Top-k Indexes Made Small and Sweet

Yufei Tao

University of Queensland, Brisbane, Queensland, Australia
`taoyf@itee.uq.edu.au`

──── **Abstract** ────

Top-k queries have become extremely popular in the database community. Such a query, which is issued on a set of elements each carrying a real-valued *weight*, returns the k elements with the highest weights among all the elements that satisfy a predicate. As usual, an index structure is necessary to answer a query substantially faster than accessing the whole input set.

The existing research on top-k queries can be classified in two categories. The first one, which is system-oriented, aims to devise indexes that are simple to understand and easy to implement. These indexes, typically designed with heuristics, are reasonably fast in practical applications, but do not necessarily offer strong performance guarantees – in other words, they are *small* but not *sweet*. The other category, which is theory-oriented, aims to develop indexes that promise attractive bounds on the space consumption and query overhead (sometimes also update cost). These indexes, unfortunately, are often excessively sophisticated in the adopted techniques, and are rarely applied in practice – they are *sweet* but not *small*.

This talk will discuss the progress of an on-going project that strives to take down the barrier between the two categories, by crafting a framework for acquiring simple top-k indexes with excellent performance guarantees – namely, small and sweet. This is achieved with reductions that produce top-k indexes automatically from the existing data structures for conventional reporting queries on unweighted elements (i.e., finding all elements satisfying a predicate), and/or the existing data structures on top-1 queries. Our reductions promise nearly no performance deterioration with respect to those existing structures, are general enough to be applicable to a huge variety of top-k problems, and work in both the external memory model and the RAM model.