SciBite
an ELSEVIER company

Use case

Enabling better data discoverability:
Leveraging semantics for intelligent scientific search

Enabling better data discoverability:
Leveraging semantics for intelligent scientific search

SciBite Search delivers intelligent scientific search to everyone, you don't have to be a search expert. Its modern, easy to use interface provides scientists with access to domain specific ontology and AI-powered search capabilities to quickly find the answers they need while also providing sophisticated text mining capabilities for deeper and more exploratory queries.

SciBite Search enables users to search both unstructured and structured information across:
• Public biomedical sources, such as Medline and Clinical Trials.gov,
• Subscription content from providers such as Elsevier, Wiley and Springer, and
• Internal and external documents.

Document-level security and role-based permissions ensure that access to proprietary information is controlled.

Here we highlight just a selection of the many applications of SciBite Search.

Intelligent scientific search, available to everyone

Most search applications are limited to rudimentary text and keyword searches and lack the scientific intelligence needed to answer common research questions.

For example, a search for the gene PDE5A would miss references to synonyms such as CGB-PDE, Phosphodiesterase 5A and CN5A.
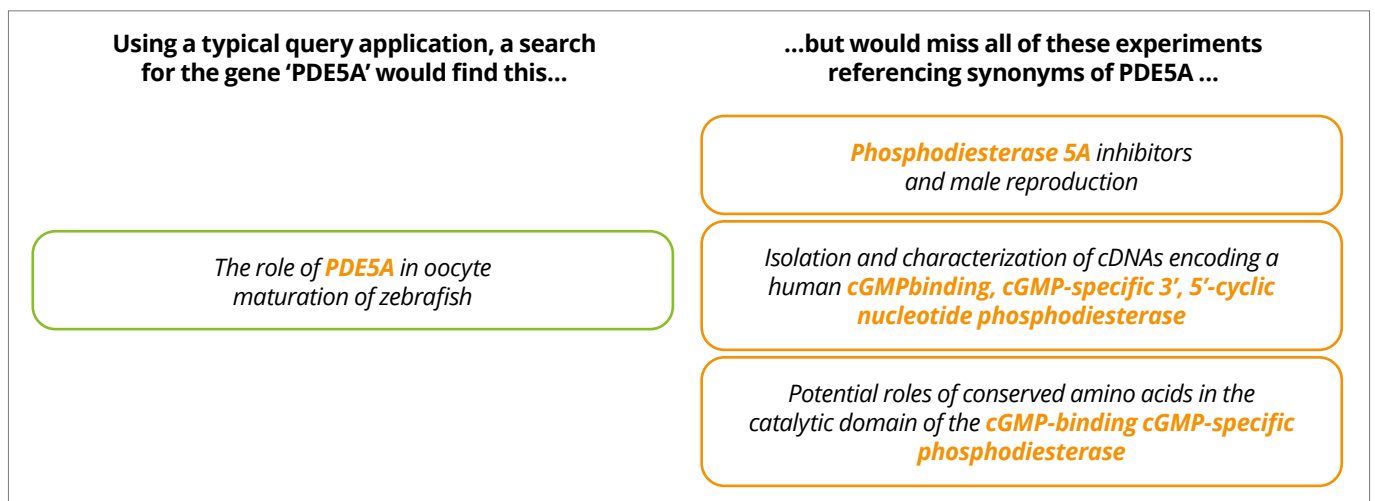
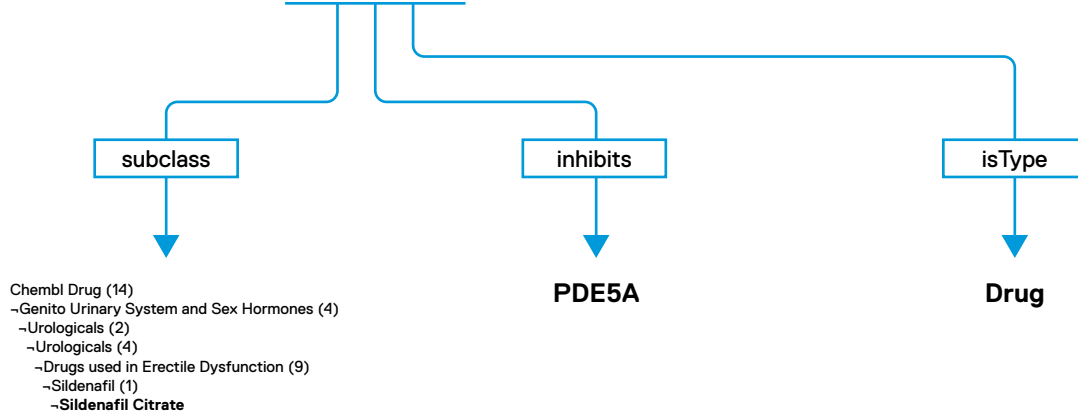| Using a typical query application, a search for the gene 'PDE5A' would find this...	...but would miss all of these experiments referencing synonyms of PDE5A ...
	Phosphodiesterase 5A inhibitors and male reproduction
The role of **PDE5A** in oocyte maturation of zebrafish	Isolation and characterization of cDNAs encoding a human **cGMPbinding, cGMP-specific 3', 5'-cyclic nucleotide phosphodiesterase**
	Potential roles of conserved amino acids in the catalytic domain of the **cGMP-binding cGMP-specific phosphodiesterase**

