DATASHEET

TEXT ANALYTICS COMPONENT COLLECTION

The Text Analytics Collection for Pipeline Pilot brings together the utility of literature search and text mining with the power of process automation and data integration in Pipeline Pilot. Achieve your most challenging text mining objectives by linking together search and characterization steps into automated routines. Integrate literature mining with your existing scientific protocols, and run them interactively or automatically every night.

WITH THE TEXT ANALYTICS COLLECTION, YOU CAN:

- Search PubMed, US and European Patents, NIH grants, Google, Yahoo, or local files, and integrate with other third party search engines
- Use a single query language with phrase, wildcard, fielded, and synonym matching
- Extract key concepts or find correlations in documents and online literature
- Enhance existing chemistry or bioinformatics routines with annotations or correlations from online literature
- Discover emerging trends and new relationships, analyze patents in your area, and track funding

SEARCH ONLINE DOCUMENTS

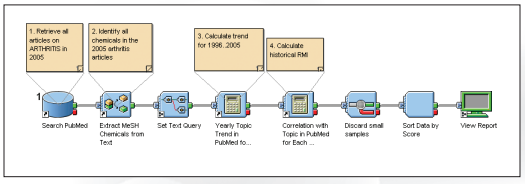
The Text Analytics Collection gives you the power to extract knowledge from important online document resources such as PubMed, US and European Patents, Wikipedia and Yahoo (user extendable to other remote text data sources). Search these databases with interactive queries, or mine them with large-scale document retrieval and characterization routines.

SEARCH LOCAL DOCUMENTS

You can also search and mine your locally stored documents in exactly the same way. The Text Analytics Collection indexes and searches folders that contain PDF, Microsoft Office, HTML, or text files on your local disk (extendable to other file formats). You can even store the results of online searches in local repositories for speedy retrieval and postprocessing. Local databases of documents stay current automatically by monitoring the folder contents for the introduction of any new or edited documents.

ANNOTATE SCIENTIFIC RESULTS

When reporting the results of pipelined data analyses, it is often useful to include additional information about the output data points. With the Text Analytics Collection, you can easily add a few steps at the end of any Pipeline Pilot protocol and have each data point serve as a query to search a



database of literature. For example, after clustering a set of genes with the Sequence Analysis Collection, you can annotate each gene with summary information from its top reference in PubMed (and a link to further search results). This kind of enhancement makes for more easily interpretable results.

IDENTIFY EMERGING TRENDS

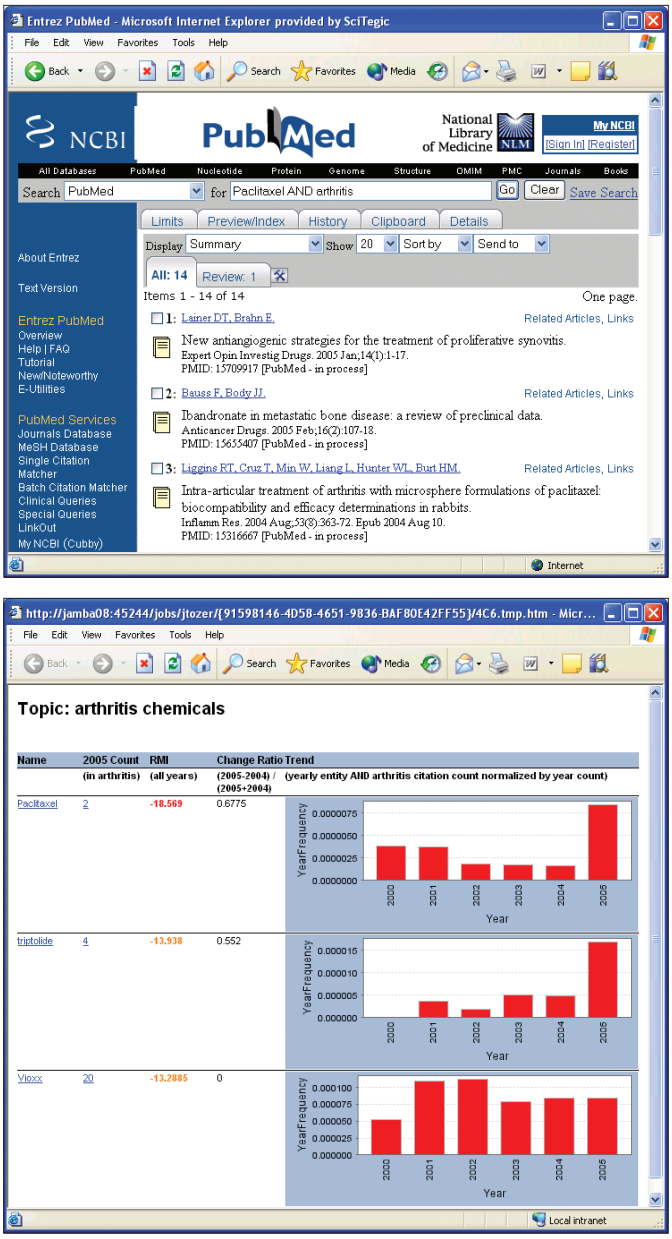
The Text Analytics Collection can monitor the scientific literature for topics of interest, and it can even alert you when new concepts are emerging for those topics. The latter is achieved by searching for new articles about your topic of interest and detecting the concept words they contain. The association of each concept with the topic of interest is calculated over time to detect emerging new relationships. This allows you to stay on top of a broader class of topics, and learn about breakthroughs before they become widely known.

MINE PATENT DATABASES

The Text Analytics Collection provides you with the tools necessary to characterize research and intellectual property trends in a field of interest. You can search and process the U.S. patent databases (extendable to other patent databases) for trends reflecting the quantity of patents, application areas, companies engaged, and more. For example, by building a protocol to process patents in the field of fuel cells, you can discover how rapidly this emerging field is growing. You can also see that applications for automobiles have come to dominate the area, and that Honda and General Motors are leading innovators.

ABOUT PIPELINE PILOT

Pipeline Pilot is an enterprise scalable scientific informatics platform that enhances research and development organizations' ability to innovate by uncovering scientific value locked in disparate data silos, automating scientific workflows, and facilitating collaboration throughout the wider scientific community. Pipeline Pilot's Component Collections contain the "scientific building blocks" of the platform and are grouped by category of science or function. By graphically combining components, you can construct workflows for data retrieval, filtering, analysis, and reporting.



Find important documents in the scientific literature (or your local files), detect and extract key concepts, and derive correlations and trends that may provide new insights.

To learn more about Pipeline Pilot, go to **accelrys.com/pipeline-pilot**