DASSAULT SYSTEMES

BIOVIA

DOCUMENTS AND TEXT COLLECTION

The Documents and Text Collection for Pipeline Pilot brings together the utility of document search, analysis and manipulation with the power of process automation and data integration in Pipeline Pilot. The Doc and Text collection uses natural language processing, advanced statistical modeling, and subject-specific ontologies and thesauri to uncover patterns and extract information from text content. Additionally, it provides extensive capabilities for managing and manipulating documents. The Documents and Text Collection can help you achieve your most challenging document analysis and manipulation objectives by linking together search, analysis and reporting steps into automated routines. Integrate literature mining and text analytics with your existing scientific protocols, and run them interactively or automatically every night.

THE DOCUMENTS AND TEXT COLLECTION, PROVIDES A SUITE OF CAPABILITIES FOR:

Documents and Search

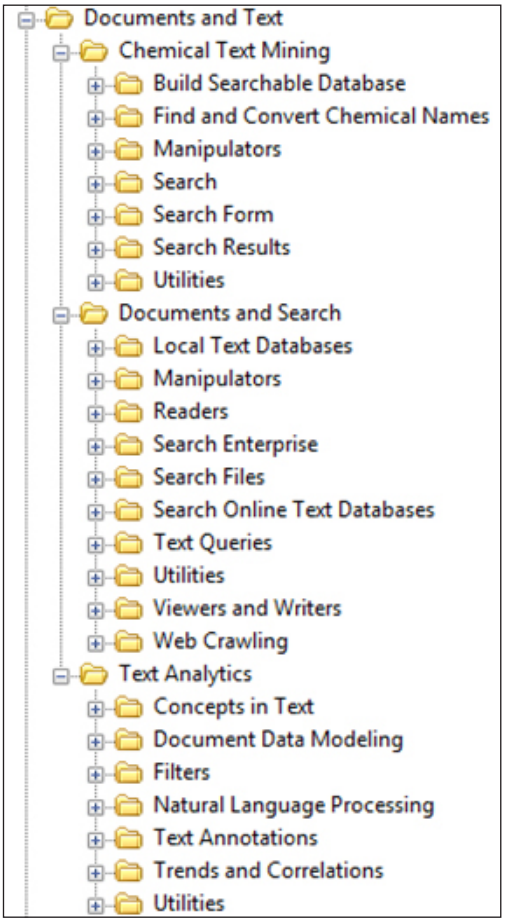
- Search PubMed, US and European Patents, Twitter, Bing, websites, SharePoint documents, local files and enterprise databases and integrate with third-party search engines, to find the documents of highest value and importance
- Search multiple data sources easily using a single query language with phrase, wildcard, fielded and synonym matching, allowing searches to be created and maintained efficiently
- Create searchable databases of key documents for ongoing review during the course of a project
- Create rich, interactive reports to explore and mine your documents and text analyses, including specialist visualizations such as Tag Clouds, highlighted terms in reports and document clustering
- Generate Word reports from templates by automatically adding content to Microsoft Word documents to streamline the creation of internal reports and external regulatory documents, reducing error-prone manual document editing
- Do web crawling to extract content and tabular data from the Web

