SciBite
an ELSEVIER company

Whitepaper

Ontology Management: A Simple, Collaborative and Robust Approach

SciBite
an ELSEVIER company

Ontology Management: A Simple, Collaborative and Robust Approach

Ontologies play a critical role in semantic enrichment, enabling unstructured scientific text to be transformed into clean, contextualised data which can be understood and exploited by computational approaches, such as machine learning. However, the process of managing these ontologies presents a challenge to most Life Sciences organisations. Maintaining multiple, evolving ontologies from both public and proprietary sources, requires significant effort from ontology experts.

SciBite's mission is to simplify both the management and use of ontologies. In our vision of a world with effective ontology management, there would be a centrally available source of information for all users, regardless of their particular use case or application. Ontologies would blend both public and internal terminology to facilitate their use across multiple different groups. Scientists would become an integral part of the process by requesting updates to ontologies. All editing would be in a purpose-built user interface with no extraneous or unnecessarily complex functionality. Plus there would be a fully traceable log to manage evolving versions of an ontology.

Here we expand upon this vision and describe SciBite's integrated approach to simple, collaborative and robust ontology management.

Centralised Ontology Management

There are often several independently developed ontologies within a single domain which often contain overlapping and even conflicting information. In addition, many pharmaceutical companies also have their own proprietary lists, such as project codes and compound IDs.

SciBite uses open standards, ensuring there is no 'lock in' to proprietary language or formats. Users can upload a locally-stored ontology file and to pull from public repositories whilst specifying if an entire ontology or a more specific subset of branches is imported.

SciBite also provides version control between releases of ontologies, including the provenance of the change. For example, if users edit public ontologies, the changes can be reconciled with new releases. A merged view of the combined ontology highlights any conflicts, enabling curators to focus on things that need to be addressed.

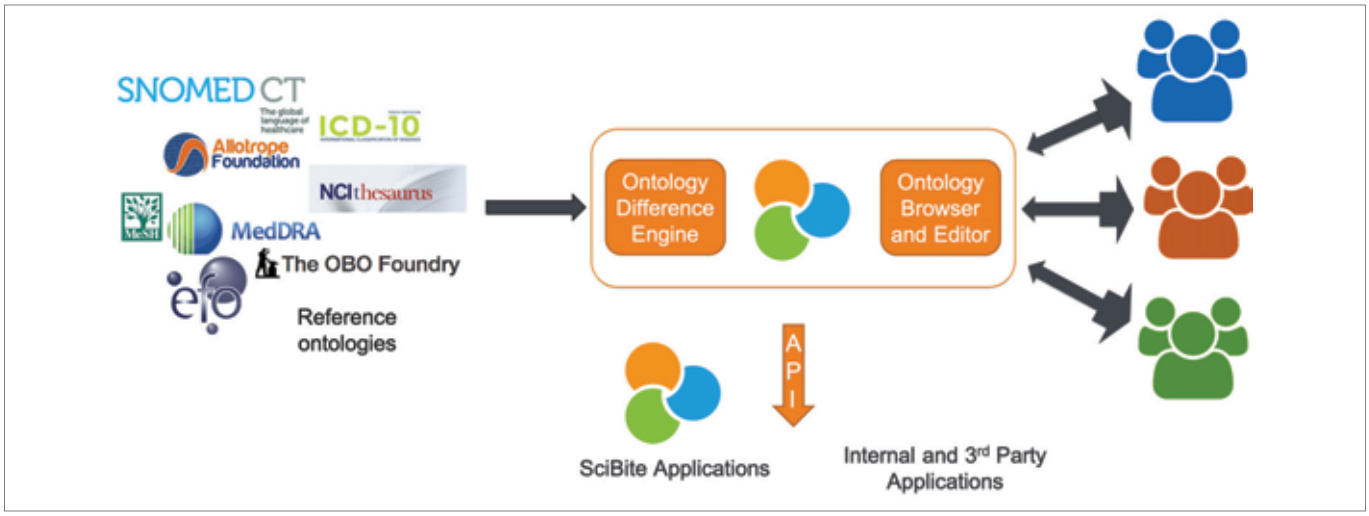

Figure 1: Integrated ontology management

Simplified Visualisation and Curation

As with most visualisation techniques, there is no single 'best' way to view an ontology that fits all users and all applications. Some applications benefit from a tree view, while views providing an easily digestible high-level summary, such as heat maps or bubble diagrams will make more sense for others. For instance, in the example illustrated below, Crohn's disease is a subclass of digestive system disease.

