

WHITEPAPER

# A Comparison of Emtree® and MeSH®



## WHAT ARE THE DIFFERENCES BETWEEN EMTREE AND MESH?

Emtree and MeSH are both comprehensive biomedical and life science thesauri, respectively used to index the biomedical literature in Embase® and MEDLINE®. This short paper examines the differences between them, highlighting their relative merits in terms of the discoverability of data.

# Emtree is used for full-text indexing of all journal articles in Embase. MeSH is used to index articles for MEDLINE.

“Emtree really helps me because if I find an interesting paper, I can review similar papers through the terms of the entry. Compared to the way of searching in other tools, it’s much easier.”

Dr. Federico Aletti, Research Fellow at the Politecnico Di Milano

The Emtree thesaurus is a hierarchically structured, controlled vocabulary for biomedicine and the related life sciences. It includes a whole range of terms for drugs, diseases, medical devices and essential life science concepts (Figure 1). Elsevier’s biomedical experts use Emtree for deep, full-text indexing of all journal articles in Embase, ensuring maximum discoverability of biomedical evidence.

MeSH stands for medical subject headings. MeSH is a controlled vocabulary health sciences thesaurus used by U.S. National Library of Medicine experts to index articles for MEDLINE/PubMed. It consists of comprehensive sets of terms with descriptors in a hierarchical structure and its purpose is also to support discoverability of data and evidence.

Emtree and MeSH have similar facet structures. Emtree was modeled on MeSH in 1988. They both include broader and narrower terms and synonyms, and are linked to CAS registry and Enzyme Commission numbers.

The screenshot displays the 'Browse Emtree' interface. At the top, there is a 'Query Builder' section with a search bar containing 'rofecoxib' and a 'Find Term' button. Below this, there are options to 'Extend your search' with checkboxes for 'Explosion' (checked) and 'As major focus'. Two buttons, 'Take this query to Drug Search' and 'Add to Query Builder', are visible. The main part of the interface shows a hierarchical tree structure of terms for 'rofecoxib'. The tree starts with 'chemicals and drugs' and branches into several categories: 'organic compound', 'agents affecting metabolism', 'analgesic, antiinflammatory, antirheumatic and antigout agents', and 'antirheumatic agent'. Each branch further subdivides into more specific terms, with 'rofecoxib' listed at the bottom of each branch, accompanied by a record count of 9,756.

Figure 1. An Emtree search for rofecoxib reveals these hierarchies of terms.

## THE ADVANTAGES OF EMTREE

**Table 1** summarizes the benefits that Emtree confers. Essentially, the design and content of Emtree makes it easier to use, with terminology given in a natural language order, no need to read supplementary files to understand terminology, and better support for explosion searches. Furthermore, Emtree has more up-to-date drug and medical device terminology, and more preferred terms and synonyms than MeSH, greatly increasing the discoverability of articles.

Emtree	MeSH	Emtree benefits for users
Terminology is given in natural language order (e.g., myeloid leukemia)	Terminology may be inverted (e.g., leukemia, myeloid)	Emtree is more intuitive with the terms being more easily recognized.
Over 300,000 synonyms (including over 190,000 drug synonyms)	Includes fewer synonyms (or entry terms): ca. 220,000	There is a higher probability that search terms chosen by users are in Emtree.
Includes all of the MeSH terms, many as Emtree synonyms	Does not include all the Emtree terms	MeSH users find Emtree easy to search.
Relies upon meaning invested in terms by the authors use of them	Has many scope notes to describe how terms are used or intended	Embase users do not need to look up scope notes to understand terms.
Larger: over 71,000 preferred terms	Smaller: over 27,000 preferred terms	Emtree gives users a better chance of finding both drug and non-drug terminology.
Extensive drugs facet with over 31,000 preferred terms	A more limited drugs facet with around 9,250 preferred terms	Emtree has a more comprehensive and up-to-date drugs terminology.
New drug terms updated earlier, and updates are made three times per year	Only established drug terms are added	Emtree gives better results for new drugs.
Over 3,000 specific terms for general and medical devices (e.g., endoscopes, catheters, prostheses)	Fewer medical device terms	Emtree is the best available resource for medical device information.
Polyhierarchical structure with duplicated trees	Polyhierarchical structure with differences between trees	Emtree enables unambiguous and context-free explosion searches.
All drug and chemical information included	Detailed drug information in a supplementary file	All the necessary drug information is in Emtree with no need to resort to other files.
Detailed drug and device trade and manufacturer name indexing	Indexes fewer terms that are not searchable in MEDLINE	Embase users can search or filter for specific drug or device trade or manufacturer names.
Updated three times per year: the latest drugs, diseases, organisms and procedures are indexed and added with back-posting of older records	Updated once per year	Emtree's update schedule ensures that new information is added quickly so that Embase users can easily find current literature.

