2020-02-10

Term Harvesting & Generation Guide

MacEachern, Mark; Saylor, Kate; Anderson, Patricia; Townsend, Whitney; Ginier, Emily

Term Harvesting & Generation Guide

Generating a comprehensive list of relevant search terms can be challenging and can vary from database to database. This document includes a list of strategies that expert searchers use to expand the list of terms and serves as a guide for generating search terms for comprehensive searching; depending on your topic, not all strategies will be relevant or necessary.

General rule: A single term is **NOT** sufficient to represent all published studies on a topic.

Considerations when identifying terms

- What are the main searchable concepts in your question?
- Is the term or combination of terms truly relevant? Is it or are they representative of one of your question's main concepts?
- What terms do the experts use? For example, if your question involves graduate medical education, then authors who publish in education journals will likely use more education jargon and less clinical jargon, than those who publish in medical education journals.
- How comprehensive of a search are you undertaking? If comprehensive, then err on the side of more terms. If comprehensiveness is not essential, then aim to include the most important terms.
- Does the term have a common acronym/abbreviation? Include it unless the acronym is common in other contexts.

Strategies to find and include terms

- Identify relevant terms from titles and abstracts of relevant articles
- Identify controlled vocabulary terms from relevant articles
- Use entry terms and synonyms included within databases to identify additional terms (see Appendix for more information)
- Use truncation in databases that allow it, but do not shortened the word too much

Common databases (and their controlled vocabularies)

- CINAHL (CINAHL headings)
- Cochrane databases (MeSH)
- Embase (EMTREE)
- ERIC (Subjects)
- Google Scholar (None)
- Medline databases (MeSH)
- PsycINFO (Subjects/Descriptors)
- PubMed (MeSH)
- Scopus (None)
- SportDiscus (Subjects/Descriptors)
- Web of Science (None)

Use the following table as a guide to identify terms for each concept of your search. The relevance of each type of term listed in the table varies from concept to concept.

Type of term	Examples	Notes
Controlled Vocabulary	MeSH: • Breast neoplasms • Education, medical EMTREE • Breast tumor • Medical education	Ideally one controlled term will represent a concept, but sometimes you have to combine two.
Term variations, including plurals	Diabetes, diabetics	
Synonyms	Cancer, cancerous, cancers, carcinoma, carcinomas, malignancies, malignancy, malignant, neoplasm, neoplasms, neoplastic, tumor, tumors, tumour, tumours	What terms are closely related? 'Malignant' and 'cancer' are not synonymous, but the former often relates to the latter.
Alternate spellings	Pediatrics, paediatrics Tumors, tumours	
Acronyms / Abbreviations	SSRI, selective serotonin reuptake inhibitor "T2 DM", T2DM, DMT2, "DM T2", DM2, "DM 2", "T2 Diabetes", "type 2 diabetes"	Consider possible issues when using these. 'Operating room' is often shortened to OR, but OR is a poor search word (too many meanings, is a stop word). Also, DFS, for example, has different meanings in dentistry and oncology.
Concept combinations	The concept of "shoulder pain" could be searched in PubMed as MeSH Shoulder Pain or as a combination of shoulder and pain terms.	This is more likely to be important when one concept, like shoulder pain, can be divided into two, shoulder & pain
Historic/archaic terminology	Burma & Myanmar	Consider if terminology has changed over time. How were diseases or populations described in previous eras?
Slang, jargon, common vernacular	Swine flu Caries/cavities	
Common spelling errors	Alzheimers vs Alzheimer's vs Alzheimer; crowdsourcing vs "crowd-sourcing"	If a term can conceivably be hyphenated, consider including both variations

Appendix: Entry terms / Synonyms

Databases with controlled vocabularies will often allow you to browse and search through the vocabulary for information about individual terms. Sometimes this information includes an internal set of associated terms that serve some backend purpose, but can be used as a brainstorming tool for searchers. PubMed, for example, includes entry terms in the MeSH database, which is a list of the controlled terms (i.e., MeSH) that PubMed indexers assign to citations. The EMTREE vocabulary in Embase includes a similar list of synonyms for many of its controlled terms. By exploring these pre-defined sets of synonyms and including some of the individual terms in their search will help researchers search more comprehensively.

Furthermore, MeSH and EMTREE are both hierarchical structures and seeing an individual term within context of other terms can provide further suggestions on additional terms.