SciBite provide a simple interface to navigate and make common changes to ontologies, lowering the barrier for users to contribute to editing and reducing the bottleneck of having to rely on ontology experts.

Users can add, replace or remove an ontology class, change its location in a hierarchy to reflect a new sub-class relationship, and enrich it with new metadata such as synonyms or text definitions.

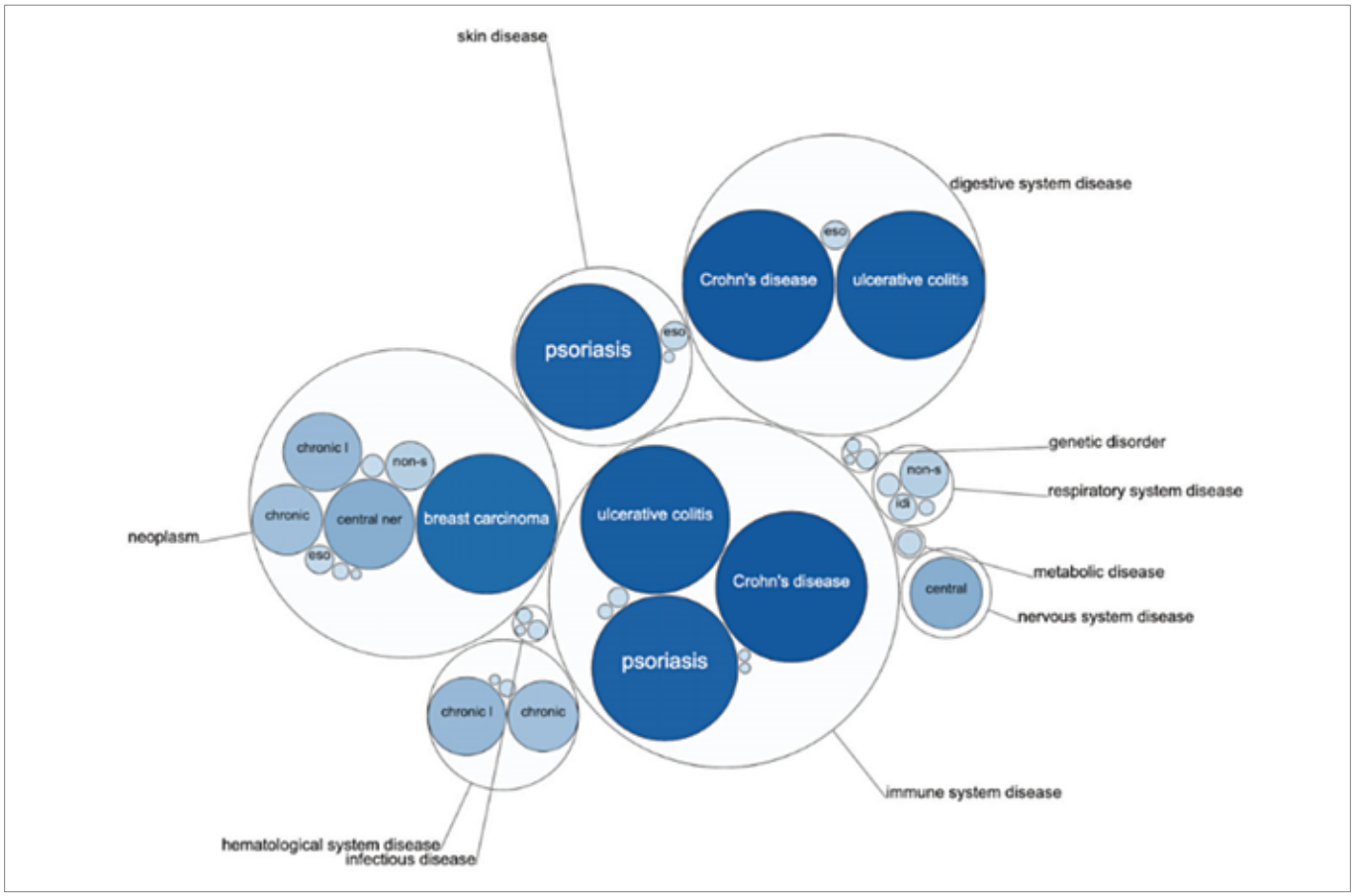

Figure 2: An example of Ontology hierarchies represented as a Venn diagram. Taken from the OpenTargets project.

Facilitating Collaboration and Integration

It is increasingly common for there to be multiple contributors, often from different locations. However, ontologies are often managed by applications that lack traceability, which hinders collaborative working. From the outset, SciBite's ontology management capabilities were built from the perspective of a multi-user environment.