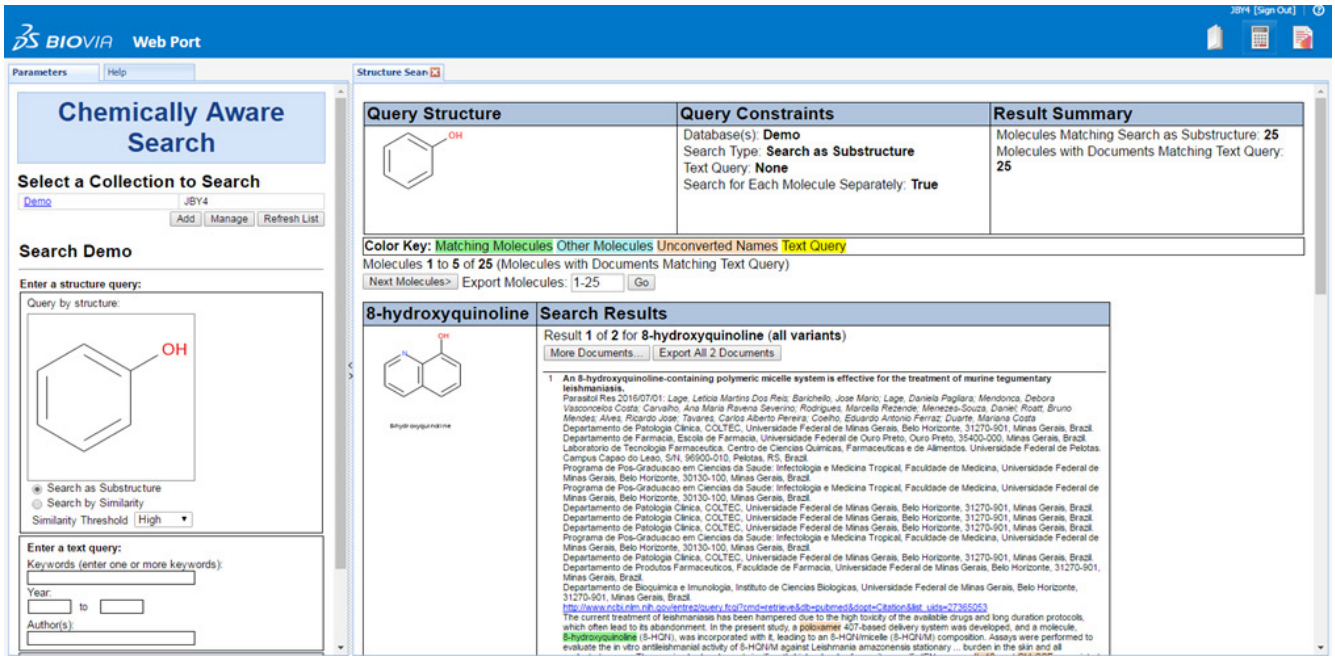
Text Analytics

- Analyze documents / text to extract key concepts to find correlations in documents and online literature for competitive intelligence and to support primary research findings
- Automate information extraction and summarization
- Use relationships between documents for greater understanding

Chemical Text Mining

- For biology, identify Amino Acids, Proteins, DNA, RNA, Cell Lines, and Cell Types.
- For chemistry, identify chemical names and embedded structures in text and convert these to structures with the included OpenEye name-to-structure converter.
  - Names: systematic names (IUPAC, InChI, and SMILES), formulae, family names, abbreviations, identifiers, CAS numbers, and non-systematic (trivial) names
  - Embedded Structures: BIOVIA Draw, ISIS/Draw, ChemDraw and Accord molecules
- Create chemically-aware search applications by mapping structures to documents in a structure-searchable text database.





SEARCH: LOCAL, ONLINE & ENTERPRISE DOCUMENTS

The Documents and Text Collection gives you the power to extract knowledge from important online document resources such as PubMed, US and European Patents, Bing, TOXNET, Twitter and Wikipedia (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. You can also search and mine internal documents in exactly the same way. The Documents and Text Collection indexes and searches folders that contain PDF, Microsoft Office, HTML, or text files (extendable to other file formats). You can even store the results of online searches in local repositories for speedy retrieval and post processing. Local databases of documents stay current automatically by monitoring the folder contents for the introduction of any new or edited documents.

CHEMICALLY-AWARE SEARCH

The Documents and Text Collection uses a chemically aware search algorithm to identify and convert chemical names to structures and enable researchers to query for information in a more intuitive and effective way with substructure, similarity, SMILES, and IUPAC name searches. Also included is a complete end-user application, "Chemically Aware Search", for indexing and searching collections of documents containing structures. It allows you to identify a set of interesting documents, download the content to a local database and index the documents for text or structure search, all without the need to author protocols or additional configuration.

