# FOM Hochschule für Oekonomie & Management

## university location Bonn

# Exposé

in the study course Wirtschaftsinformatik

on the subject

## Development of a Query Language for Full-Text Search in Relational Databases

by

## Sebastian Bunge

| | |
|---|---|
| Advisor: | Prof. Dr. Peter Steininger |
| Matriculation Number: | 539441 |
| Submission: | August 1, 2022 |

# Contents

# 1 Problem of the Thesis and Scientific Question

Current relational databases go far beyond the transactional processing of data with their functionalities. One of these functionalities is the possibility of a full-text search over data. For this full-text search, powerful extensions of the database query language (SQL) are available, which generate complex query terms by an extensive combination of operators. The goal of this thesis is to generate query terms for the database based on a search engine-like input. The starting point for the input can be e.g. the search operators of Google.

The scientific question may be defined as: In what way can a software component be developed using a custom query language and interpreter to enable full-text search on Microsoft SQL servers?

# 2 Initial Situation and Classification in the Existing Research

The thesis can be classified into the scientific areas of interpreter construction and full-text search.

The topic of full-text search, with possibilities such as word and phrase-based search and the ability to index documents in their native form, such as office documents and pdfs, is an increasingly important feature of modern databases and document management systems and thus of great relevance. Making such a powerful tool, through easy-to-use search engines and grammar, as easily accessible as possible, without requiring prior knowledge of SQL, allows it to be used by individuals and companies with little IT knowledge. Just one example of why domain-specific languages and interpreters play an important role in today's IT world.

There are already many examples of how such custom languages and interpreters make writing code more accessible or support experienced software engineers in their development. Examples in the context of full-text search are search engines, which generate SQL statements from search queries.

# 3 Approach and Methodology

To achieve the objective of my thesis I will refer to literature that defines concepts of full-text search and the development of interpreters. Based on this I use the method of prototypical implementation and try to apply these concepts practically.

# 4 Outline

1. Introduction

2. Theory

   a) Elaboration of the language elements for full-text search in common relational databases (Microsoft SQL Server)

   b) How to build interpreters

3. Approach

   a) Definition of grammar for full-text search input and schematic description using the extended Backus-Naur form

   b) Conversion of the grammar into a software component in a web application

4. Source Code

5. Examples and Usage

6. Conclusion

# 5 Literature

The core of my literature will be "Pro full-text search in SQL Server 2008" by Michael Coles and "Crafting interpreters" by Robert Nystrom, which define the most important elements of full-text search and interpreter development. In addition, I will refer to articles from projects similar to mine and review their approach and sources for my purposes. Articles I have found recently on this topic are several projects by Holger and Hannah Bast, who have developed tools to assist in full-text search.

# 6 Working Title

The title of my thesis is "Development of a Query Language for Full-Text Search in Relational Databases".

# 7 Objective / Expected Result

The objective of my thesis should be an artifact that allows you to define certain search criteria in a search box on a website. These search criteria should enable functions of the full-text search, e.g. reflections and phrase-based search. From this, appropriate SQL code is generated and applied to a Microsoft SQL server. As a result, tuples of server content are to be displayed that match the search query, e.g. excerpts from Word documents or pdfs.

# 8 Project / Time Schedule

Schedule here