Informatics and Technology in Clinical Care and Public Health
J. Mantas et al. (Eds.)

doi:10.3233/SHTI210939

A Multilingual Browser Platform for Medical Subject Headings

Raphael SCHEIBLE[a,b,1], Philip STRECKER[b], Suhail YAZIJY[b],
Fabian THOMCZYK[b], Ruslan TALPA[c], Andrea PUHL[d] and Martin BOEKER[a,b]
[a] *Institute of Artificial Intelligence and Informatics in Medicine, Chair of Medical Informatics, University Hospital rechts der Isar, Technical University of Munich, Munich, Germany*
[b] *Institute of Medical Biometry and Statistics, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany*
[c] *subZero LLC, Oradea, Romania*
[d] *Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany*

Abstract. The National Library of Medicine (NLM) controls and publishes the thesaurus Medical Subject Headings which is used for indexing PubMed. Besides an XML export, the NLM offers a web based MeSH browser. The platform contains English terms. The German Institute for Medical Documentation and Information (DIMDI) partially translated and published these terms. Recently, the German National Library of Medicine (ZB-MED) overtook the translation of MeSH. However, there is no dedicated platform which focuses on MeSH and covers multiple languages. Here, we address this gap, by offering a modern multilingual searchable MeSH browser. A modular platform using open source technology is presented. The frontend enables the user to search and browse terms and switch between different languages. The current version of the presented MeSH browser contains English and German MeSH terms and can be accessed at https://mesh-browser.de.

Keywords. MeSH-Browser, Web Service, Databases, Medical Subject Headings

1. Introduction

Medical Subject Headings (MeSH) is a hierarchical thesaurus for medical terminology which is controlled and published by the National Library of Medicine (NLM). It is used for indexing PubMed and annually released in XML format. For public access, the NLM offers a web-based MeSH Browser for English terms. The German Institute for Medical Documentation and Information (DIMDI) has been translating MeSH for many years, revising it annually and adding new terms. However, after the National Library of Medicine (NLM) stopped providing the translation system for MeSH in 2019, DIMDI halted its translation activities after releasing their translated 2019 version of MeSH. Mid 2020, the German National Library of Medicine (ZB-MED) decided to take on the MeSH translation task with a semi-automated approach [1]. Nevertheless, new datasets were not published until now, making the aforementioned DIMDI 2019 version the most

[1] Corresponding author, AIIM, Klinikum rechts der Isar der Technischen Universität München, Ismaninger Str. 22, 81675 München, Germany; E-mail: raphael.scheible@tum.de.

recent and reliable German MeSH data to work with. The objective of this work is to present a modern multilingual MeSH browser, fully implemented with open-source technology, based on the DIMDI 2019 dataset in combination with the 2021 English version of MeSH.

2. Methods

2.1. System Design

PostgreSQL has been selected as database software because it is designed to handle a wide range of workloads, from single machines to large-scale web services, while natively supporting JSON. To store the data of MeSH exports, we modeled a multilingual database layout. With PostgREST [2] we added a RESTful API on top of the database. Dedicated views and procedures, which access multiple internal data tables, were exposed as API endpoints for which an OpenAPI documentation was generated. ElasticSearch (ES) was used to provide a fuzzy full-text search. For this purpose, a specially defined search container with searchable content from various data tables was connected to ES by ZomboDB [3], a third party PostgreSQL extension. Unfortunately, PostgREST does not support ZomboDB's operator and query DSL. Thus, to adjust requests to the search endpoint, we leveraged OpenResty. The entire implementation of the backend was accelerated by the PostgREST Starter Kit [4] with its associated CLI tool [5]. The frontend was developed as a single page application (SPA) using the React Admin framework [6]. In order to communicate to an API, the React Admin app requires a compatible data provider which generates HTTP requests depending on the API architecture. In this case, a custom data provider for PostgREST was developed and used [7]. The overall system design is illustrated in Figure 1.