**Table 1.** Advantages of Emtree when compared to MeSH

## THE ADVANTAGES OF MESH

MeSH has extensive history notes that can be used to track earlier literature predating the introduction of particular terms. It also has extensive scope notes. However, these include restrictions in term usage which users need to be aware of, as well as term definitions.

It is also worth noting that MeSH has extensive terminology in nursing, veterinary medicine and dentistry. While all of these terms are in Emtree, many are synonyms.

---

## WHY CHOOSE EMTREE INSTEAD OF MESH?

The Emtree terminology:

- Includes all of MeSH (i.e., all of the MeSH terms are mapped to Emtree terms)
- Is readily extensible in response to user needs, subject to agreement with overall policy
- Is optimally embedded into Embase, Elsevier's premier biomedical research solution

In addition, the drug and medical device terminology in Emtree is:

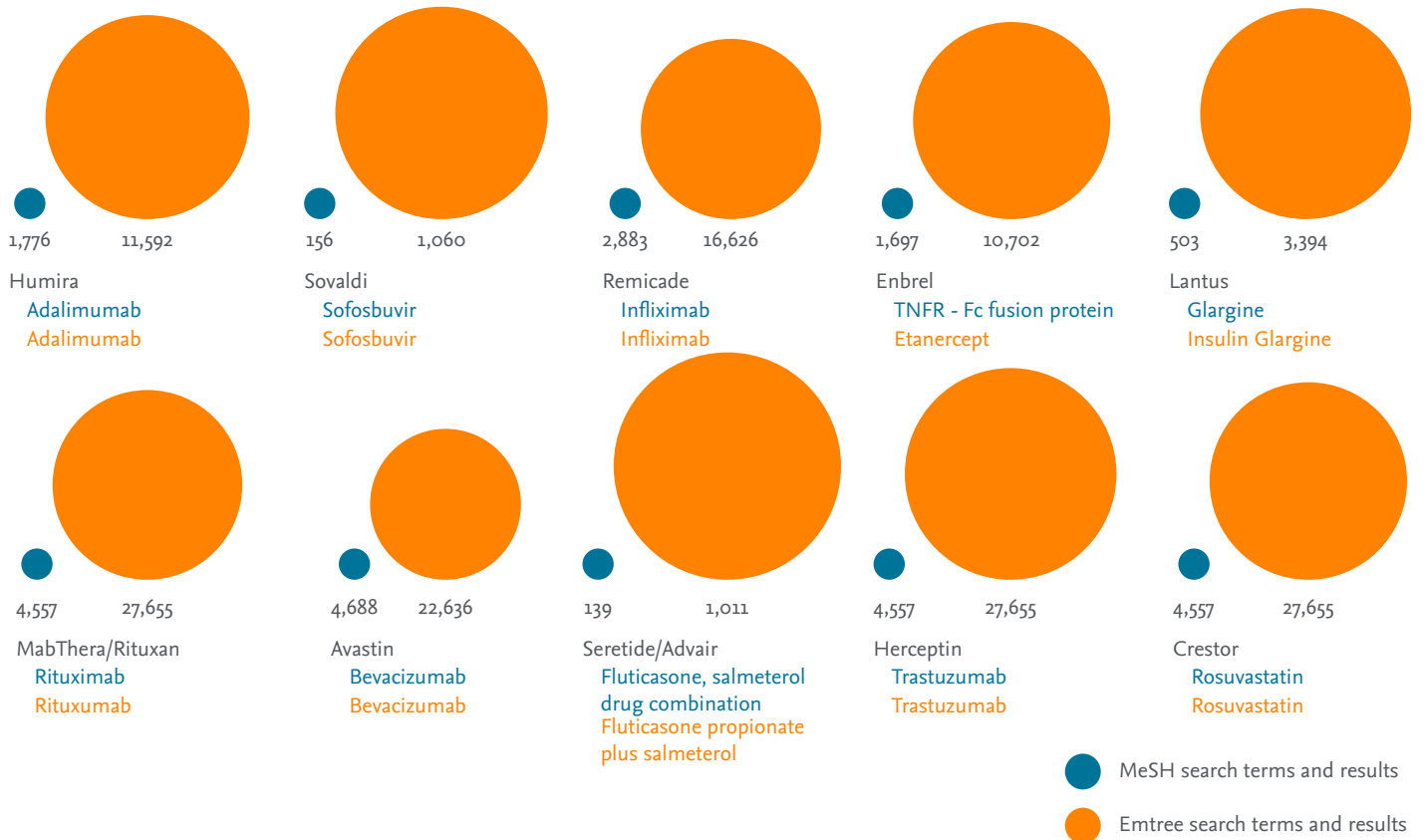
- More extensive, with many synonyms (including many trade names) mapping to each drug or device term
- Up to date, making Emtree especially suited for finding information on new drugs and devices
- Polyhierarchical, meaning each Emtree term is accessible from multiple points of view (e.g., for drug terms: structure, activity, therapeutic use)

