First study:

Manual assessment

Massonnaud, C., Lelong, R., Kerdelhué, G., Lejeune, E., Grosjean, J., Griffon, N., & Darmoni, S. J. (2019). Performance evaluation of three semantic expansions to query PubMed. Health information and libraries journal, 10.1111/hir.12291. Advance online publication. doi:10.1111/hir.12291

Method

Queries were built from the 40 MeSH descriptors most frequently used for indexation, in the MeSH Diseases category C.
The queries were expanded using the three strategies: CISMeF, MeSH and UMLS.
Three boolean queries were built out of these expansions for each of the 40 descriptors:

1. Q(CISMeF not MeSH)
2. Q(CISMeF not UMLS)
3. Q(UMLS not CISMeF)

The boolean queries were performed in PubMed and the 20 first citations were retrieved.
The relevance of each citation was evaluated by three independent evaluators, on a 3-modality scale (fully, partially, not relevant), based on title and abstract only.

Results

Table 4 Mean precision, by expert and relevance

Expert	Q(CISMeF NOT MeSH+) (SD)	Q(CISMeF NOT UMLS) (SD)	Q(UMLS NOT CISMeF) (SD)	*P*-value*
Precision for partial and full relevance (%)				
cm	67.05 (± 21.10)	57.08 (± 21.16)	60.40 (± 26.14)	0.683
gk	79.12 (± 24.14)	73.75 (± 28.93)	77.20 (± 32.25)	0.747
el	88.03 (± 11.28)	80.00 (± 13.98)	84.97 (± 21.48)	0.422
all	78.07 (± 10.17)	70.28 (± 11.85)	74.19 (± 12.56)	
Precision for full relevance (%)				
cm	31.09 (± 17.71)	24.17 (± 16.63)	33.00 (± 20.67)	0.175
gk	40.57 (± 28.80)	30.83 (± 29.91)	46.40 (± 33.87)	0.169
el	51.41 (± 23.91)	39.17 (± 23.63)	49.61 (± 29.89)	0.26
all	41.02% (± 10.17)	31.39 (± 7.52)	43.00 (± 8.81)	

*Student's t-test for the mean differences between Q(CISMeF NOT UMLS) and Q(UMLS NOT CISMeF) (R 3.4.3 package *stats)*.

Table 5 Kappa statistic for inter-rater agreement and pairwise agreement (Light's method), computed depending on different groupings of the levels of the relevance assessment

	Levels of relevance assessment*		
Kappa statistic	Po, Pa, F	(Po + Pa), F	Po, (Pa + F)
Inter-rater agreement			
cm – el – gk	0.30	0.50	0.28
Pairwise agreement			
el – gk	0.43	0.60	0.30
cm – gk	0.25	0.47	0.25
cm – el	0.22	0.43	0.21

*Po: poor relevance, Pa: partial or full relevance, F: full relevance. (Po + Pa) or (Pa + F) means the levels have been aggregated into one category.

Queries built with the UMLS expansion provided new citations with a slightly higher mean precision (74.19%) than with the CISMeF expansion (70.28%), although the difference was not significant.

Inter-rater agreement was 0.28.

Results varied greatly depending on the descriptor selected

Discussion

The number of citations retrieved by the three strategies and their precision varied greatly according to the descriptor. This heterogeneity could be explained by the quality of the synonyms.
Limitations: indirect comparison of the strategies, only 40 descriptors evaluated, manual assessment with poor inter-rater agreement.

Assessment of semantic expansion strategies to query Pubmed

Objectives

To propose an alternative optimisation of the standard PubMed ATM query strategy, by expanding the end-user query with synonyms generated by different mappings: MeSH, UMLS, and CISMeF.

Semantic expansion strategies

1. All queries were based on a MeSH descriptor (e.g. *diabetes mellitus*)

2. They were all first expanded by PubMed's ATM (Automatic Term Mapping):

   "diabetes mellitus"[MeSH Terms] OR ("diabetes"[All Fields] AND "mellitus"[All Fields]) OR "diabetes mellitus"[All Fields]

3. Then they were further expanded using synonyms:

   "diabetes mellitus"[MeSH Terms] OR ("diabetes"[All Fields] AND "mellitus"[All Fields]) OR "diabetes mellitus"[All Fields] OR "synonym 1"[All fields"] OR "synonym 2"[All field] OR ...