ANNOTATE SCIENTIFIC RESULTS

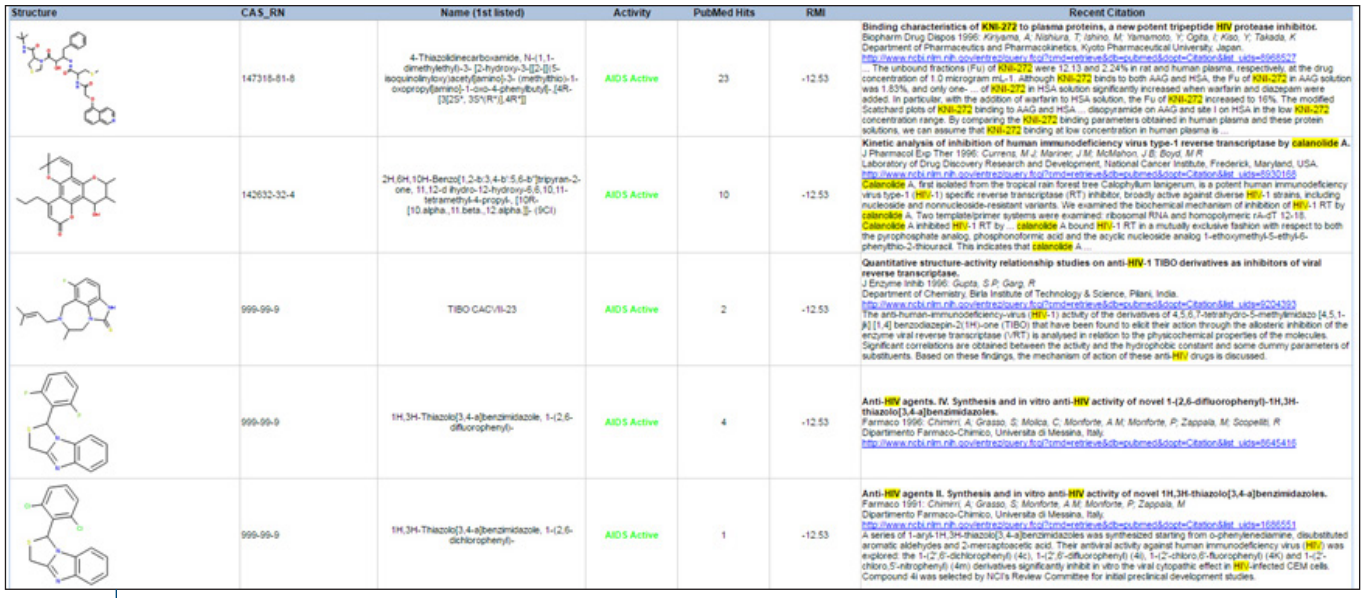
When reporting the results of pipelined data analyses, it is often useful to include additional information about the output data points. With the Documents and Text Collection, you can easily add a few steps at the end of any Pipeline Pilot protocol and have each data point serve as a query to search a database of literature. For example, after clustering a set of genes with the Biology Collections, 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. Also, you can create your own custom document templates in Microsoft Word, and use the Documents and Text Collection to surgically fill in scientific data and charts to generate beautifully styled final documents.

IDENTIFY EMERGING TRENDS

The Documents and Text 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 Documents and Text 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.



