

PubChem

Site search for over 100 million chemical compounds from 741 different data sources



Company

"PubChem" group in the National Center for Biotechnology Information (NCBI) is part of the National Library of Medicine (NLM), a branch of the United States National Institutes of Health (NIH). PubChem is the largest free chemical informatics web site in the world and has detailed information from 741 Data Sources for over 103M chemical compounds, 254M Substances, 269M Bioactivities, 31M Literature, 3M Patents, 1M BioAssays and more.

Siqian He,
US National Institutes of
Health



"We cannot achieve such success without the Sphinx/Manticore search engine!"

Thank you all for making such a powerful search engine!"



Challenge

Manticore used to do full-text search among all collections (chemical compound, chemical substance, biological assay, patent, pubmed, protein, gene, taxonomy, disease, literature, pathway, pathway reaction, ...), total 10TB data in size. They first tried Solr, but it did NOT scale with their data growth. Sphinx/Manticore turned out to be the light-weight but powerful search engine that exactly suits their needs.



Contact or visit us if you have any questions.

E-mail: contact@manticoresearch.com

Services page: <https://manticoresearch.com/services/>



Solutions and results

- **Calling c++ sphinx client lib:** 'api/libsphinxclient/sphinxclient.h' to do a search related query
- **Autocomplete-** is a feature when an application predicts the rest of a not typed word. So the user can type only 2 or 3 first letters of a compound and be able to see different suggestions
- **Faceting** - gives the ability to make filtration of search results by different properties of searched items



How it looks

SEARCH FOR

citra

Compound	Gene	-Autocomplete	Taxonomy
Citraconic anhydride	citrate synthase		Stomoxys calcitrans
CITRAZINIC ACID	citrate synthase like		Diaporthe citri
Citraconic acid	citrate lyase beta like		Panonychus citri
Citramalic acid	citrate (5i)-synthase CIT1		Diaphorina citri
Citral	citrate (5i)-synthase CIT2		Alternaria citri
CITRAL DIETHYL ACETAL	ATP citrate lyase		Xanthomonas citri
Citral dimethyl acetal	isocitrate lyase		Spiroplasma citri
citrate	Isocitrate dehydrogenase		Citrus
citranaxanthin	homoisocitrate dehydrogenase		Naja atra
Citraconimide	CIT1		Indian cobra

Summary Similar Structures Search Related Records PubMed (MeSH Keyword)

Faceting

Compounds	Substances	Proteins	Pathways	BioAssays	Literature	Patents
(123)	(550)	(1)	(24)	(1,778)	(66,311)	(531)

Searching descriptions and metadata of bioassay records submitted by PubChem's contributors. [Read More...](#)

1,778 results Filters SORT BY Relevance Download



Contact or visit us if you have any questions.

E-mail: contact@manticoresearch.com

Services page: <https://manticoresearch.com/services/>