Example 1: Myocardial infarction entry terms and location in MeSH

Myocardial Infarction

NECROSIS of the MYOCARDIUM caused by an obstruction of the blood supply to the heart (CORONARY CIRCULATION).
Year introduced: 1979

Entry Terms:

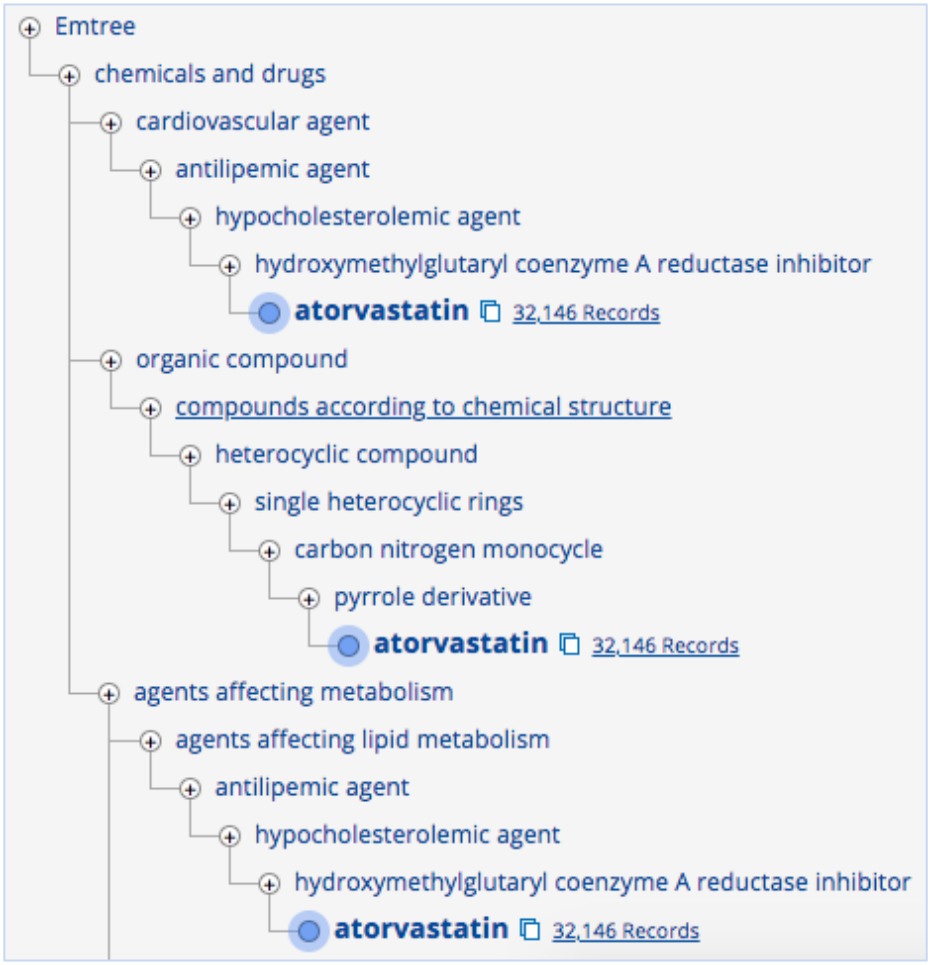
- Infarction, Myocardial
- Infarctions, Myocardial
- Myocardial Infarctions
- Cardiovascular Stroke
- Cardiovascular Strokes
- Stroke, Cardiovascular
- Strokes, Cardiovascular
- Heart Attack
- Heart Attacks
- Myocardial Infarct
- Infarct, Myocardial
- Infarcts, Myocardial
- Myocardial Infarcts

See Also:

- Heart Rupture, Post-Infarction

All MeSH Categories
Diseases Category
Cardiovascular Diseases
Heart Diseases
Myocardial Ischemia
Myocardial Infarction
Anterior Wall Myocardial Infarction
Inferior Wall Myocardial Infarction
Non-ST Elevated Myocardial Infarction
Shock, Cardiogenic
ST Elevation Myocardial Infarction

Example 2: Atorvastatin Emtree location and synonyms, including CAS numbers



History

This term was added to Emtree in 1990

Synonyms

2 (4 fluorophenyl) beta, delta dihydroxy 5 isopropyl 3 phenyl 4 phenylcarbamoyl 1h pyrrole 1 heptanoic acid; atorlip; atorvastatin calcium; atorvastatin calcium trihydrate; atovarol; cardyl; ci 981; ci981; glustar; lipibec; lipitor; liprimar; liptonorm; lowlipen; sortis; storvas; tahor; torvast; totalip; xarator; ym 548; ym548; zarator

CAS Registry Numbers

134523-00-5; 134523-03-8