Datasheet

Enhancing semantic search and discovery

Explore TERMite

- ULTRA-FAST: indexes at up to 1 million words per second
- SMART: detects ambiguity in natural language, supports multiple in/output formats
- SCALABLE & FLEXIBLE: powerful API that embeds into your workflow

Text, text, and more text

In biomedical science and healthcare, text remains the most common medium for recording and sharing information. In the digital age, the ease of contributing text has led to an explosion of content that tends to be less structured than traditional formats (e.g. Medline articles, patent documents). Scientists nowadays need better tools to not only narrow down the search space, but also recognise and extract nuggets of key information from text for spotting hidden trends, hypothesis generation, and business intelligence. Frustratingly, many existing tools have a steep learning curve or aren’t tuned for analyzing biological text. How can SciBite help?

Scan fast, tag right

We developed TERMite (TERM Identification Tagging and Extraction), a named entity recognition (NER) tool that rapidly scans and semantically annotates raw text (up to 1 million words per second) with entities from over 50 biopharma and biomedical topics. The entities are drawn from VOCabs, which is SciBite’s flagship collection of manually curated vocabularies with >20 million synonyms, specifically tuned for NER text analytics. Since VOCabs are based on public ontologies, such as gene ontology, tagging entities with TERMite will map your data to those ontologies, align data to public standards, and improve data interoperability.

Text can come from many disparate sources

- electronic lab books
- social media posts
- doctors’ notes
- RSS feeds
- spreadsheets of free-text survey responses
- clinical trial records
- comments in patient’s forums
- internal databases

In biomedical science and healthcare, text remains the most common medium for recording and sharing information. In the digital age, the ease of contributing text has led to an explosion of content that tends to be less structured than traditional formats (e.g. Medline articles, patent documents). Scientists nowadays need better tools to not only narrow down the search space, but also recognise and extract nuggets of key information from text for spotting hidden trends, hypothesis generation, and business intelligence. Frustratingly, many existing tools have a steep learning curve or aren’t tuned for analyzing biological text. How can SciBite help?

Scan fast, tag right

We developed TERMite (TERM Identification Tagging and Extraction), a named entity recognition (NER) tool that rapidly scans and semantically annotates raw text (up to 1 million words per second) with entities from over 50 biopharma and biomedical topics. The entities are drawn from VOCabs, which is SciBite’s flagship collection of manually curated vocabularies with >20 million synonyms, specifically tuned for NER text analytics. Since VOCabs are based on public ontologies, such as gene ontology, tagging entities with TERMite will map your data to those ontologies, align data to public standards, and improve data interoperability.
Understands nuances
Tagging the correct terms is only the first step. TERMite has built in relevance detection, distinguishing between terms that are “throwaway mentions” and those that really matter to your subject of interest.

Powerful API for any use case
TERMite is our core platform from which we deliver additional capabilities for specific use cases. Developed in Java, the hassle-free set-up typically takes only a few minutes. Once installed, you will have the freedom to run one-off analyses on the user interface, or embed TERMite into your analysis workflow to harness its full potential. Popular use cases include:

- Mine the entire Medline database for gene-phenotype-disease correlations
- Analyse grants to discover new trends
- Scan internal documents to find hidden target-drug-indication relationships for drug repurposing

The possibilities are endless. What's yours?

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.