Figure 1: Most search tools miss synonyms of search terms

At the core of SciBite Search are our scientific vocabularies, or VOCabs, and Named Entity Recognition (NER) engine, TERMite. SciBite's team of experienced ontologists create and maintain VOCabs by augmenting public ontologies using a combination of manual and machine curation to ensure quality and accuracy and ensure they have comprehensive synonym coverage. Using these VOCabs, TERMite rapidly processes all ingested content and applies an explicit, unique meaning and description to scientific terms and concepts found within the text, contextualizing them so that they can be understood as "things, not strings"1. For example, SciBite Search understands that PDE5A is a protein rather than just a collection of characters.

This ensures that all relevant data will be found, regardless of which synonym is used as the search term.
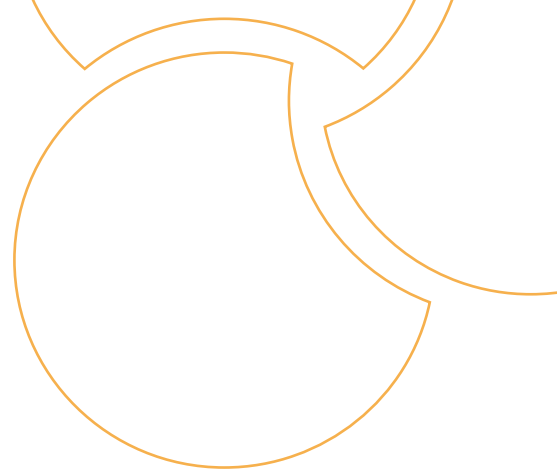
This contextualisation also enables SciBite Search to answer semantic questions. For example, an article describing "The Effect of Sildenafil on Selenite-Induced Cataract in Rats" would be found by a simple text search using any of the terms in the sentence. However, SciBite Search understands, for example, that Sildenafil is a drug which inhibits the target PDE5A and that rats are rodents. So, a query for 'phosphodiesterase inhibitors in rodents' would return exactly the same article.

Things, not strings

"The effect of **Sildenafil** on selenite-induced cataracts in rats."

| subclass | inhibits | isType |

Chembl Drug (14)
¬Genito Urinary System and Sex Hormones (4)
 ¬Urologicals (2)
  ¬Urologicals (4)
   ¬Drugs used in Erectile Dysfunction (9)
    ¬Sildenafil (1)
     ¬**Sildenafil Citrate**

PDE5A

Drug

Figure 2: SciBite Search understands scientific terms as "things not strings"

1   https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html

SciBite Search also enables scientists to use natural language scientific queries, surfacing powerful query capabilities while hiding the complex logic from end users. This ranges from questions that have a 'yes/no' answer to those that generate set of facts and those that require aggregation of information, to the such as:

