

Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature.

The biomedical literature continues to grow at a rapid pace, making the challenge of knowledge retrieval and extraction ever greater. Tools that provide a means to search and mine the full text of literature thus represent an important way by which the efficiency of these processes can be improved. We have developed the next generation of the Textpresso information retrieval system, Textpresso Central (TPC). TPC builds on the strengths of the original system by expanding the full text corpus to include the PubMed Central Open Access Subset (PMC OA), as well as the WormBase *C. elegans* bibliography. In addition, TPC allows users to create a customized corpus by uploading and processing documents of their choosing. TPC is UIMA compliant to facilitate compatibility with external processing modules, and takes advantage of Lucene indexing and search technology for efficient handling of millions of full text documents. Like Textpresso, TPC searches can be performed using keywords and/or categories (semantically related groups of terms), but to provide better context for interpreting and validating queries, search results may now be viewed as highlighted passages in the context of full text. To facilitate biocuration efforts, TPC also allows users to select text spans from the full text and annotate them, create customized curation forms for any data type, and send resulting annotations to external curation databases. Future steps of development include a set of robust text mining modules that can be applied to full text articles by the user through an work flow manager. TPC is an online literature search and curation platform currently featuring over 1.5 million full text articles and can be accessed at <http://www.textpresso.org/tpc>.