For example, governance and provenance are an intrinsic part of the editing process, with the ability to manage permissions and prevent issues associated with the same term being edited by two different people simultaneously.

Ontologies can be accessed programmatically via an Application Programming Interface (API), not only by SciBite's applications, such as TERMite, our named entity recognition (NER) and extraction engine, but also by third party tools. This enables an 'ontology as a service' approach, ensuring that all applications make consistent use of up-to-date ontologies.

Ontology Enrichment

To keep ontologies up to date, curators typically need to maintain awareness of the evolving language in their particular domain, particularly in new scientific fields, such as "CRISPR", "10x Genomics" and "Drop-Seq". This is a laborious, manual task, which can result in potentially important terms or changes being missed.

To simplify this process, SciBite leverages machine learning techniques to support the enrichment process in a controlled way. For example, to suggest terms that are similar to a word of interest because they are used in a similar context.

We also leverage machine learning approaches to manage the disambiguation process by identifying words that commonly co-occur with a term of interest and give users the option to use these as positive and negative 'boosters' that add context to a term. For example, the word 'chip' has multiple meanings, including a foodstuff, a computer component and a microarray. However, when 'chip' used in combination with the word 'DNA' then its ambiguity is significantly reduced. Users can also specify where to look for such 'boosters', such as limiting the search for common co-occurrences of words to scientific sources such as MEDLINE or more widely, for example to include news sources.

We're always evaluating innovative technologies and continue to explore additional applications of machine learning to simplify ontology management. For example to identify gaps in ontologies such as a missing branch. This concept has the potential to be extended to use seed words as the basis of automatically building a new ontology automatically and fill gaps where no public ontology currently exists, for example for clinical measurements such as body mass index.

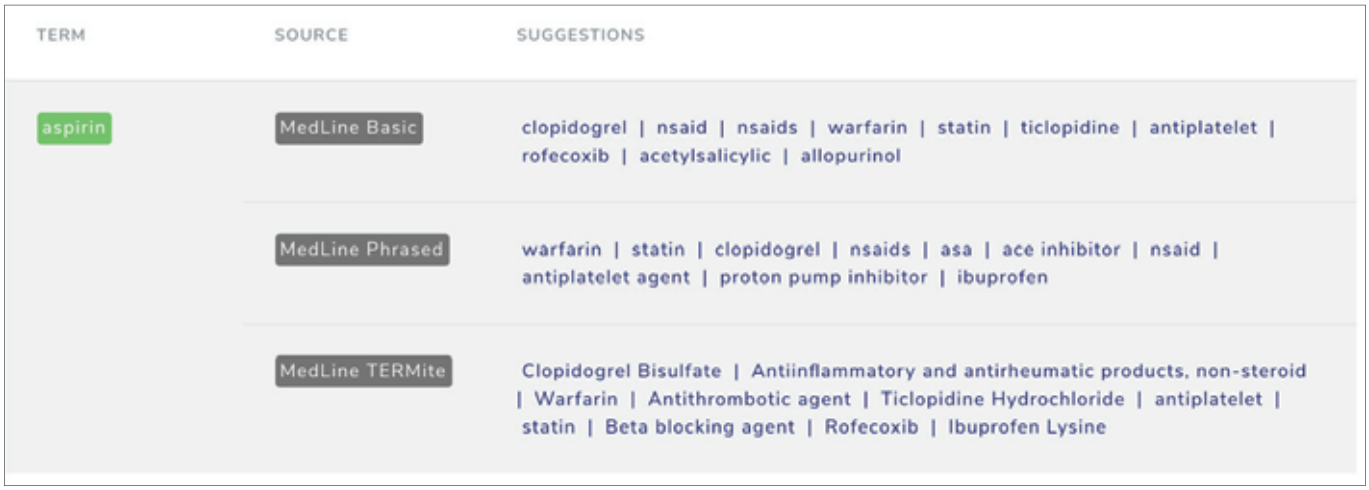

Figure 3: Term suggestion

Summary

The challenges associated with managing an evolving ontology are amplified when organisations rely on multiple ontologies from different sources. SciBite addresses this problem with simple to use ontology management capabilities built on a robust enterprise-ready platform.

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

Head Office:
SciBite Limited
BioData Innovation Centre
Wellcome Genome Campus
Hinxton, Cambridge CB10 1DR
United Kingdom

www.scibite.com
contact@scibite.com
LinkedIn: SciBite
Twitter: @SciBite
+44 (0)1223 786 129