location-based services are developed to provide people with information of their surrounding environments. Two types of location-based queries are developed for the geo-textual data and the geotagged images respectively. Recent study on spatial keyword search focused on the processing of spatial keyword queries which retrieve objects that match particular keywords within a spatial region. Besides, spatial image search is proposed to find similar images and information about the query image. However, such a huge amount of geographic data also pose a question: how can we find the exact information we want? Interactive methods have been studied widely to find the right information based on human-computer interaction. In this thesis, we focus on the location-based services and study the advanced interactive methods in these services using inverted index.

We propose a general framework to process the location-based keyword queries interactively. The proposed framework adopts a unifying strategy for processing different variants of spatial keyword queries. We adopt the autocompletion paradigm that generates the initial query as a prefix matching query. If there are few matching results, other variants of spatial keyword search are performed as a form of relaxation that reuses the processing performed in the earlier phase. The types of relaxation allowed include spatial region expansion and exact/approximate prefix/substring matching. Moreover, since the autocompletion paradigm allows appending characters after the initial query, we look at how query processing performed for the initial query and relaxation can be reused in such instances. Compared to existing work which processes variants of spatial keyword query as new queries over different indexes, our approach offers a more compelling way to efficient and effective spatial keyword search.

For the location-based image search, it can be modeled as the nearest neighbor search in high-dimensional spaces. Due to the "curse of dimensionality" problem, it is very expensive
to process the nearest neighbor (NN) query in high-dimensional spaces; and hence, approximate approaches, such as Locality-Sensitive Hashing (LSH), are widely used for their theoretical guarantees and empirical performance. Current LSH-based approaches target at the L1 and L2 spaces, while, as shown in the existing work, the fractional distance metrics (Lp metrics with $0 < p < 1$) can provide more insightful results than the usual L1 and L2 metrics for data mining and multimedia applications. However, none of the existing work can support multiple fractional distance metrics using one index. We propose LazyLSH that answers approximate nearest neighbor queries for multiple Lp metrics. Different from previous LSH approaches which need to build one dedicated index for every query space, LazyLSH uses a single base index to support the computations in multiple Lp spaces, thereby significantly reducing the index maintenance overhead. LazyLSH can provide more accurate results for approximate kNN search under fractional distance metrics.

Finally, we design and develop an interactive augmented reality system for shopping, called ARShop, to provide the above location-based services. The main objective of our system is to enhance users' shopping experience. A user can input a shopping list and issue a location-based query using a photo of his/her surrounding environment. Based on the shopping list and the user-uploaded photo, the ARShop system can direct the user to shops that contain items in his/her shopping list.