# Emtree Vocab Datasheet

## Context-aware computational search

Scientists today are confronted with overwhelming volumes and sources of data to analyse. Whilst computational searching is possible, this quickly proves to be inadequate when tackling the deluge of scientific text with variations of spellings and synonyms.

To help address this problem, SciBite has created a series of optimized vocabularies that enable computers to uncover relevant information. Based on public ontologies or reference databases from a wide range of topics, our curation team transform and further enrich these ontologies into specialized vocabularies which are consumed by our entity extraction engine, TERMite. These capabilities provide our customers with the ability to identify and extract relevant scientific information from millions of records at the click of a button.

### Enrich your data with Emtree

SciBite is proud to offer Elsevier's Emtree thesauri as a SciBite VOCabulary to be used alongside our existing VOCabs. With Emtree, users can index their own sources using these standards via SciBite's proven technology, in pursuit of FAIR, harmonized data.

- **COMPREHENSIVE** – Unrivalled coverage of drug, disease and medical device terminology
- **CONTEXT-AWARE** – Enriched semantic content ensures maximum capture of subject area and context
- **GOLD STANDARD** – Access terms registered with regulatory bodies including WHO, EMA and FDA.

### Emtree VOCab pack

Originally created by subject matter experts for the biomedicinal and life science community, Emtree includes 90,000 drug, disease, medical device and essential life science entities. These thesauri are further enriched and optimized using our proprietary tools and are manually curated with rules and additional vocabularies to create the Emtree VOCab pack with 2.5 million synonyms, which when coupled with our entity extraction engine, TERMite, can index text-based documents to generate a rich, annotated dataset that aligns to Emtree terminology. The Emtree VOCab pack includes a full version of Emtree, as well as separate medical device and combination drug vocabularies that incorporates additional SciBite content.

The Emtree vocabulary includes:
- All drug generic names described by the FDA and EMA
- All drug international non-proprietary names (INNs) described by the WHO
- Drug trade names for most major pharmaceutical companies
- Medical device trade and general names, plus medical procedures
- Global medical device nomenclature (GMDN) names used by the FDA
- Diseases, organisms, biological functions, and medical and biological parameters
- Terms for traditional Chinese medicine
- All terms from the MEDLINE thesaurus MeSH

Established in 1947, Emtree continues to be updated three times a year with new terms included primarily on the frequency of use as candidate terms. Emtree's comprehensive hierarchy provides an intuitive context to the relationship of terms within the VOCab and allows users to filter results with any branch of the thesauri, as shown in Figure 1.
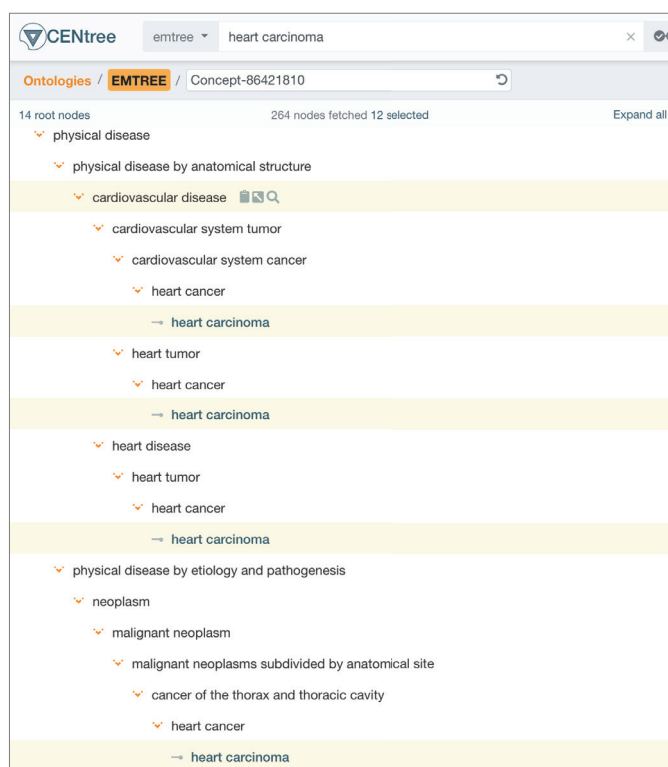


*Figure 1. Emtree comprehensive hierarchical structure.*

## Deploy Emtree in your use cases

As with all SciBite VOCabs, the Emtree module can be used on its own or in conjunction with our other VOCabularies depending on your project requirements. Emtree can support a host of use cases, as described in Table 1.

| Use case | Description |
|---|---|
| Clean data entry | Create application ontologies from Emtree thesauri (drug, compounds, disease etc.) for use in LIMS or ELNs |
| Search | Disambiguate biomedical entities in clinical literature, company websites, press releases etc. for competitive intelligence analysis |
| Current Awareness | Deliver integrated information from search to keep workgroups up-to-date with their field |
| Knowledge graph | Connect compounds with their activity on a target or role for drug repurposing or to identify potential adverse events |
| Pharmacovigilance | (Semi-)automate monitoring of products for safety signals via search i.e., track drug-disease relations or drug-drug interactions |
| Regulatory submissions | Include terminology used by major regulatory bodies like the EMA and FDA |
| Medical devices | Perform clinical efficacy investigations of medical devices using Emtree's extensive lists of medical devices (e.g., endoscopes, catheters, prostheses) |
| Clinical studies | PICO (Patient/ Problem, Intervention, Comparison and Outcome) effectiveness studies, as well as post-market surveillance |
| Model | Create entity recognition AI models |

*Table 1. How Emtree can be deployed in your use cases*

Emtree can be used to annotate biomedical hits in unstructured data sources using TERMite, SciBite's named entity recognition tool (see Figure 2).



*Figure 2. Context-aware enrichment – use TERMite, SciBite's named entity recognition tool, to annotate unstructured data sources with Emtree.*

## About SciBite

SciBite is an award-winning semantic software company offering an ontology-led approach to transforming unstructured content into machine-readable, clean data. Supporting the top 20 pharma with use cases across life sciences, SciBite empowers customers with a suite of fast, flexible, deployable API technologies, making it a critical component in scientific data-led strategies. Contact us to find out how we can help you get more from your data.