## COMPARATIVE DRUG SEARCHES

To illustrate the differences between Emtree and MeSH when searching for drugs, a series of drug searches were performed using Embase (via [embase.com](http://embase.com)) and MEDLINE (via PubMed). The searches were performed on July 1, 2015. Each database was searched using the Emtree preferred term and the NLM Supplementary Concept data (there were no MeSH preferred terms for the drugs on the list). The results are illustrated in [Figure 2](#) and listed in [Table 2](#).

The drugs chosen were the top 10 leading products based on global pharmaceutical sales according to the PM Live Top Pharma List 2014 ([http://www.pmlive.com/top\\_pharma\\_list/Top\\_50\\_pharmaceutical\\_products\\_by\\_global\\_sales](http://www.pmlive.com/top_pharma_list/Top_50_pharmaceutical_products_by_global_sales)).

The search period was literature published between January 1, 2010 and December 31, 2014.



**Figure 2.** Comparison of the volumes of results found with drug searches in MEDLINE/PubMed and Embase.

Product name	Search term		Number of citations	
	Emtree preferred term	MeSH substance name	Embase	MEDLINE
Humira	Adalimumab	Adalimumab	11,592	1,776
Sovaldi	Sofosbuvir	Sofosbuvir	1,060	156
Remicade	Infliximab	Infliximab	16,626	2,883
Enbrel	Etanercept	TNFR - Fc fusion protein	10,702	1,697
Lantus	Insulin Glargine	Glargine	3,394	503
MabThera/Rituxan	Rituxumab	Rituximab	27,655	4,557
Avastin	Bevacizumab	Bevacizumab	22,636	4,688
Seretide/Advair	Fluticasone propionate plus salmeterol	Fluticasone, salmeterol drug combination	1,011	139
Herceptin	Trastuzumab	Trastuzumab	13,091	1,933
Crestor	Rosuvastatin	Rosuvastatin	5,465	865

**Table 2.** Results of comparative drug searches

---

## SUMMARY OF FACTS ABOUT THE TWO DATABASES

### Embase

- Biomedical database produced by Elsevier
- Comprehensive coverage of biomedicine with a focus on drugs, pharmacology and medical devices
- Indexed with Emtree, which has over 71,000 preferred terms, including over 31,000 drugs and chemicals)
- Over 30 million records from 1947 to present, currently covering over 8,500 journals
- Covers all the journals that are in MEDLINE and has 2,900 unique journals that are not in MEDLINE
- Unique records include pharmacology journals and European literature

