Linguamatics

datasheet

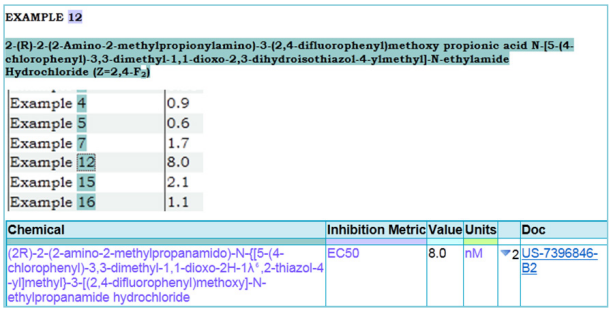
I2E Chemistry

Chemical Text Mining

I2E Chemistry provides the first interactive text mining system designed for chemistry. By integrating chemical search from **ChemAxon** and text mining from **Linguamatics**, users can identify chemical structures and understand their role within documents such as patents, scientific articles, and internal documents.

I2E Chemistry allows users to extract all chemical structures within documents or to filter based on substructure or similarity. It allows extraction of chemicals, associated properties and any relationship with other entities. I2E's agile NLP-based text mining approach allows users to discover a wide range of chemical information, such as:

- Structure activity relationships (SAR)
- Melting point or boiling point of a chemical
- The role of a chemical in a reaction
- Concentration of chemicals



EXAMPLE 12

2-(R)-2-(2-Amino-2-methylpropionylamino)-3-(2,4-difluorophenyl)methoxy propionic acid N-[5-(4-chlorophenyl)-3,3-dimethyl-1,1-dioxo-2,3-dihydroisothiazol-4-ylmethyl]-N-ethylamide Hydrochloride (Z=2,4-F_2)

Example 4	0.9
Example 5	0.6
Example 7	1.7
Example 12	8.0
Example 15	2.1
Example 16	1.1

Chemical	Inhibition Metric	Value	Units		Doc
(2R)-2-(2-amino-2-methylpropanamido)-N-{[5-(4-chlorophenyl)-3,3-dimethyl-1,1-dioxo-2H-1λ⁶,2-thiazol-4-yl]methyl}-3-[(2,4-difluorophenyl)methoxy]-N-ethylpropanamide hydrochloride	EC50	8.0	nM	2	US-7396846-B2

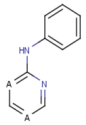
Figure 1 shows extracting SAR information from tables

Chemical	Feature	Value	Doc	Hit
4-hydroxy-5-phenyl-4,5-dihydro-2H-spiro[1λ⁴-benzothiepine-3,1'-cyclohexane]-1,1-dione	mp	209-210°	US-6943189-B2	1 Example 31 cis-4-Hydroxy-5-phenyl-2,3,4,5-tetrahydro spiro(benzothiepine-3,1'-cyclohexane)-1,1-dioxide (60) ... mg of white crystal, mp 209-210° C. Proton and carbon NMR ...
2'-phenyl-3'-oxa-7'λ⁴-thiaspiro[cyclohexane-1,5'-tricyclo[6.4.0.0²,⁴]dodecane]-1'(8'),9',11'-triene-7',7'-dione	mp	154-155°	US-6642268-B2	1 8b-Phenyl-1a,2,3,8b-tetrahydrospiro (benzothiepino[4,5-b]oxirene-2,1'-cyclohexane)-4,4-dioxide (58) ...) of yellow solid, mp 154-155° C. Proton and carbon NMR ...
4-hydroxy-5-phenyl-4,5-dihydro-2H-spiro[1λ⁴-benzothiepine-3,1'-cyclohexane]-1,1-dione	mp	99-100°	US-6642268-B2	1 trans-4-Hydroxy-5-phenyl-2,3,4,5-tetrahydro spiro(benzothiepine-3,1'-cyclohexane)-1,1-dioxide (59) ... as a white solid, mp 99-100° C. Proton NMR showed this ...

Figure 2 shows extracting melting point information

I2E Chemistry provides a wide range of chemical search capabilities to help find both known and novel compounds:

- Exact, substructure and similarity search
- Dictionary based search for known chemicals e.g. common names
- Name to structure to find novel chemicals