• Is Sildenafil an inhibitor of Phosphodiesterase 5A?
• What are the known inhibitors of Phosphodiesterase 5A?
• What is the top drug for Phosphodiesterase 5A?
• What are the top 5 journals for cancer research?
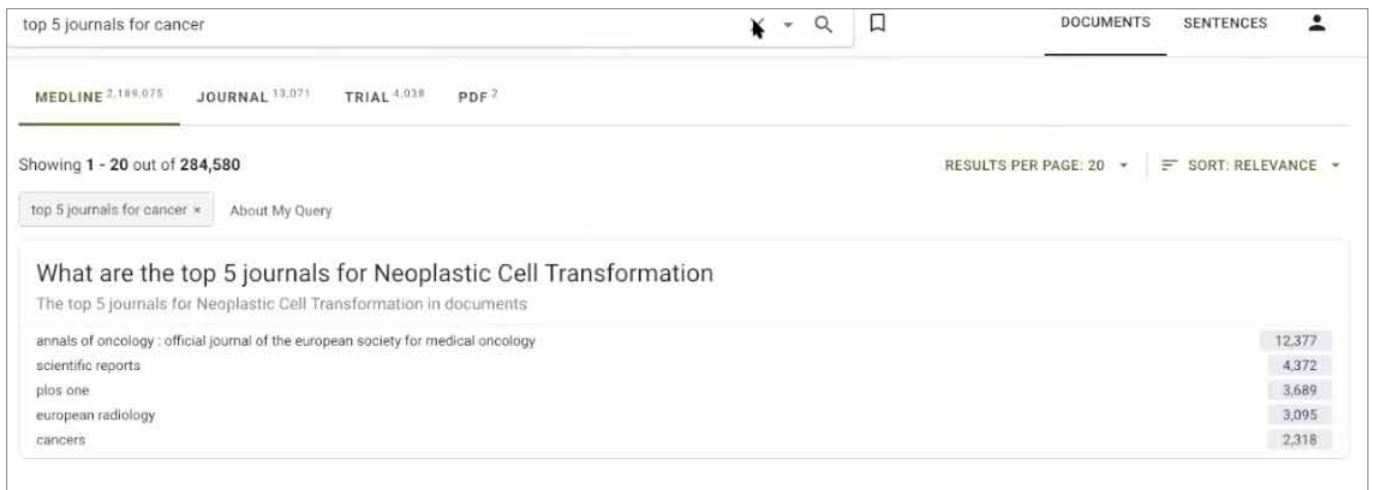
Figure 3: *SciBite Search enables users to easily answer complex questions*

Deeper search and exploration

As well as improving the search experience for scientists, SciBite Search also provides powerful text mining capabilities to support more sophisticated use cases.

Document search tools are limited to the content of the document index. For example, such tools will only know that Viagra is made by Pfizer if this relationship is explicitly stated within a document. SciBite Search leverages knowledge graphs and deep learning to augment the scientific search experience, enabling users to make connections between scientific concepts and find information based on relationships.

For example, users can ask questions that rely on the metadata of the entities mentioned in documents, such as looking for "pharmaceutical companies in the south of the United Kingdom mentioned in the same sentence as cancer".

Users can perform boolean queries and quickly navigate to the most interesting results based on taxonomies. SciBite Search also enables users to instantly identify terms that co-occur within a sentence or within a document. For example, by generating a list of genes which are mentioned most frequently with a disease of interest, potential new avenues for research can be revealed.
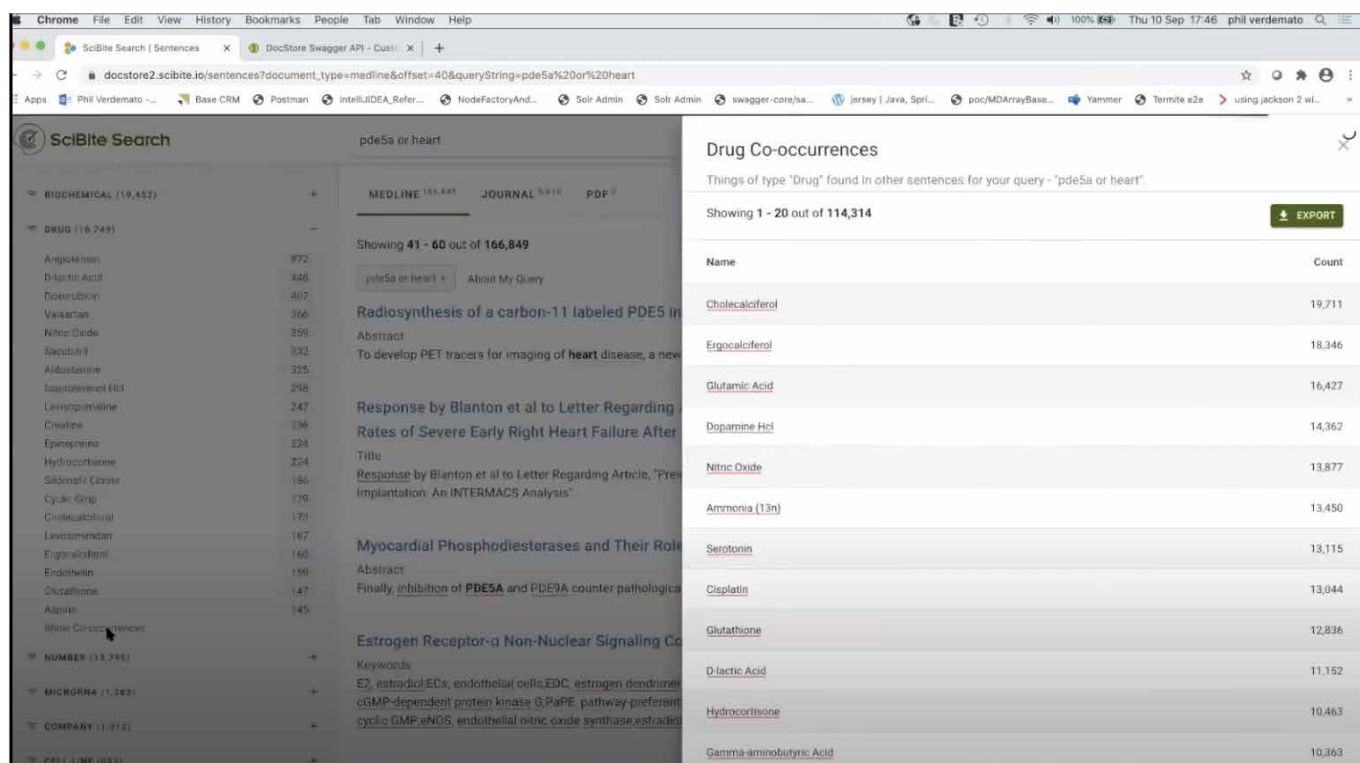
Figure 4: *A ranking of the drugs that co-occur most often with a target of interest*

