Project update

Outline

- Introduction
- Process flow of the overall system
- Retrieving the mappings
  - UMLS
  - BioPortal
  - SNOMED CT mapping tool
- Retrieving UMC publications
  - Querying PubMed

Introduction

Goal: build an evaluation measure that shows how well the research output of Dutch UMCs match their different patient groups.

- Publications are indexed with MeSH terms (with a vast subset that describes diseases)
- Patient diagnoses are labeled with ICD10 codes

How?: Get mappings between MeSH and ICD10 that represent the same disease

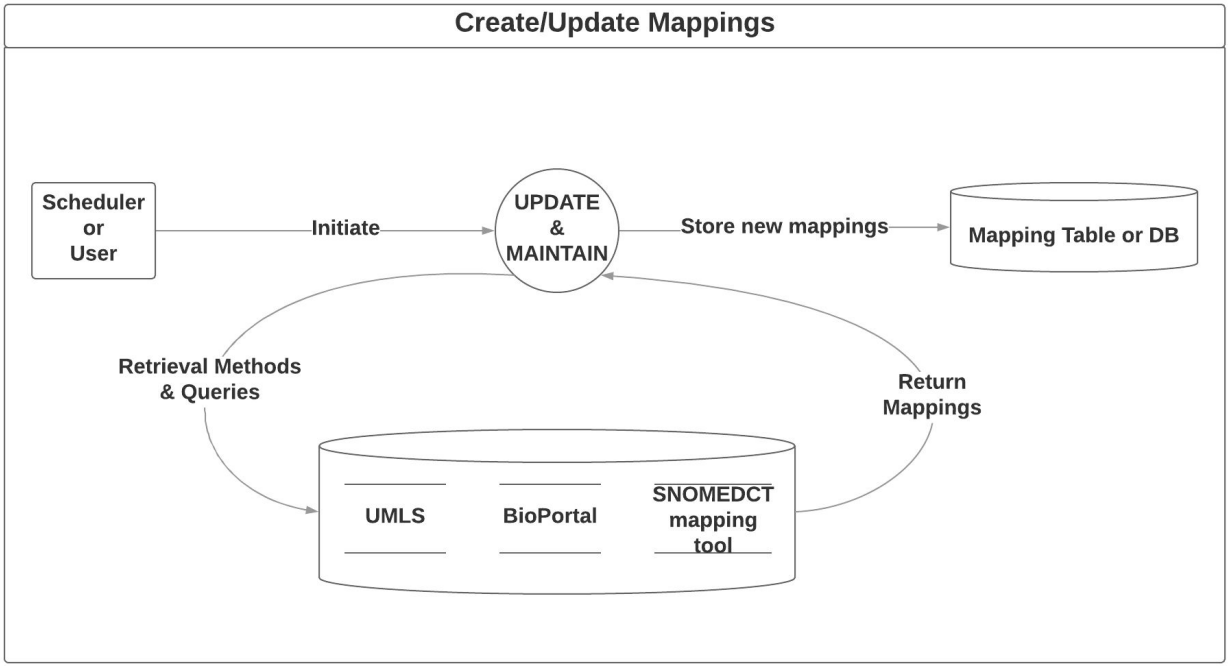
Where?: Find several communities that have created mappings between medical vocabularies.

