

# CHEMICAL NAMING TOOLS

Daniel Bonniot  
Budapest Annual Meeting  
2016

# *The naming family*

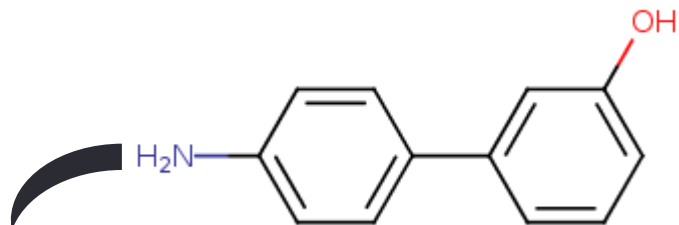
- Convert structures to IUPAC and common names (Structure to Name)
- Convert names to structures (Name to Structure)
- Extract structures from documents (Document to Structure)
- Display documents in web browser (Document annotator)
- Index and search document collections (Plexus Mining)

## Agenda

1. Recent improvements (focus on s2n)
2. Architecture for custom document processing
3. Integration story by Questel

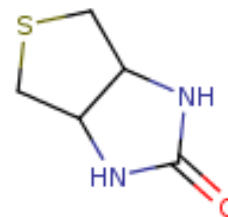
# Structure to Name

## Ring assemblies



4'-amino-[1,1'-biphenyl]-3-ol

## Saturated rings in fusion



hexahydro-1H-thieno[3,4-d]imidolidin-2-one

hexahydro-1H-thieno[3,4-d]imidazol-2-one

General for all types of rings  
(*binaphthalene*, *bicyclohexane*,  
*bifuran*, ...)

### Astex Therapeutics:

“It has reduced the number of roundtrip failures on our corporate database **by about half!**”

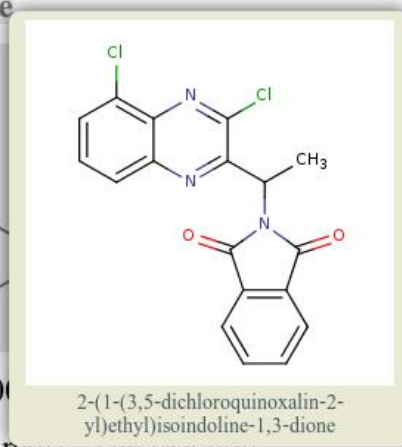
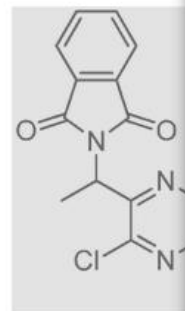
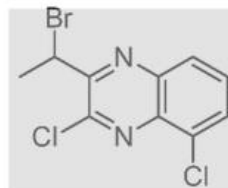
# Document Annotator

WO 2013/152150

PCT/US2013/035203

- 39 -

2-(1-(3,5-dichloroquinoxalin-2-yl)ethyl)isoindoline-1,3-dione



To a solution of 2-(1-bromoethyl)-3,5-dichloroquinoxaline (1.00 mol) in DMF (8.2 L) was added potassium phthalimide (1.21 Kg, 6.54 mol). The reaction

5 mixture was stirred for 3 h. At this time LC-MS analysis showed that the reaction

# *Document Annotator*

Input: HTML, XML, Text or PDF document

Output: HTML, with chemical structures highlighted

Usage: Java and .NET API for integration in custom applications  
(also used in ChemCurator)

# *Document to Structure*

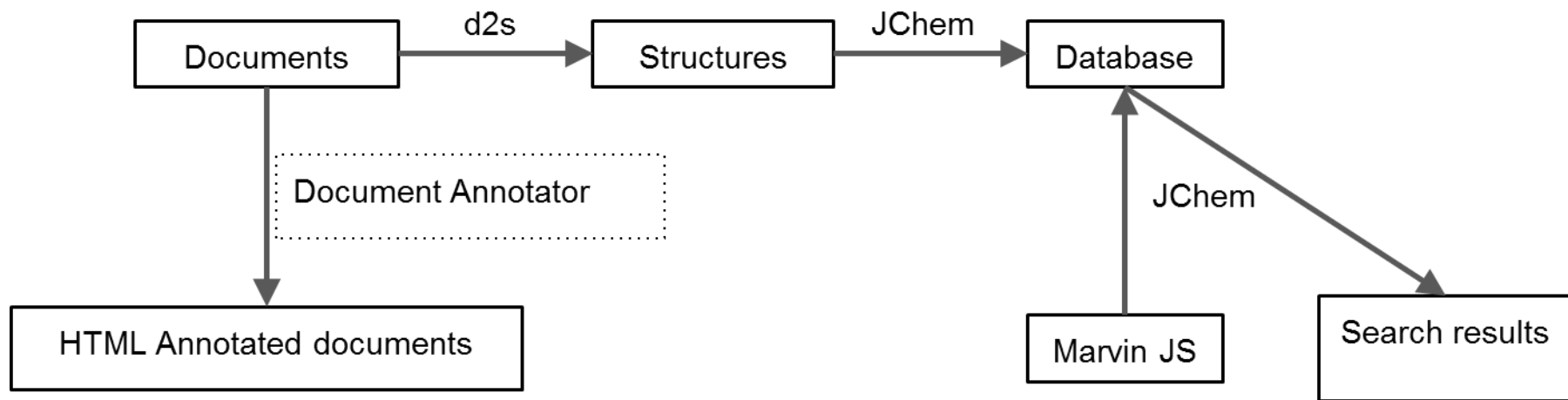
Detect and extract chemical structures from a document

API to process a document, a single sentence or short paragraph

Tens of options to control behaviour:

- include chemical elements?
- include fragments?
- ...