### MEDLINE

- Biomedical database produced by U.S. National Library of Medicine
- Covers all the fields of biomedicine including nursing, dentistry and veterinary science
- Indexed with MeSH, which has over 27,000 terms
- Over 22 million records from 1947 to present, currently covering over 5,600 journals
- All MEDLINE journals are included in Embase

## CONCLUSION

While MeSH offers historical notes and extensive terminology for nursing, dentistry and veterinary science, Emtree is a more comprehensive and up-to-date thesaurus for biomedical research. In addition to its superior content, Emtree has a more refined design that ensures better discoverability. This makes Embase a better resource for applications such as pharmacovigilance, systematic reviews for evidence-based medicine and gathering information for medical device development and post-market surveillance.

---

## GLOSSARY

**Facet** - The levels of a thesaurus subject hierarchy are called facets. Each facet represents a broad category of subjects (anatomy, organisms, drugs/chemicals, diseases, medical techniques and equipment, patient groups and so on).

**Indexing** - The process of reviewing literature and adding terms to it to make it discoverable is called indexing. Fully indexed records in Embase (excluding MEDLINE records licensed from the U.S. National Library of Medicine) are manually indexed using the full text of each article. Index terms identified by trained indexers with a biomedical background are checked against Emtree before being added to the records. For licensed MEDLINE records, index terms assigned by the U.S. National Library of Medicine from the MeSH thesaurus are mapped to Emtree.

**Natural language order** - When words are in the order they would appear in an English sentence, they are in natural language order. Spinal cord injury is the natural language order, but injury, spinal cord is inverted.

**Scope note** - A note explaining the meaning of a term is called a scope note. MeSH uses scope notes to explain the intended meaning of all terms. In Emtree, scope notes are used to define the meaning and usage of checktags (e.g., human) and subheadings (e.g., adverse drug reaction).

**Term** - A word or phrase in a thesaurus is called a term. Terms are used to index literature so that it is discoverable.

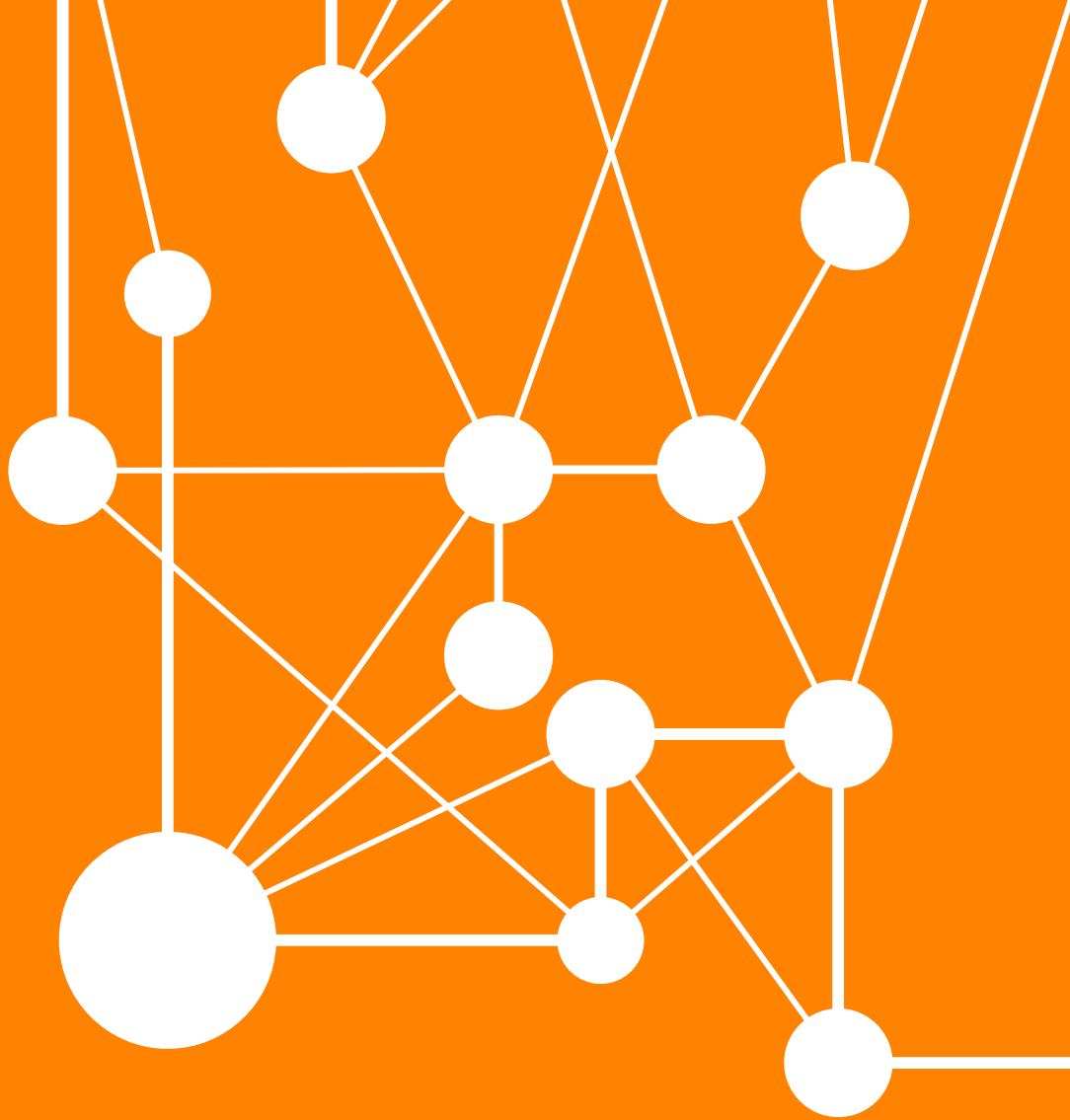
**Synonym** - A term that has the same meaning as another term is called a synonym. Cardiac infarct, myocardial infarction and heart attack are synonyms of the preferred term heart infarction in Emtree. Synonyms (or entry terms) are mapped to the preferred term in Emtree and can be used both to identify preferred terms. They can also be used to retrieve articles in which the preferred term is indexed.