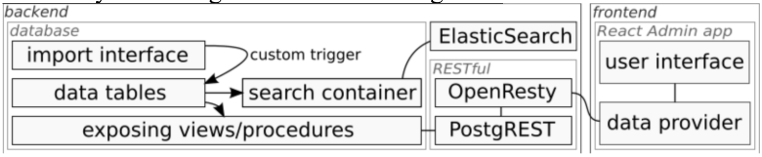


Figure 1. The architecture of the application.

2.2. Data Import

The NLM designed an XML format in which both the English and German MeSH exports are published as individual files. Consequently, a generic import procedure was developed. Specifically, each XML file was flattened into CSV files, partially containing JSON content in order to represent nested structures of the XML format. Subsequently, CSV data was directly written into the database using custom views of the underlying PostgreSQL database. JSON content was processed by trigger functions distributing nested data into the designated data tables. As the translations of MeSH qualifiers were missing in the German XML file, a CSV database export published by the DIMDI was processed to integrate the missing data. The import workflow was implemented using snakemake [8].

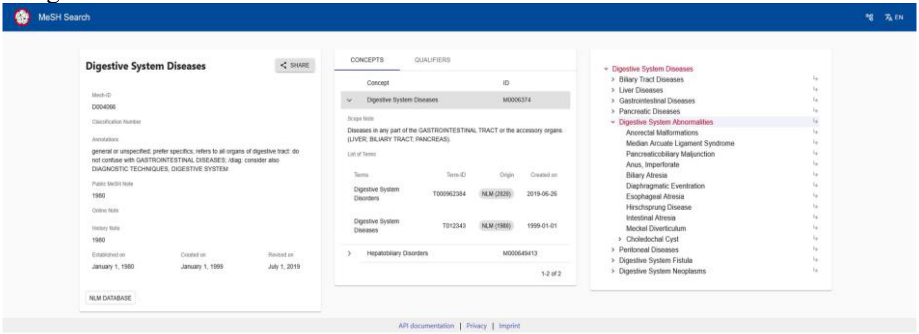
3. Results

3.1. REST Interface

The REST interface exposes 10 public endpoints: 7 of them to query specific structures from underlying data tables, 1 to provide answers to search queries and 2 to serve database procedures for querying the tree structure of MeSH. The internal structures consist of 16 tables, 5 views, 5 triggers and 8 procedures. One of the 5 views is for importing data using a customized trigger. Another one is the search container.

3.2. User Interface

The current version of the interface is built with minimal visual complexity and contains four pages to show MeSH information: (1) a landing page with a search field, (2) a list of search results, (3) detailed view for MeSH descriptors (cf. Figure 2) and (4) a MeSH tree browser. The interface of these pages is designed to minimize the number of steps required to reach a desired MeSH entity. Both language selection and link to the MeSH tree browser are accessible on each page. On the landing page, the query input field lists suggestions by utilizing ElasticSearch. When querying, the results are displayed in a list that allows further customization of the search term and sorting by score, MeSH term and entity type. On (3) and (4), detailed information about a specific MeSH descriptor, including related concepts and associated MeSH qualifiers, is displayed. Generally, the user interface labels are internationalized and designed for adding more translations in the future. Additionally, the web application offers consistent URLs and views of MeSH terms, providing shareable links. As a bonus, an integrated Swagger UI displays the OpenAPI documentation and enables users to consume the API endpoints directly through the browser.



Digestive System Diseases

Mesh-ID: D004066

Classification Number

Annotations: general or unspecified; prefer specifics; refers to all organs of digestive tract; do not confuse with GASTROINTESTINAL DISEASES; /diag: consider also DIAGNOSTIC TECHNIQUES; DIGESTIVE SYSTEM

Public MeSH Note: 1980

Online Note

History Note: 1980

Established on	Created on	Revised on
January 1, 1980	January 1, 1999	July 1, 2019

Concept	ID
Digestive System Diseases	M0006374
Hepatobiliary Disorders	M000649413

Scope Note: Diseases in any part of the GASTROINTESTINAL TRACT or the accessory organs (LIVER; BILIARY TRACT; PANCREAS).

