Automating search term identification with MeSH entry terms via API

Helge Knüttel, Elke Hausner, Claudia Kapp

Riga, EAHIL 2024

Agenda

- Background

- Presentation of tools for capturing entry terms from PubMed

  - Command-line tool: "Fetch MeSH entry terms"

  - R-based tool: "R MeSH entry terms"

- Use cases

- Outlook

Why are entry terms important?

- A search in bibliographic databases is performed for systematic reviews

- The search consists of free-text terms and keywords (e.g. MeSH terms)

- Relevant free-text terms can be determined via various sources, e.g. entry terms in the MeSH database

3.2.2 Identifying relevant text words

Relevant text words (i.e. free-text terms) can be identified by checking the terms used in the title, abstract and other relevant fields (e.g. author keywords) of a sample of relevant references. It is important to be aware of the fact that natural language allows concepts to be expressed in different words. It is essential, therefore, to look up synonyms for each concept describing the review topic. Medical dictionaries can be used to clarify definitions and identify synonyms. The MeSH database also offers both definitions (Scope Notes) and a listing of synonyms and related terms for each MeSH term ('Entry terms'), which lists different terms being used for a concept. Likewise, Elsevier's Emtree thesaurus for Embase also lists synonyms for each term. Synonyms of pharmaceutical substances can be effectively searched via PubChem (https://pubchem.ncbi.nlm.nih.gov/). A third approach for identifying text words consists of checking search strategies from other systematic reviews on a similar topic.

Lefebvre C, et al. Technical Supplement to Chapter 4: Searching for and selecting studies. In: Higgins JPT et al. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 6.4 (updated February 2024). Available from www.training.cochrane.org/handbook.

What are MeSH entry terms?

Entry Terms and Other Cross-References (https://www.nlm.nih.gov/mesh/intro_entry.html)

- MeSH ist part of the NLM knowledge resources

- "Entry terms [...] are synonyms, near-synonyms alternate forms, and other closely related terms in a MeSH record.

- "Entry terms [...] are equivalent to the preferred term for purposes of cataloging, indexing and retrieval"

- "Each entry term expression in MeSH is unique and carefully chosen to avoid expressions that are multimeaning." (https://www.nlm.nih.gov/mesh/meshsugg.html)

MeSH in Entrez

Access to MeSH via API

- **Entrez**: Query and database system of 40 databases from NLM/NCBI

    - PubMed, MeSH & many molecular databases

- **Entrez Programming Utilities (E-utilities):** Web-APIs to the Entrez system

    - API = application programming interface → to be used by other programs

- **Entrez Direct (EDirect)**: Programs to access E-utilities on the Unix **command-line**

    - Provided by NLM/NCBI, open source

    - Must be installed on computer

    - Easy and powerful access to the databases plus tools to process the data

    - → Used here

Presentation of tools for capturing entry terms from Pubmed

COMMAND-LINE TOOL: FETCH MESH ENTRY TERMS

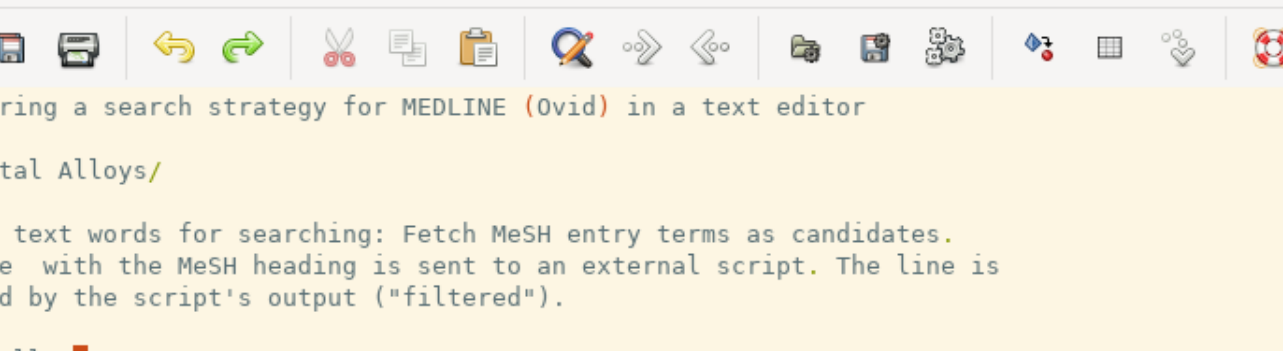# Command-Line Tool: Fetch MeSH entry terms

- **Goal**: Use entry terms as candidates for text word searches

    - Efficient source of terminological breadth for a concept

    - Identify terms that may not be present in the seed articles (if you have any at all)