**Preferred term** - The Emtree term that is used for indexing is called the preferred term. These usually have a number of synonyms that can be used as entry points in Emtree to identify them. Search syntax is available on most systems to enable the use of either preferred terms or synonyms to retrieve articles in which the preferred term is indexed.

**Explosion search** - A search that retrieves all the information indexed to a preferred term and to all its narrower terms is called an explosion search.

**Narrower term** - A preferred term that is hierarchically related to (and more specific than) another preferred term is called a narrower or child term. For example, brain injury and spinal cord injury are narrower terms for nervous system injury.

**Search term** - The word or phrase that a user of Embase or PubMed uses in constructing a search is called a search term.



## LEARN MORE

To request information or a product demonstration, please visit [elsevier.com/embase](http://elsevier.com/embase) or email us at [embase@elsevier.com](mailto:embase@elsevier.com).

Visit [elsevier.com/rd-solutions](http://elsevier.com/rd-solutions) or contact your nearest Elsevier office.

### ASIA AND AUSTRALIA

Tel: +65 6349 0222

Email: [sginfo@elsevier.com](mailto:sginfo@elsevier.com)

### JAPAN

Tel: +81 3 5561 5034

Email: [jpinfo@elsevier.com](mailto:jpinfo@elsevier.com)

### KOREA AND TAIWAN

Tel: +82 2 6714 3000

Email: [krinfo.corp@elsevier.com](mailto:krinfo.corp@elsevier.com)

### EUROPE, MIDDLE EAST AND AFRICA

Tel: +31 20 485 3767

Email: [nlinfo@elsevier.com](mailto:nlinfo@elsevier.com)

### NORTH AMERICA, CENTRAL AMERICA AND CANADA

Tel: +1 888 615 4500

Email: [usinfo@elsevier.com](mailto:usinfo@elsevier.com)

### SOUTH AMERICA

Tel: +55 21 3970 9300

Email: [brinfo@elsevier.com](mailto:brinfo@elsevier.com)