TECHNISCHE UNIVERSITÄT DRESDEN

FACULTY OF COMPUTER SCIENCE
INTERNATIONAL CENTER FOR COMPUTATIONAL LOGIC

Master Thesis

Master Computational Logic

Translating Natural Language to SPARQL

Xiaoyu Yin
(Born 13. June 1994 in Zhumadian, Mat.-No.: 4572954)

Supervisor: Dr. Dagmar Gromann

Dresden, September 7, 2018

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die von mir am heutigen Tag dem Prüfungsausschuss der Fakultät Informatik eingereichte Arbeit zum Thema:

Translating Natural Language to SPARQL

vollkommen selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Dresden, den September 7, 2018

Xiaoyu Yin

Abstract

English abstract here

Contents

1 Introduction

This chapter provides an introduction of the thesis. We introduce the motivation of this thesis in section 1.1, and the outline of this thesis in section 1.2.

1.1 Motivation

In recent years, numerous documents have been published onto the web in a fast-growing speed, forming a global information space which worldwide people have access to.

1.2 Thesis Outline

Chapter 2 presents the background of this thesis.

Chapter 3 describes the research method used in this thesis.

Chapter 4 shows the experiments conducted during this thesis.

Chapter 5 depicts the analysis derived from the experiment results.

Chapter 6 concludes the whole thesis and provides potential directions for the future work.

2 Background

This chapter introduces the background knowledge involved in this thesis. Semantic Web Technologies is briefly introduced in section 2.1, including the notion of Linked Data in section 2.1.1 and SPARQL in section 2.1.2. Section 2.2 introduces the field of neural machine translation. Related research is discussed in section 2.3.

2.1 Semantic Web Technologies

2.1.1 Linked Data

2.1.2 SPARQL

2.2 Neural Machine Translation

2.2.1 Sequence to Sequence Learning

2.2.2 Recurrent Neural Network

2.2.3 Long Short-Term Memory

2.3 Related Work

[CXY+17] proposed an enhanced encoder-decoder framework for the task of translating natural language to SQL, a similar query language with SPARQL but targeting structured databases instead of knowledge bases. They used not only bleu, but also query accuracy, tuple recall, and tuple precision for measuring the quality of output queries, and achieved good results.

[SMM+18, SMV+18] proposed a generator-learner-interpreter architecture, namely Neural SPARQL Machines to translate any natural language expression to encoded forms of SPARQL queries. They designed templates with variables that can be filled with instances from certain kinds of concepts in target

knowledge base and generated pairs of natural language expression and SPARQL query accordingly. After encoding operators, brackets, and URIs contained in original SPARQL queries, the pairs were fed into a sequence2sequence learner model as the training data. The model was able to predict on unseen natural language sentence, and generate encoding sequence of SPARQL for interpreter to decode.

3 Methodology

3.1 Model

3.2 Evaluation

4 Experiments

4.1 Implementation

4.2 Results

5 Analysis

6 Conclusion

Bibliography

[CXY+17] Ruichu Cai, Boyan Xu, Xiaoyan Yang, Zhenjie Zhang, Zijian Li, and Zhihao Liang. An Encoder-Decoder Framework Translating Natural Language to Database Queries. 2017.

[SMM+18] Tommaso Soru, Edgard Marx, Diego Moussallem, Gustavo Publio, André Valdestilhas, Diego Esteves, and Ciro Baron Neto. SPARQL as a foreign language. In *CEUR Workshop Proceedings*, volume 2044, aug 2018.

[SMV+18] Tommaso Soru, Edgard Marx, André Valdestilhas, Diego Esteves, Diego Moussallem, and Gustavo Publio. Neural Machine Translation for Query Construction and Composition. 2018.

List of Figures

List of Tables

Acknowledgments

Acknowledgments here blablabla

Copyright Information