Example use cases

## Extract more insights from electronic notebooks

Most Life Sciences organisations will have, at some point, deployed an Electronic Laboratory Notebook (ELN) with the goal of centralising R&D data. ELNs have become an important source of not only key experimental data points but also the development of methods and SOPs. However, much of the information stored within an ELN is captured as qualitative free text or as a collection of attachments and, as a result, the ability to mine it is typically limited to rudimentary text and keyword searches.

SciBite Search opens up new possibilities to search the ELN data more effectively and derive valuable insights. It accurately marks-up all relevant terms and concepts present within an experiment, without being limited by the indexing terms used by the ELN.

SciBite Search not only provides a better search experience for ELN data, it also enables users to answer more complex questions, such as those outlined below.

Examples of questions that can be rapidly answered with SciBite Search

• Find all references to project ABC-101, regardless of the syntax used by the author (e.g. ABC-101, ABC101 and ABC 101)

• Find all experiments for a specific target across the organisation, regardless of which synonym was used by the author of the experiment.

• Which projects are investigating potential biological therapeutics?

• Which targets have we studied that are associated with inflammatory disorders?

• Which diseases have we studied for both a target of interest and other targets in the same class and what were the outcomes?

• Which pre-clinical studies have utilised a specified mouse model?

• Which experimental techniques are growing across the organisation and would benefit from a core facility?
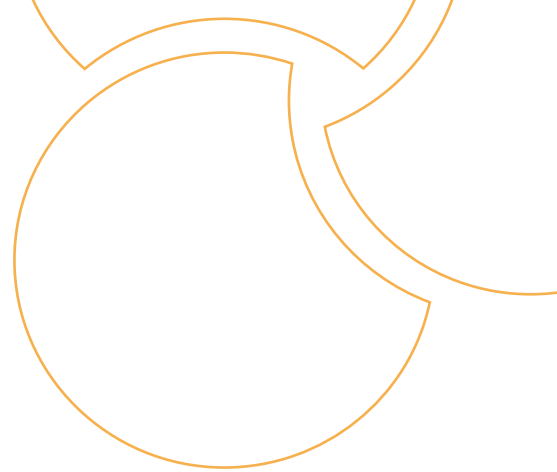
Most ELNs only enable users to search for exactly what was written by the author of an experiment. The inconsistent use of synonyms during data entry makes it difficult to identify and collate all relevant data for a disease or target of interest. For example, an experiment describing work on 'muscarinic acetylcholine receptor M1' will not be found by a scientist who performs a search using the commonly used synonym 'cholinergic receptor muscarinic 1'. Even for more defined entries, the meaning of a particular field or its contents may be ambiguous, imprecise or contain multiple different data types, such as Gene, Assay Type and Species.

For more information, read the full use case.

