

Querying Graphs with Data

Libkin, L., Martens, W., & Vrgoc, D. (2016). Querying graphs with data. *Journal of the ACM (JACM)*, 63(2), 1-53. <10.1145/2850413> Accessed 29 Jan 2021.

Abstract

Graph databases have received much attention as of late due to numerous applications in which data is naturally viewed as a graph; these include social networks, RDF and the Semantic Web, biological databases, and many others. There are many proposals for query languages for graph databases that mainly fall into two categories. One views graphs as a particular kind of relational data and uses traditional relational mechanisms for querying. The other concentrates on querying the topology of the graph. These approaches, however, lack the ability to combine data and topology, which would allow queries asking how data changes along paths and patterns enveloping it. In this article, we present a comprehensive study of languages that enable such combination of data and topology querying. These languages come in two flavors. The first follows the standard approach of path queries, which specify how labels of edges change along a path, but now we extend them with ways of specifying how both labels and data change. From the complexity point of view, the right type of formalisms are subclasses of register automata. These, however, are not well suited for querying. To overcome this, we develop several types of extended regular expressions to specify paths with data and study their querying power and complexity. The second approach adopts the popular XML language XPath and extends it from XML documents to graphs. Depending on the exact set of allowed features, we have a family of languages, and our study shows that it includes efficient and highly expressive formalisms for querying both the structure of the data and the data itself..