

Logical approaches to query optimization over web-based sources

Michael Benedikt

Department of Computer Science, University of Oxford
`michael.benedikt@cs.ox.ac.uk`

Abstract. Much of the data needed to answer user queries resides on the web – accessible via web apis or web forms. The setting is answering queries over sources that are connected to one another by semantic relationships (that is, integrity constraints). The sources may have overlap, may have multiple means of access with different requirements (e.g. number of required arguments), and the means of access within and across relations may differ radically in cost. This work presents a method for generating query plans that run on top of these sources, obeying the access patterns and exploiting the integrity constraints. The approach is based on “implementing” interpolation and definability results in predicate logic, while extending them to take into account access restrictions, modifying them work in the presence of background integrity constraints, combining them with cost considerations. In this talk I will give a flavor of the theory behind the logic-based approach, and then discuss a variety of algorithmic and system issues. This is joint work with Balder ten Cate, Julien Leblay, and Efi Tsamoura.