4. Three different strategies were used to identify synonyms:

   MeSH: mappings using synonyms provided by the MeSH thesaurus

   UMLS: mappings using the Concept Unique Identifiers (CUI) of the UMLS metathesaurus

   CISMeF: mappings between the terms of the 71 KOS available in the HeTOP (Health Terminology/Ontology Portal) cross lingual terminology server:

   - manual mappings
   - supervised mappings: automatic alignment *via* UMLS or using Natural Language processing, then manually curated.

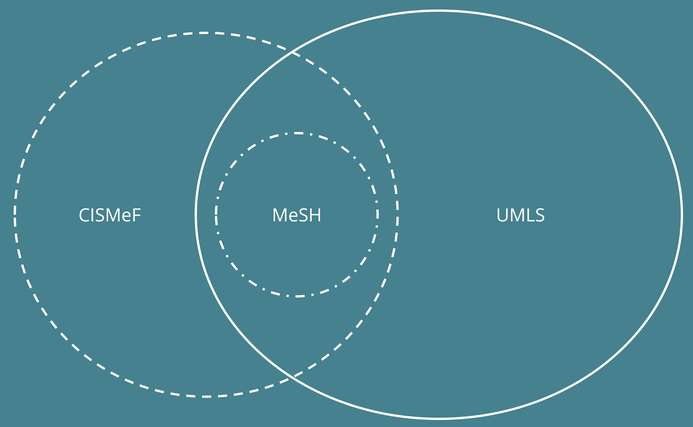


Figure: representation of the scope and relations between the CISMeF, MeSH, and UMLS terminologies.

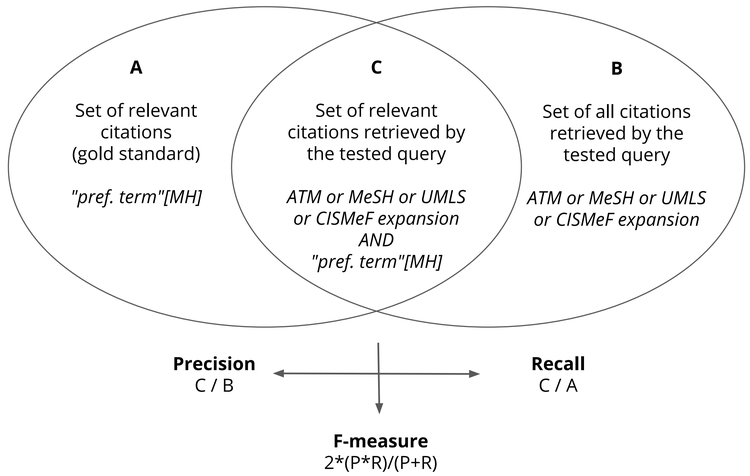
Second study:

Automatic assessment

Massonnaud C., Kerdelhué G., Grosjean J., Lelong R., Griffon N., Darmoni SJ. Finding the best semantic expansion to query PubMed: automatic performance assessment of four search strategies on all MeSH descriptors.
In re-review at JMIR Medical Informatics.

Method

Innovative method to compute automatically information retrieval metrics : precision, recall, and F-measure.
Therefore the expansion strategies can be evaluated on all the 26,636 MeSH descriptors.



Results

Table 2. Mean performances of the four search strategies for the 26,636 MeSH descriptors

KOS	Precision, % (sd)	Recall, % (sd)	F-measure, % (sd)
ATM	44.24 (± 24)	31.12 (± 29)	28.41 (± 23)
MeSH	50.93 (± 23)	38.01 (± 31)	34.59 (± 24)
CISMeF	49.20 (± 23)	40.11 (± 31)	35.10 (± 24)
UMLS	49.20 (± 23)	40.57 (± 31)	35.51 (± 24)

KOS: knowledge organisation system

The performances of the semantic expansion strategies vary greatly depending on the MeSH descriptor.

Discussion

These results confirm there is no ideal search strategy for all descriptors. Different semantic expansions should be used depending on the descriptor and the user's objectives.

This led our team to develop a tool that allows users to input a descriptor and choose which metric they wish to maximize, automatically build the query with the appropriate expansion strategy, and send it to PubMed.

This tool is freely available on the HeTOP website, and could benefit to all researchers to improve their information searching in day-to-day use.