- The **manual way**:

    - Find relevant MeSH Heading in browser → copy *Main Heading* → paste → edit

    - Copy the other *Entry Terms* → paste → edit

    - Repeat for all exploded descriptors

    - Check exploded terms in all trees: Trees may differ

Command-Line Tool: Fetch MeSH entry terms

Use the **Entrez API/Edirect** to fetch MeSH Entry Terms:

- After identifying relevant MeSH Heading(s):

  - Search MeSH → fetch XML data → extract entry terms → output

  - Explosion: Extract tree-numbers from XML → search again with truncation → …

- This works by chaining a few freely available programs:
  Output of one program is input of the next one (Unix philosophy)

- Combination of all steps into a single script makes it easy to use

- Can be used from other tools

- Explosion (in all trees) is the default

Command-Line Tool:
Fetch MeSH entry terms from within text editor

Command-Line Tool:
Fetch MeSH entry terms from within text editor

Command-Line Tool:
Fetch MeSH entry terms from within text editor

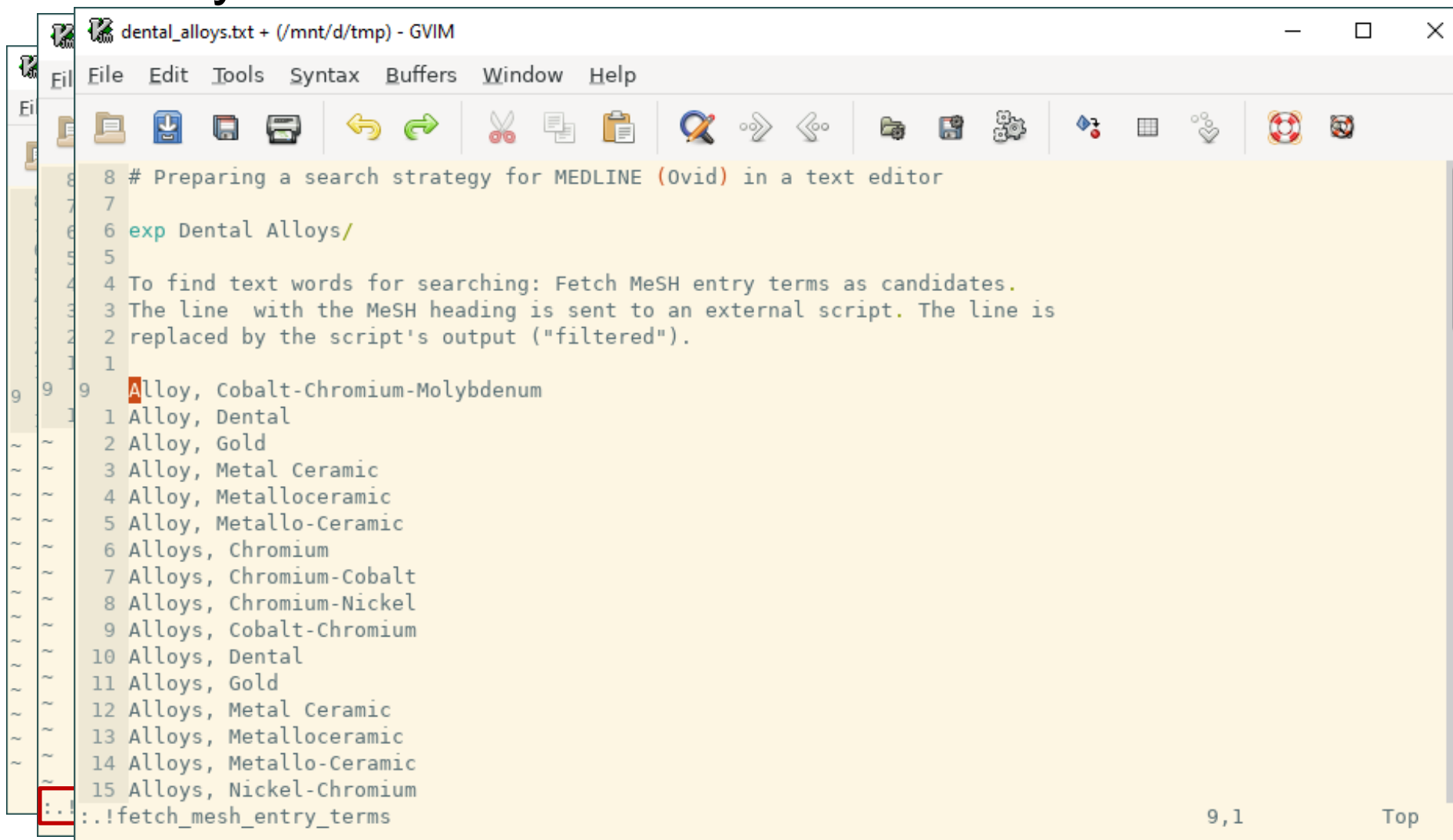dental_alloys.txt + (/mnt/d/tmp) - GVIM

File   Edit   Tools   Syntax   Buffers   Window   Help

```
8 # Preparing a search strategy for MEDLINE (Ovid) in a text editor
7
6 exp Dental Alloys/
5
4 To find text words for searching: Fetch MeSH entry terms as candidates.
3 The line  with the MeSH heading is sent to an external script. The line is
2 replaced by the script's output ("filtered").
1

  Alloy, Cobalt-Chromium-Molybdenum
1 Alloy, Dental
2 Alloy, Gold
3 Alloy, Metal Ceramic
4 Alloy, Metalloceramic
5 Alloy, Metallo-Ceramic
6 Alloys, Chromium
7 Alloys, Chromium-Cobalt
8 Alloys, Chromium-Nickel
9 Alloys, Cobalt-Chromium
10 Alloys, Dental
11 Alloys, Gold
12 Alloys, Metal Ceramic
13 Alloys, Metalloceramic
14 Alloys, Metallo-Ceramic
15 Alloys, Nickel-Chromium
```

:.!fetch_mesh_entry_terms   9,1   Top

Command-Line Tool: **User Feedback**

- Helge plus 1 other user: Neat little tool that saves time. ☺

- Use of command-line not common among expert searchers.

- → Hard to find testers.

- Please, try it out and give feedback!

- Thus, another approach by Claudia …

Presentation of tools for capturing entry terms from Pubmed

# R-BASED TOOL "R MESH ENTRY TERMS"

R-based Tool "R MeSH entry terms"

- Developed by Claudia Kapp

- Proof of concept created: February 2024

- Shiny App (access via local network)

- Functionalities

    - browse and export MeSH entry terms

    - compare MeSH entry terms to a user's test set (RIS/PubMed format) -> not presented today