UMLS

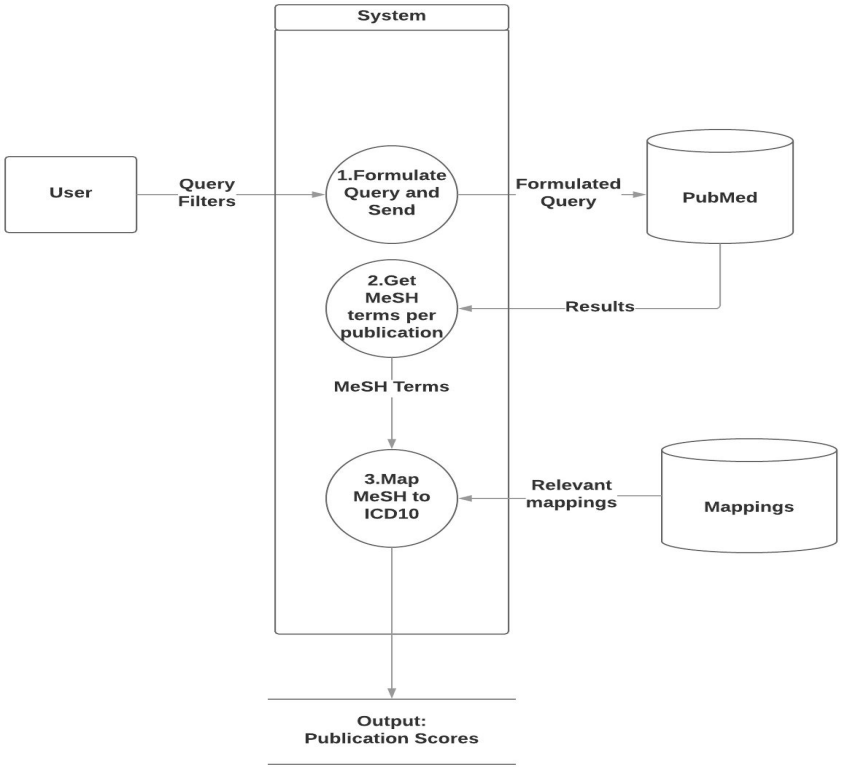
BioPortal

SNOMED CT

Process Flow of the Overall System(1/2)



Process Flow of the Overall System(2/2)

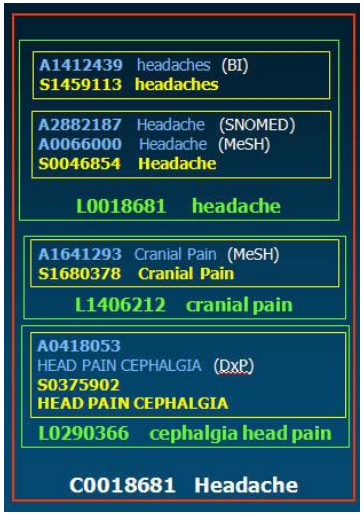


Retrieving the mappings

UMLS (1/3)

UMLS includes Metathesaurus software:

- A large biomedical thesaurus that organizes terms from several sources by concept, meaning and links similar names for the same concept.



A frame structure of 4 levels:

1. Concept Unique Identifier (CUI) ⇒ **Meaning**
2. Lexical Unique Identifier (LUI) ⇒ **Lexical Variants**
3. String Unique Identifier (SUI) ⇒ **Further variations**
4. Atom Unique Identifier (AUI) ⇒ **Building Blocks**

Image source: https://www.nlm.nih.gov/research/umls/new_users/online_learning/Meta_005.html

UMLS (2/3)

Atoms in one concept have the same meaning. Hence, ICD10 and MeSH terms are part of the same concept can be considered as a mapping.

For the current project, all the mappings can be found in a single file of Metathesaurus: **MRCONSO.RRF**

CUI	AUI	LAT	SAB	CODE	STR	…..
C0000727	A8183940	DUT	MSHDUT	D000006	Bulk, acute	
C0000727	A0017734	ENG	MSH	D000006	Abdomen, Acute	
C0000727	A0639289	ENG	ICD10	R10.0	Abdomen, Acute	

UMLS (3/3)

Results:

- Found 1782 CUIs that contain both MeSH and ICD10.
- Found **2116** unique MeSH-ICD10 pairs.
- Added Dutch translation wherever it was present.
- Note: There are restrictions of publishing ICD10 and translations of any source General terms + additional restrictions in category 12.3 (https://uts.nlm.nih.gov/help/license/licensecategoryhelp.html)

BioPortal

BioPortal is a repository of biomedical ontologies that can be accessed through Web Services.

- It offers a REST api in order to access its resources. An API KEY is required!
- BioPortal creates the mappings through several methods:
  - Using UMLS
  - Using NCBO loom algorithm
  - OBO mappings
- Retrieval methods:
  - Method 1: Immediate mappings between ICD10 and MeSH (22 out of 2119)
  - Method 2: Mappings between MeSH <-> ICD10CM <-> ICD10 (779 out of 2538)
  - Method 3: Mappings between MeSH <-> SNOMED <-> ICD10 (On progress)

SNOMED CT

Mappings between SNOMED CT to ICD10CM

“The purpose of the SNOMED CT to ICD-10-CM map is to support semi-automated generation of ICD-10-CM codes from clinical data encoded in SNOMED CT for reimbursement and statistical purposes.”

- I-MAGIC: Algorithm that implements real-time mapping between SNOMEDCT and ICD10CM
- **Demo page:** https://imagic.nlm.nih.gov/imagic/code/map
- All mappings are in a file that can be accessed with the UMLS or SNOMED License.
- Link of release: https://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd10cm.html
- FIlename: der2_iissscRefset_ExtendedMapSnapshot_INT_YYYYMMDD.txt which is in the Snapshot\Refset\Map directory.

