IE chowdhury2010IR D41D8CD98F00B204E9800998ECF8427E

Gottfried Wilhelm Leibniz Universität Hannover
Fakultät für Elektrotechnik und Informatik

Information extraction from arcticles on the impacts of COVID-19 lockdowns on air quality

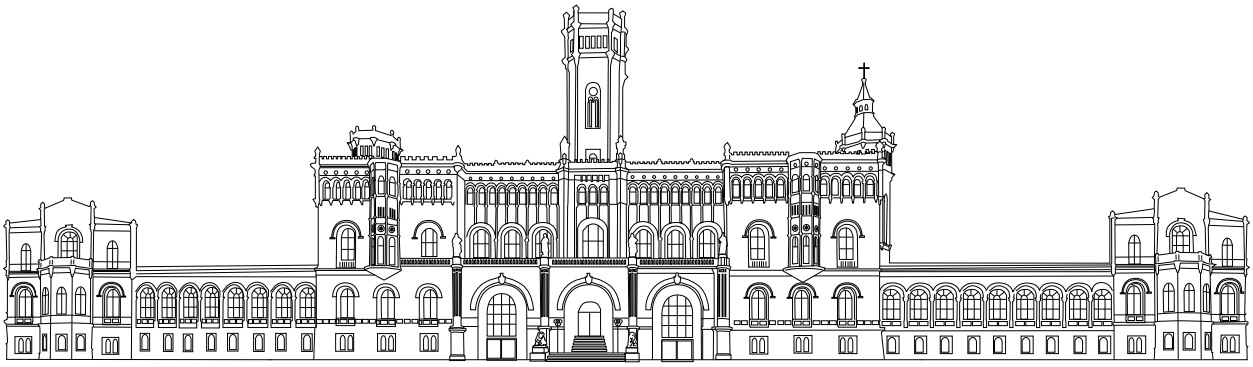
A thesis submitted in fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science

BY

Quentin Münch
Matriculation number: 10031323
E-mail: quentin.muench@stud.uni-hannover.de

First evaluator: Prof. Dr. Sören Auer
Second evaluator: Dr. Jennifer D'Souza
Supervisor: Dr. Markus Stocker

31.08.2022



Declaration of Authorship

I, Quentin Münch, declare that this thesis titled, 'Information extraction from articles on the impact of COVID-19 lockdowns on air quality' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

NAME

Signature: _____

Date: _____

OPTIONAL

Acknowledgements

XXXXX

Abstract

Your Abstract. Clearly motivate your work (WHY), state what is your problem (WHAT), and describe your solution (HOW). Also, explain how your solution was evaluated (either empirically or formally) and summarized the observed results
Keywords: KW1, KW2, KWn

Contents

List of Figures

List of Tables

Acronyms

GaV Global-as-View

GLaV Global-Local-as-View

LaV Local-as-View

LSLOD Life Science Linked Open Data

QEP Query Execution Plan

RDF Resource Description Framework

RDF-MT RDF Molecule Template

RDFS RDF Schema

SDL Semantic Data Lake

SSQ star-shaped sub-query

URI Universal Resource Identifier

Chapter 1

Introduction

The outbreak of the 2019 coronavirus disease (COVID-19) greatly impacted the entire world. Due to the virus being very contagious, hygiene became a lot more important and social interactions had to be reduced to a minimum. While the precise measures to reduce the spread of the virus differed from country to country, lockdowns were introduced in almost every part of the world. People were not allowed to go outside unless it was essential for their living. Companies were told to let the employees work from home, so they didn't need to leave the house. Clubs, bars and restaurants had to close down because of government orders as well as a lack of customers. Considering these and many other points, it is apparent that people were not driving around as much as before the pandemic. Rather they were staying at home, thus considerably reducing their transportation emissions.

Many scientists saw this as a unique opportunity to research how much of an impact such lockdowns have on air quality. While daily life was put on hold because of the restrictions, the streets of major cities were emptier than ever. Many articles regarding air quality have been published, however not all followed scientific methods and reviewing practices. In order to see all the data in one place, the research center Jülich accumulated every article pertaining to the change in air quality during lockdowns. The examined articles are all scientifically verified[**base**]. These articles contain a great amount of data, which has to be processed manually. The research center Jülich processed each article individually and collected the air quality data. They set up a website which presents the results in various charts.

However, the process of manually having to check each research article proves to be very inefficient. Therefore, models for automated information extraction of such air quality data points could provide a fitting solution. Since scientists are still conducting new research regarding the connection between lockdowns and air quality,

these models would also be useful for later database updates. By automatically extracting the data, the time spent skimming through the papers will be significantly reduced. On the contrary, the technology is errorprone. It is not expected to find every single data point that exists in the text. However, at the very least, it will provide a solid baseline for further examination.

Chapter 2

Background

This chapter introduces the main topics needed to understand the development of this thesis.

There are huge amounts of text data on the internet. In fact, there is so much of it, that no human can possibly ever read and understand everything. That is why we try to use the computer to help us guide through the data.
There are several strategies for establishing order across texts, the most common being information retrieval (IR), information filtering (IF) and information extraction (IE)[**IE**]. Information retrieval concerns itself with all the activities related to the organization of, processing of, and access to, information of all forms and formats. It can also be seen as a document retrieval system, since it is designed to retrieve information about the existence of documents relevant to a user query[**chowdhury2010IR**]. Information filtering aims to remove irrelevant data from incoming streams of data items [**hanani2001IF**]. Information Extraction refers to the automatic extraction of structured information such as entities or relations from unstructured sources[**sarawagi2008IE**]. In contrast to IR systems, IE systems need to extract facts from the documents itself. The extracted data is often used for filling in databases, which are then available for various applications to further process the data[**klgl2014IE**]. Since information is often spread across multiple sentences, understanding natural language is fundamental to IE[**IETaC**].

Chapter 3

Related Work

Topics related to this thesis have been extensively treated in the literature. This chapter presents an overview of what has been done

Chapter 4

Approach

This chapter states the problem statement and proposed solution

Chapter 5

Implementation

This section presents your implementation

Chapter 6

Experimental Evaluation

The experimental evaluation is reported in this section. Please, include your research questions.

The research questions addressed by this thesis are: **RQ1)** YYY **RQ2)** XXX **RQ3)** TT **RQ4)** OPP

The remainder of this chapter is structured as follows: First, the used benchmark is described. Second, the data preparation is presented. Afterwards, the setup of the experiment is depicted. Finally, the results are shown and analyzed.

Benchmark:

Metrics:

Implementations:

Chapter 7

Conclusions and Future Work

This chapter presents the lessons learned and future work