R-based Tool "R MeSH entry terms"

R-based Tool "R MeSH entry terms"

R-based tool: "Find MeSH entry terms"

R-based tool: "Find MeSH entry terms" (txt-file)

Test user feedback

So far 4 people have tested the tool

What did you like?

Suggestions for improvements

„Easy to use"

"Workflow is not intuitive: App needs more documentation about what it does :)"

„Does what it should"

"Error messages should be more meaningful to an end user and less technical"

"Very simple"

"What is that good for?"

Use Cases

- IQWiG

  - For quick orientation and familiarisation with a new topic

  - <u>Not</u> a primary source for free text terms

  - Refinement of the phrase search

- Helge

  - One source of several for candidate search terms.

  - Like quick and seamless integration.

  - MeSH entry terms can be reported for development of search strategy.

  - Usefulness of the MeSH entry terms depends on topic.

Outlook

- Implementation of other thesauri

  - Emtree: More comprehensive but without API ☹

  - Snomed CT

  - UMLS Metathesaurus

- R-based Tool

  - Integrate in searchbuildR if necessary

Tool Access

- Command-Line Tool

  - https://github.com/knh11545/cli4es

  - Try it out in the browser (BinderHub)!

  - Contact: helge.knuettel@ur.de

- R-based Tool

  - Source code is currently not yet open source and can be shared upon request.

  - Contact: Claudia Kapp (recherche@iqwig.de)

Our context: IQWiG

Claudia Kapp
- Master in Clinical Psychology

Information Specialist at IQWiG since 2021
- Development of 10-20 search strategies per year
- Programming and Data analysis with R (since 2019)
- Development of the package and Shiny app searchbuildR

Elke Hausner
- Bachelor in Nursing Science
- Master in Public Health

Information Specialist at IQWiG since 2006
- 30-40 peer reviews on search strategies per year
- No programming skills

Our Context: Helge Knüttel

- Subject & liaison librarian
- University of Regensburg Library, Germany
- Part of my work:
    - Systematic searches
    - Methodological support for evidence synthesis projects
    - ~12-25 searches/year

- The tools I am using keep changing …
- For search strategy development & documentation:
    - Switched from Word to a text editor (vim): Plain text & Markdown-formatted text files
    - Prepare and keep search strategies in text files
    - Git/GitLab for organising projects & sharing files with researchers
    - Unix-style command-line tools:
        - Replace common manual tasks with automation
        - Quality checks
        - Augment documentation