

A new search strategy for m-closest keywords query by using dynamically-created grid partitioning over spatial datasets

Yuan Qiu*, Hui Zhai*, Dang Hoang Anh**, Tadashi Ohmori*, Hideyuki Fujita* and Takahiko Shintani*

* Graduate School of Information System, The University of Electro-Communications, Japan

** NTT Data, Japan

Abstract. In current search engine researches, new abilities of spatial keyword queries are significant. We focus on the issue of the m-closest keywords (mCK) query, which is one of spatial-keyword queries over spatial data in the Web. The mCK query is a problem to find the optimal set of records in the sense that they are the spatially-closest records that satisfy m user-given keywords. This query is used to identify an optimal spatial location under given m keywords. The mCK query was originally proposed by Zhang et al. [ICDE2009]. They assumed one specialized R*-tree in the data-server side and proposed an Apriori-style enumeration of MBRs. However, in practice, we cannot expect many existing spatial Web-data servers such as Flickr or Twitter to maintain all data in such an R-tree. Furthermore we want to issue mCK queries over heterogeneous data sets.

To overcome these limits, our study assumes that we create a grid partitioning from necessary data only when a mCK query is given, and we propose a new search-strategy termed Diameter Candidate Check for the mCK query over such dynamically-created grid partitioning. In the poster presentation, we will demonstrate our system over web data from Flickr, with sophistication for getting diversified top-K answers for the mCK query problem.

Keywords. spatial database, spatial keyword query, mCK query