Faculty of Sciences

Stream join processing in RDF mapping engines

by

Sitt Min Oo
Student number: 01503244

Supervisor: Prof. Dr. Ruben Verborgh
Supervisor: Dr. Anastasia Dimou
Counsellor: Gerald Haesendonck

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Computer Science

Academic Year 2020–2021

Preface

Acknowledgement

Sitt Min Oo, December 2020

Stream join processing in RDF mapping engines

by

Sitt Min Oo

Master's dissertation submitted in order to obtain the academic degree of
Master of Science in Computer Science

Academic year 2020–2021

Supervisor: Prof. Dr. Ir. Bjorn De Sutter
Counsellor: Dr. Tim Besard

Faculty of Sciences
University Ghent

Abstract

Here comes abstract.

Keywords

RDF, RMLStreamer, RML, Adaptive windows, Stream joins.

Contents

Chapter 1

Introduction

A large volume of data is generated daily on the web in a variety of domains. These data are often structured according to an organization's specific needs or formats: Leading to a difficulty in integrating the data across the different applications. These generated data might have to be associated with archival data, also of heterogeneous formats, to provide a coherent view required by analysis tasks. Heterogeneous web data formats, such as CSV or HTML, are not explicitly defined to enable linking entities in one document to other related entities in external documents. Based on W3C standard, semantic data formats such as RDF triples [1], are a solution to this particular problem by enriching the data with knowledge and associations across different domains, through the usage of common ontologies. RDF triples also form the basic building blocks of knowledge graphs. Knowledge graphs are extensively used in social networks like Facebook[2] and especially with Google's search engine[3], it enables machines to understand the data and perform complex automated processing on the data. Therefore, there is a need to transform non-RDF data to RDF compliant format.

There have been recent state-of-the-art techniques to solve this task of consolidating heterogeneous data and transforming them to a RDF compliant format. In this thesis, we will focus on one such format called RDF triples. These RDF processing engines are

1.1 Terminology and definitions

Een sectie [4].

1.2 Nog een sectie

Nog een.

1.2.1 Een subsectie

Bibliography

[1] E. Miller, "An introduction to the resource description framework," *Bulletin of the American Society for Information Science and Technology*, vol. 25, no. 1, pp. 15–19, 1998. DOI: https://doi.org/10.1002/bult.105. eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/bult.105. [Online]. Available: https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/bult.105.

[2] J. Weaver and P. Tarjan, "Facebook linked data via the graph api," *Semantic Web*, vol. 4, pp. 245–250, 2013.

[3] A. Singhal. (2012). "Introducing the knowledge graph: Things, not strings," [Online]. Available: https://blog.google/products/search/introducing-knowledge-graph-things-not/ (visited on 12/25/2020).

[4] D. Phuoc, H. Nguyen Mau Quoc, H. Ngo, T. Nhat, and M. Hauswirth, "The graph of things: A step towards the live knowledge graph of connected things," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 37, pp. 25–35, Mar. 2016. DOI: 10.1016/j.websem.2016.02.003.