- On progress

Retrieving UMC publications

NCBI offers several public programmatic APIs to allow access to various Databases.

The most useful API is the **Entrez Programming Utilities** (E-Utilities):

- Allow access to all Entrez databases (PubMed,PMC, Gene etc.)
- Eight programs that are in the form of fixed URLs.
- Documentation: https://www.ncbi.nlm.nih.gov/books/NBK25501/

Using Entrez, one can access Pubmed, then ask to retrieve publications according to some *search criteria* and get back *results* (possibly indexed with *MeSH terms*).

Querying PubMed

Querying PubMed can be performed programmatically with E-Utilities, by defining a typical PubMed query (as in the PubMed site).

Current filters of interest:

1. **MeSH Categories**
   - “Diseases Category"[Mesh]
   - "Health Care Category"[Mesh]
   - "Psychiatry and Psychology Category"[Mesh]
2. **Date Ranges**
   - YYYY/DD/MM:YYYY/DD/MM[Publication Date]
3. **Authors’ Affiliation**
   - Authors with a certain affiliation can be searched with the [ad] filter
     - e.g: For Leiden: (Leiden[ad] AND ("university medical center"[ad] OR "university medical centre"[ad] OR "academic hospital"[ad] OR mc[ad] OR LUMC[ad]))

Querying PubMed

The 3 filters can formulate a single query by conjuncting them:

Query := MeSHCategories AND DateFilters AND Affiliation

Two E-utilities tools are used:

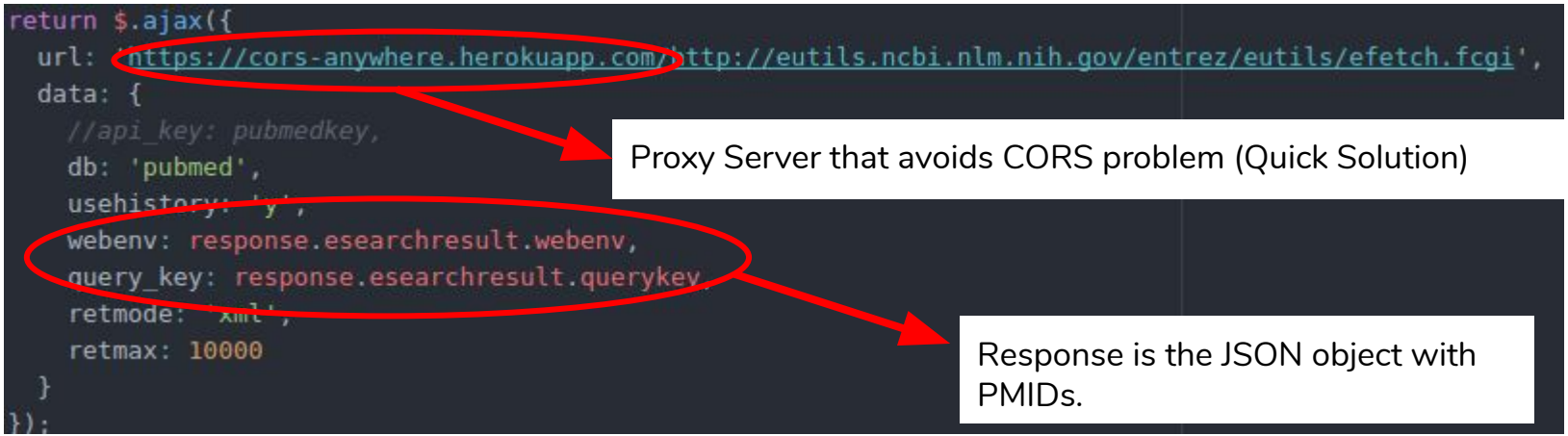
1. **Esearch:** returns a list of PMIDS that satisfy certain filter criteria (query).

2. **Efetch:** returns detailed information of publications in given PMID list.

Querying PubMed
Esearch as an ajax request

- Return a JSON object of PMIDS.

Querying PubMed
Efetch as an ajax request



- Return an XML file with PMID information (title, DOI, Publication Date, authors, MeSH, keywords etc.)

Querying PubMed

Things to concern:

- An API-KEY is not necessary, but it makes the querying process faster.
- The current implementation is coded with JavaScript, which is a bad practice. Back-end implementations should be developed.
- E-utilities limit the results to a maximum of 10000. An iteration process of getting the next results is necessary.