Unlock the potential of departmental scientific documents

Much of the knowledge found within departments or cross-functional project teams is contained in the plethora of documents, reports and emails that they produce and the scientific articles they have downloaded. The unstructured nature of these files, the range of formats used and the fact that they are typically spread across different locations limits the ability to search them for useful information.
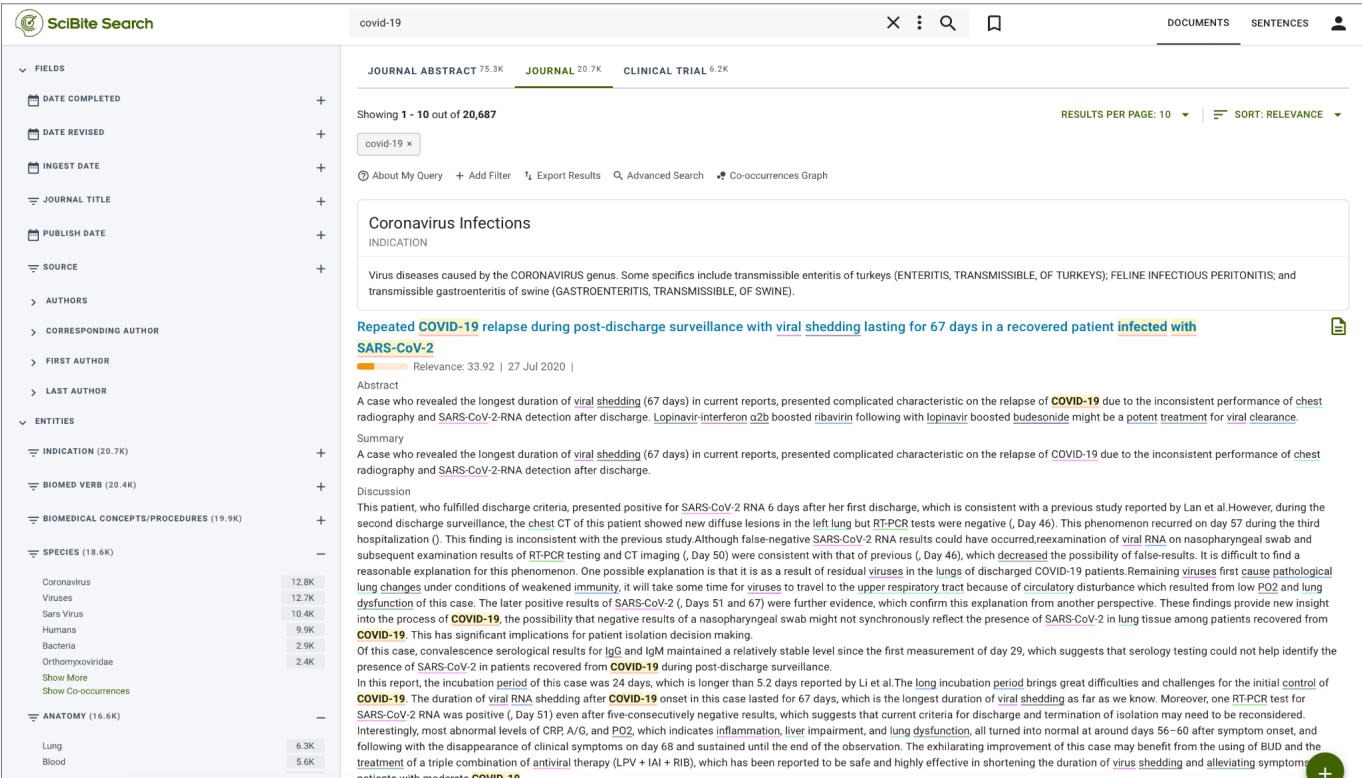
Even where such documents are organised in a file store, or structured in a document management system or SharePoint, such systems typically offer a very poor search experience for users, usually limited to exact matches of what was written by the document author. Similarly, inconsistent use of synonyms when writing a document makes it difficult to subsequently identify and collate all relevant data for a disease or target of interest.

SciBite Search is a fully fledged scientific search system which delivers intelligent query capabilities for individual teams or entire organisations. It's simple to deploy and easy for departments to upload their own documents whilst ensuring access is controlled appropriately.

Not only does SciBite Search make it simpler to interrogate the information managed within these documents, it also enables users to answer more complex ontology-based questions and find all documents related to a specific topic e.g. all those that talk about mode of action, all those that are about toxicology, etc.

SciBite Search injects semantic annotations into PDFs, whilst retaining their original format, making it easy for users to instantly get a feel for what a document is about.

For more information, read the full use case.

Figure 5: Summarising all Terms Identified within a document (left hand side of the screen)

SciBite

an ELSEVIER company

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. At SciBite we believe data fuels discovery and we are leading the way with our pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. Supporting the world's leading scientific organisations with use-cases from discovery through to development, SciBite's suite of fast, flexible, deployable API technologies empower our customers, 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.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at contact@scibite.com