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 19, 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 'To what extent can a software component be developed using a custom query language and code generator 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 code generators 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 will make use of a prototypical implementation and try to apply these concepts practically. A software prototype is defined as an executable model with essential properties of the target system, which is the basis for the system specification. The model reproduces essential properties of the planned system in a descriptive and easily modifiable form. The prototype should represent the functionality much more effectively than textual descriptions or purely static models. (cf. Connell, Shafer 1989) The implementation of a prototype is an iterative process: Implementation followed by evaluating and adapting to new problems and specifications. In each case, as much functionality is implemented as is necessary to verify the targeted phase result. (cf. Pomberger, Pree, Stritzinger 1992, p. 3)

4 Outline

The following is a draft of the structure of the thesis, which is subject to change.

1. Abstract
2. Theory

    - What is Full-Text Search
    - Full-Text Search on MS SQL Server
    - How to design a language
        - DSL
        - EBNF
    - How to build a code generator

3. Implementation
4. Source Code
5. Examples and Demonstration
6. Summary

5 Literature

The core of my literature will be "Pro full-text search in SQL Server 2008" by Michael Coles (Coles, Cotter 2009) and "Crafting interpreters" by Robert Nystrom (Nystrom 2021), 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 H. Bast, who has developed tools to assist in full-text search. (Bast, Buchhold 2013)(Bast, Bäurle, et al. 2013)

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 is a prototype 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

August 3rd - September 2nd | Literature research and writing of the theoretical part
September 3rd - October 2nd | Developing the grammar and prototype
October 3rd - October 16th | Writing down the development process and demonstration
October 17th - November 3rd | Buffer and Proofreading

Bibliography

BAST, Hannah; BÄURLE, Florian, et al.: *Broccoli: Semantic Full-Text Search at your Fingertips*. arXiv:1207.2615 [cs]. July 2013. URL: http://arxiv.org/abs/1207.2615 (visited on 07/28/2022)

BAST, Hannah; BUCHHOLD, Björn: An Index for Efficient Semantic Full-Text Search. In: *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. CIKM '13. event-place: San Francisco, California, USA. New York, NY, USA: Association for Computing Machinery, 2013, pp. 369–378. ISBN: 978-1-4503-2263-8. DOI: 10.1145/2505515.2505689. URL: https://doi.org/10.1145/2505515.2505689

COLES, Michael; COTTER, Hilary: *Pro full-text search in SQL Server 2008*. The expert's voice in SQL server. Berkeley, CA: Apress, 2009. ISBN: 978-1-4302-1594-3

CONNELL, John L.; SHAFER, Linda: *Structured rapid prototyping: an evolutionary approach to software development*. Yourdon Press computing series. Englewood Cliffs, N.J: Yourdon Press, 1989. ISBN: 978-0-13-853573-5

NYSTROM, Robert: *Crafting interpreters*. eng. Daryaganj Delhi: Genever Benning, 2021. ISBN: 978-0-9905829-3-9

POMBERGER, Gustav; PREE, Wolfgang; STRITZINGER, Alois: Methoden und Werkzeuge für das Prototyping und ihre Integration. In: *Inform., Forsch. Entwickl.* 7 (1992) Nr. 2, pp. 49–61