List of Terms

Terms	Term-ID	Origin	Created on
Digestive System Disorders	T000962384	NLM (2020)	2019-06-26
Digestive System Diseases	T012343	NLM (1980)	1999-01-01

1-2 of 2

- Digestive System Diseases
  - Biliary Tract Diseases
  - Liver Diseases
  - Gastrointestinal Diseases
  - Pancreatic Diseases
  - Digestive System Abnormalities
    - Anorectal Malformations
    - Median Arcuate Ligament Syndrome
    - Pancreaticobiliary Maljunction
    - Anus, Imperforate
    - Biliary Atresia
    - Diaphragmatic Eventration
    - Esophageal Atresia
    - Hirschsprung Disease
    - Intestinal Atresia
    - Meckel Diverticulum
    - Choledochal Cyst
  - Peritoneal Diseases
  - Digestive System Fistula
  - Digestive System Neoplasms

Figure 2. Screenshot of the detailed view of a MeSH descriptor showing the tree and detailed information.

4. Discussion

The proposed system design provides a feasible solution to implement a multilingual MeSH browser. Based on fuzzy search functionality of ES, the platform provides a simple, yet powerful access to the MeSH thesaurus and offers streamlined adaptation of additional languages. This initial use case was restricted to help the team at “Wissen Was Wirkt” [9] - a German-language blog run by Cochrane Germany - to efficiently find German tags and categories for their blog posts using the comprehensive MeSH vocabulary. Another application could be the preparation of English search queries for MEDLINE, similar as in the abandoned work of the French MeSH Browser [10].

5. Conclusion

The presented MeSH browser was designed for multilingual functionality with modern open-source technology. In addition to a modern user interface, a RESTful API with OpenApi documentation was realized. The current published version contains English and German descriptors and is publicly available at https://mesh-browser.de.

Acknowledgements

This work was supported by the German Ministry for Education and Research (BMBF FKZ 01ZZ1801B, 01ZZ1804A). Ruslan Talpa is the CEO of subZero LLC, the company that develops the PostgREST Starter Kit and its associated CLI tool.

References

[1] Langnickel L, Baum R, Wollnik-Korn G, Fischer-Wagener B, Madan S, Fluck J. The future of German MeSH: a new semi-automatic translation process and new services for search and annotation. GMS Med Inform Biom Epidemiol. 2020 Aug 25;16(2):Doc07.

[2] PostgREST/postgrest: REST API for any Postgres database [Internet]. [cited 2021 Sep 1]. Available from: https://github.com/PostgREST/postgrest

[3] ZomboDB. zombodb/zombodb [Internet]. 2021 [cited 2021 Aug 24]. Available from: https://github.com/zombodb/zombodb

[4] PostgREST Starter Kit [Internet]. subZero; 2021 [cited 2021 Aug 24]. Available from: https://github.com/subzerocloud/postgrest-starter-kit

[5] subZero CLI [Internet]. subZero; 2021 [cited 2021 Aug 24]. Available from: https://github.com/subzerocloud/subzero-cli

[6] marmelab/react-admin: A frontend Framework for building B2B applications running in the browser on top of REST/GraphQL APIs, using ES6, React and Material Design [Internet]. [cited 2021 Aug 24]. Available from: https://github.com/marmelab/react-admin

[7] PostgREST Data Provider For React-Admin [Internet]. raphiniert.com; 2021 [cited 2021 Aug 24]. Available from: https://github.com/raphiniert-com/ra-data-postgrest

[8] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable data analysis with Snakemake. F1000Research. 2021 Jan 18;10:33.

[9] Blog Wissen Was Wirkt [Internet]. Wissen Was Wirkt. [cited 2021 Aug 23]. Available from: https://wissenwaswirkt.org/

[10] Thirion B, Pereira S, Névéol A, Dahamna B, Darmoni S. French MeSH Browser: a cross-language tool to access MEDLINE/PubMed. AMIA Annu Symp Proc AMIA Symp. 2007 Oct 11;1132.