

発表概要

# 要約情報を用いた SPARQL クエリ分散処理における 集約計算高速化手法の検討

金子 舟<sup>1,a)</sup> 高野 保真<sup>2,b)</sup> 千代 英一郎<sup>1,c)</sup>

2017年11月1日発表

Resource Description Framework (RDF) は主語, 述語, 目的語の 3 つ組で情報を統一的に表現し, SPARQL と呼ばれる標準クエリ言語を用いることでマッチする情報を引き出すことができる. 大規模な RDF データに対しては, SPARQL クエリの処理効率が実用上の問題となるため, 複数のサーバを用いた分散処理手法が広く研究されている. その手法の 1 つに, RDF データの要約情報を利用したクエリの分散処理手法が提案されている. この手法では, 値や述語などを基準に RDF データをグループとしてまとめた要約情報を利用し, 与えられたクエリから解の範囲の絞り込みや負荷の分散を考慮して, 検索効率の向上を目指している. 分散時のクエリ処理効率が向上することがすでに確認されているものの, 各サーバで分散して求めた結果を集約する方法が明らかではなく, 一部の集約計算等を含むクエリを分散検索できなかった. そこで本発表では, クエリの解析時に集約計算を考慮した変換処理を行うことで, 各結果集合の統合時に集計処理を行う手法を提案する. SPARQL クエリ言語の定義に基づいて, 分散処理時の変換規則を定義する. 変換時には, RDF データの分割方法についても考慮する必要がある. また, 本手法を用いて集約関数を含むクエリによる集計処理を行った場合の実行結果については実データを用いて評価した.

## Improving the Efficiency of Aggregate Operations in Distributed SPARQL Query Processing Using Summary Information

SHU KANEKO<sup>1,a)</sup> YASUNAO TAKANO<sup>2,b)</sup> EIICHIRO CHISHIRO<sup>1,c)</sup>

Presented: November 1, 2017

The Resource Description Framework (RDF) is used as a general method for modeling of information in the form of subject-predicate-object expressions, known as triples. RDF triples can be extracted matching information by using the standard query language SPARQL. Since the efficiency of SPARQL query processing is practically important for large-scale RDF graphs, many studies focus distributed SPARQL query processing. In such a situation, we have proposed a method for distributed query processing using summary information of RDF data. The summary information is a grouping of RDF data on the basis of predicates and it can be used to limit the range of RDF data to be searched. Although it has already been confirmed that the summary information contribute to the efficiency of distributed query processing, it is not clear how to integrate the results of queries including some aggregate operations. In this presentation, we propose a method to enable the summary-based distributed query processing to integrate aggregation operations. The method defines conversion rules for aggregate operations in distributed processing based on the specification of the SPARQL query language. We also show a result of processing queries including aggregate operations with real data.

<sup>1</sup> 成蹊大学大学院理工学研究科  
Graduate School of Science and Technology, Seikei University,  
Musashino, Tokyo 180-8633, Japan

<sup>2</sup> 成蹊大学理工学部  
Faculty of Science and Technology, Seikei University,  
Musashino, Tokyo 180-8633, Japan

a) dm166202@cc.seikei.ac.jp

b) yasunao-takano@st.seikei.ac.jp

c) chishiro@st.seikei.ac.jp