Structure	CAS_RN	Name (1st listed)	Activity	PubMed Hits	RMI	Recent Citation
	147316-61-8	4-Thiazolidinecarboxamide, N-(1,1-dimethylethyl)-3- [2-hydroxy-3-[[2-[[(5-isoquinolinyloxy)acetyl]amino]-3- (methylthio)-1-oxopropyl]amino]-1-oxo-4-phenylbutyl]-, [4R-[3[2S*, 3S*(R*)],4R*]]	AIDS Active	23	-12.53	**Binding characteristics of KNI-272 to plasma proteins, a new potent tripeptide HIV protease inhibitor.** Biopharm Drug Dispos 1996; *Kiriyama, A; Nishiura, T; Ishino, M; Yamamoto, Y; Ogita, I; Kiso, Y; Takada, K* Department of Pharmaceutics and Pharmacokinetics, Kyoto Pharmaceutical University, Japan. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=Citation&list_uids=8668527 ... The unbound fractions (Fu) of KNI-272 were 12.13 and 2.24% in rat and human plasma, respectively, at the drug concentration of 1.0 microgram mL-1. Although KNI-272 binds to both AAG and HSA, the Fu of KNI-272 in AAG solution was 1.83%, and only one- ... of KNI-272 in HSA solution significantly increased when warfarin and diazepam were added. In particular, with the addition of warfarin to HSA solution, the Fu of KNI-272 increased to 16%. The modified Scatchard plots of KNI-272 binding to AAG and HSA ... disopyramide on AAG and site I on HSA in the low KNI-272 concentration range. By comparing the KNI-272 binding parameters obtained in human plasma and these protein solutions, we can assume that KNI-272 binding at low concentration in human plasma is ...
	142632-32-4	2H,6H,10H-Benzo[1,2-b:3,4-b':5,6-b'']tripyran-2-one, 11,12-d ihydro-12-hydroxy-6,6,10,11-tetramethyl-4-propyl-, [10R-[10.alpha.,11.beta.,12.alpha.]]- (9CI)	AIDS Active	10	-12.53	**Kinetic analysis of inhibition of human immunodeficiency virus type-1 reverse transcriptase by calanolide A.** J Pharmacol Exp Ther 1996; *Currens, M J; Mariner, J M; McMahon, J B; Boyd, M R* Laboratory of Drug Discovery Research and Development, National Cancer Institute, Frederick, Maryland, USA. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=Citation&list_uids=8530166 Calanolide A, first isolated from the tropical rain forest tree Calophyllum lanigerum, is a potent human immunodeficiency virus type-1 (HIV-1) specific reverse transcriptase (RT) inhibitor, broadly active against diverse HIV-1 strains, including nucleoside and nonnucleoside-resistant variants. We examined the biochemical mechanism of inhibition of HIV-1 RT by calanolide A. Two template/primer systems were examined: ribosomal RNA and homopolymeric rA-dT 12-18. Calanolide A inhibited HIV-1 RT by ... calanolide A bound HIV-1 RT in a mutually exclusive fashion with respect to both the pyrophosphate analog, phosphonoformic acid and the acyclic nucleoside analog 1-ethoxymethyl-5-ethyl-6-phenylthio-2-thiouracil. This indicates that calanolide A ...
	999-99-9	TIBO CAC/VII-23	AIDS Active	2	-12.53	**Quantitative structure-activity relationship studies on anti-HIV-1 TIBO derivatives as inhibitors of viral reverse transcriptase.** *J Enzyme Inhib 1996; Gupta, S P; Garg, R* Department of Chemistry, Birla Institute of Technology & Science, Pilani, India. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=Citation&list_uids=9004393 The anti-human-immunodeficiency-virus (HIV-1) activity of the derivatives of 4,5,6,7-tetrahydro-5-methylimidazo [4,5,1-jk] [1,4] benzodiazepin-2(1H)-one (TIBO) that have been found to elicit their action through the allosteric inhibition of the enzyme viral reverse transcriptase (VRT) is analysed in relation to the physicochemical properties of the molecules. Significant correlations are obtained between the activity and the hydrophobic constant and some dummy parameters of substituents. Based on these findings, the mechanism of action of these anti-HIV drugs is discussed.
	999-99-9	1H,3H-Thiazolo[3,4-a]benzimidazole, 1-(2,6-difluorophenyl)-	AIDS Active	4	-12.53	**Anti-HIV agents. IV. Synthesis and in vitro anti-HIV activity of novel 1-(2,6-difluorophenyl)-1H,3H-thiazolo[3,4-a]benzimidazoles.** Farmaco 1996; *Chimirri, A; Grasso, S; Monica, C; Monforte, A M; Monforte, P; Zappala, M; Scopelliti, R* Dipartimento Farmaco-Chimico, Universita di Messina, Italy. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=Citation&list_uids=8645416
	999-99-9	1H,3H-Thiazolo[3,4-a]benzimidazole, 1-(2,6-dichlorophenyl)-	AIDS Active	1	-12.53	**Anti-HIV agents II. Synthesis and in vitro anti-HIV activity of novel 1H,3H-thiazolo[3,4-a]benzimidazoles.** Farmaco 1991; *Chimirri, A; Grasso, S; Monforte, A M; Monforte, P; Zappala, M* Dipartimento Farmaco-Chimico, Universita di Messina, Italy. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=retrieve&db=pubmed&dopt=Citation&list_uids=1686551 A series of 1-aryl-1H,3H-thiazolo[3,4-a]benzimidazoles was synthesized starting from o-phenylenediamine, disubstituted aromatic aldehydes and 2-mercaptoacetic acid. Their antiviral activity against human immunodeficiency virus (HIV) was explored: the 1-(2',6'-dichlorophenyl) (4c), 1-(2',6'-difluorophenyl) (4i), 1-(2'-chloro,6'-fluorophenyl) (4K) and 1-(2'-chloro,5'-nitrophenyl) (4m) derivatives significantly inhibit in vitro the viral cytopathic effect in HIV-infected CEM cells. Compound 4i was selected by NCI's Review Committee for initial preclinical development studies.

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.

ABOUT PIPELINE PILOT

BIOVIA Pipeline Pilot is a graphical programming environment that optimizes scientific innovation, increases operational efficiency and reduces costs for both research and IT. Pipeline Pilot automates the scientific analysis of data, enabling users across the enterprise to rapidly explore, visualize and report research results. Pipeline Pilot's Component Collections contain the "scientific building blocks" that allow researchers, developers, engineers and IT professionals to perform both science-specific and generic data-processing functions. By graphically combining components, you can construct workflows for data retrieval, filtering, analysis and reporting.

To learn more about Pipeline Pilot,
go to 3DSBIOVIA.com/pipeline-pilot

Our **3D**EXPERIENCE® platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE® Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 190,000 customers of all sizes in all industries in more than 140 countries. For more information, visit **www.3ds.com**.

3DEXPERIENCE®

DASSAULT SYSTEMES | The **3D**EXPERIENCE® Company

Dassault Systèmes Corporate
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

BIOVIA Corporate Americas
BIOVIA
5005 Wateridge Vista Drive,
San Diego, CA
92121
USA

BIOVIA Corporate Europe
BIOVIA
334 Cambridge Science Park,
Cambridge CB4 0WN
England