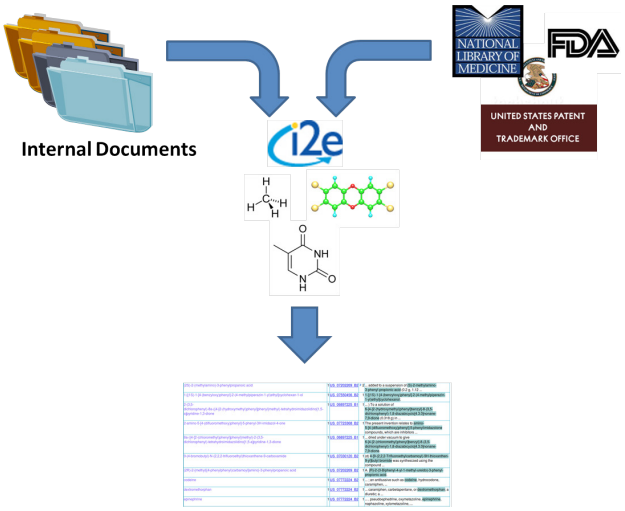


Class1	Relation	Class2		Doc		Hit
imatinib	inhibit	ABL1	2	16303243	1	Imatinib, an inhibitor of BCR-ABL tyrosine kinase, also inhibits BCRP-mediated drug transport.
imatinib methanesulfonate	inhibit	ABL1	3	15803362	1	BACKGROUND: Imatinib mesylate is a potent inhibitor of Abl, KIT, and PDGFR tyrosine kinases.
gefitinib	inhibit	EGFR	2	15692759	2	The clinical benefit and safety of the EGFR tyrosine kinase inhibitor gefitinib ('Iressa')1 was evaluated in this Phase II, multicentre study of patients with taxane and anthracycline pretreated, metastatic breast cancer.
lapatinib	inhibit	ERBB2	3	16452223	2	Alternatively, inhibition of ErbB2 signaling using lapatinib (GW572016), a reversible small-molecule inhibitor of ErbB1/ErbB2 tyrosine kinases, at pharmacologically relevant concentrations, leads to marked inhibition of survivin protein with subsequent apoptosis.

Figure 3 shows what chemicals with this substructure act as inhibitors

Identify and extract chemicals from internal and external documents

I2E Chemistry can process documents in a wide range of formats such as PDF, XML, docx, tsv, allowing internal document silos to become available for chemical text mining. Taking an automated approach to identify chemical structures avoids costly and time consuming manual chemical mark-up. Searches can also be run across both internal and external documents (see below illustration).



Key Benefits

- Faster and more efficient identification of chemicals and associated information
- Fully automated extraction of chemical information
- Avoid costly manual chemical mark-up
- Chemical searching across any document (internal and external)
- Scalable to millions of large documents e.g. full text

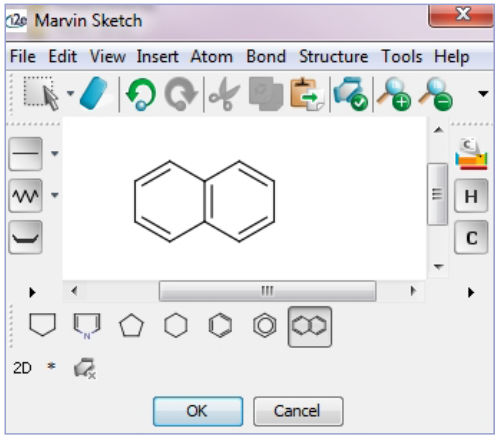
Linguamatics

Save and share searches for automation and reproducibility

Chemical substructure searches can be saved and re-run on any set of documents. The portable nature of the chemical queries allows them to be shared with people you are collaborating with.

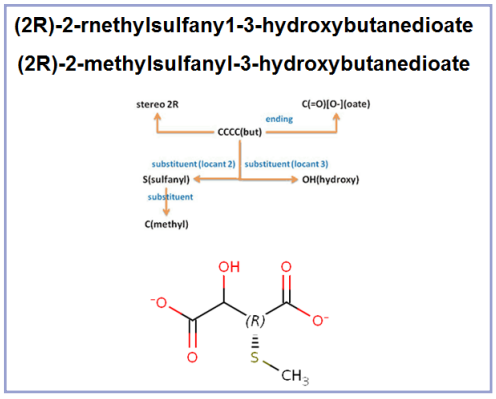
Integrated Chemical Drawing

ChemAxon's Marvin Sketch is included in the product and can be used to specify chemical structures to search.



Chemically aware OCR correction

Errors introduced by scanning documents are corrected to ensure they are valid chemical structures.



Applications include:

Scientific research, intellectual property and commercial intelligence.

Why is it different from searching using chemical structures?

Text mining can answer specific questions much more efficiently than document search e.g. show me all chemicals used for capsule coatings combined with gelatine returns a list of chemicals with associated evidence, rather than a list of documents to read.

What are the advantages of fully automatic methods for finding chemical structures?

This allows structures to be found in documents where mark-up by hand has not yet been done, or is uneconomic to do e.g. for a company's internal reports. It is highly scalable and can find chemical structures at particular points within a document so you can ask questions such as "which chemicals are mentioned in the same list as chemical X".

About ChemAxon

ChemAxon provides cheminformatics software platforms and desktop applications to optimize the value of chemistry information in pharmaceutical R&D.

Their mission is to enable scientists to manage their chemical data via a modern and cost effective suite of informatics tools and applications, developed together with their customers and partners.

Over 500 corporate clients, including 9/10 top pharmaceutical companies. 3/5 Top Pharma have migrated to ChemAxon cheminformatics.

For more information or a demonstration, please call us on
+44 1223 651910
enquiries@linguamatics.com

THE QUEEN'S AWARDS FOR ENTERPRISE: 2014

Linguamatics 324 Cambridge Science Park
Milton Road Cambridge CB4 0WG UK
Tel: +44 (0)1223 651 910

Linguamatics 1900 West Park Drive Suite 280
Westborough MA 01581 USA
Tel: +1 617 674 3256

www.linguamatics.com