An Algorithm for Extracting Schemas from External Memory Graphs
グラフデータからスキーマを抽出するための外部記憶アルゴリズム

Student No.: 201621621
Name: Yoshiki SEKINE

In recent years, graph data has been widely used and various kinds of new graph data is actively being created. In contrast to other databases such as relational databases and XML, most of graphs do not have their own schemas. Therefore, in many cases we cannot make use of schema to manage graphs effectively. Here, if we can extract a schema from a graph efficiently, we can take advantage of the extracted schema for query optimization, structure browsing, query formulation, and so on. However, extracting schemas from large graphs are difficult due to the following reasons. Firstly, most of schema extraction algorithms proposed so far are in-memory algorithms and thus cannot deal with large graphs that do not fit in main memory. Secondly, schema extraction is a complex and time-consuming task. In particular, the utility function, which is a popular function used in schema extraction, requires a large amount of computation cost as the number of unique edge labels in a graph becomes larger. To address these problems, we propose a novel schema extraction algorithm for large graphs. This is designed as an external memory algorithm using parallel processing and our novel utility function. Our utility function is designed so that less computation cost is required while schemas are extracted as appropriately as the original utility function.

We implemented our algorithm in Ruby and made evaluation experiments. The results suggest that our algorithm can extract schemas from graphs more efficiently and appropriately than the algorithm using the previous utility function, and that the parallelization of the class extraction makes the execution time faster for a real-world graph with a large number of unique edge labels.

Academic Advisors: Principal: Nobutaka SUZUKI
Secondary: Atsuyuki MORISHIMA