# Chemical indexing and searching: architecture



# *New Plexus Mining demo site*

<http://ps-mining-demo.chemaxon.com>

<http://ps-demo.chemaxon.com>

- Three document sets:
  - XML Patents (US and Asian)
  - PDF Journal articles
  - Safety Data Sheets (MSDS)
- Try it at the demo station!
  - Possibility for demos and evaluations on customer sets

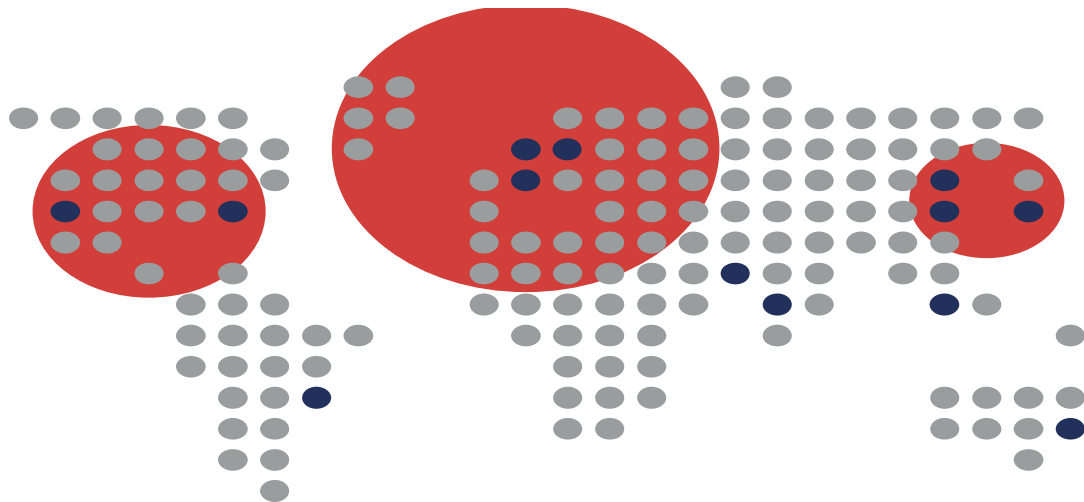




---

# Covering Intellectual Property Needs Across The Innovation Lifecycle

# 30M€



1970  
France Telecom

2001  
Spin-Off

2007  
1<sup>st</sup> LBO

2015  
2<sup>nd</sup> LBO





## HOW

---

## WHAT

---

**Find  
Technology**

**Mitigate  
Risks**

**Sustain  
Advantage**

## WHEN

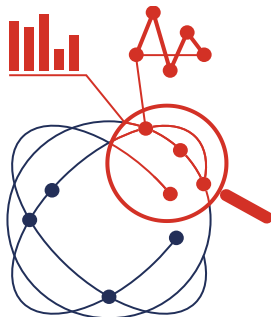
---

Research &  
development

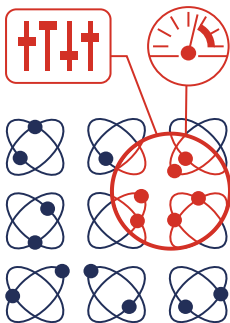
Product  
development

Market  
development

Turn insight  
into action



Intellectual Property  
Business Intelligence



Intellectual Asset  
Management

Turn insight  
into action



Merged Markush Service (MMS)

Sources: WPIM & MPharm

Producers: Derwent & French PTO

Based on the DARC system

Turn insight  
into action



Performant & flexible solutions

Complementarity (data, solutions, people)

Expertise in chemistry, patents & analysis

Reputation



Turn insight  
into action



Chemical Structure Display

Chemical Indexing

Structure Searching

# Questel-ChemAxon Partnership

## Chemical Structure Display

The screenshot displays the Intellixir software interface for document management. The main content area shows a patent document titled "Bauxite direct flotation composite collector and application method thereof". The document includes the following details:

- Date:** 2014
- Patent Number:** CN104148191 A
- Priority number:** CN20140380611
- Representative patent:** CN104148191 A
- Authors:** Cao Xuefeng, Hu Yuehua, Sun Wei, Gao Jiande, Huang Hongjun, Zhu Yiyang
- Affiliations:** CENTRAL SOUTH UNIVERSITY
- Short Affiliations:** CENTRAL SOUTH UNIV
- Descriptors:** Flotation, Flotation agent, Macromolecular, Collector
- IPC:** B03D1/016, B03D101/02

The **Abstract** describes a collector for bauxite direct flotation desliming, comprising a chemical component, an organosilicon surfactant, a polyoxyethylene ether surfactant, and a mining foam of monocarboxylic fatty acid, organosilicon surfactant, and polyoxyethylene ether surfactant. The collector has a collecting ability, dispersing performance, and aluminum oxide recovery that are improved, and the use of desliming is improved.

The **Claims (0)** section includes a claim for a kind of bauxite flotation process collector, characterized by a single cycloalkane that is synthesized from an acid and an organosilicon surface active agent and polyethyleneoxide ether class agent complex thing.

The **Chemical Structure Display** shows the structure of naphthene-1-carboxylic acid, which is a naphthalene ring system with a carboxylic acid group (-COOH) attached to the first carbon atom. The structure is highlighted with a blue box and labeled "naphthene acid".

The interface also features a navigation bar at the top with icons for Home, Browse, Patents, Text Detection, Categorize, Players, Focus, Experts, My Docs, and My Interfaces. A search bar is located in the top right corner. The document is identified as "document: 1/235".



Available now



# Questel-ChemAxon Partnership

## Chemical Indexing & Structure Searching

### Features

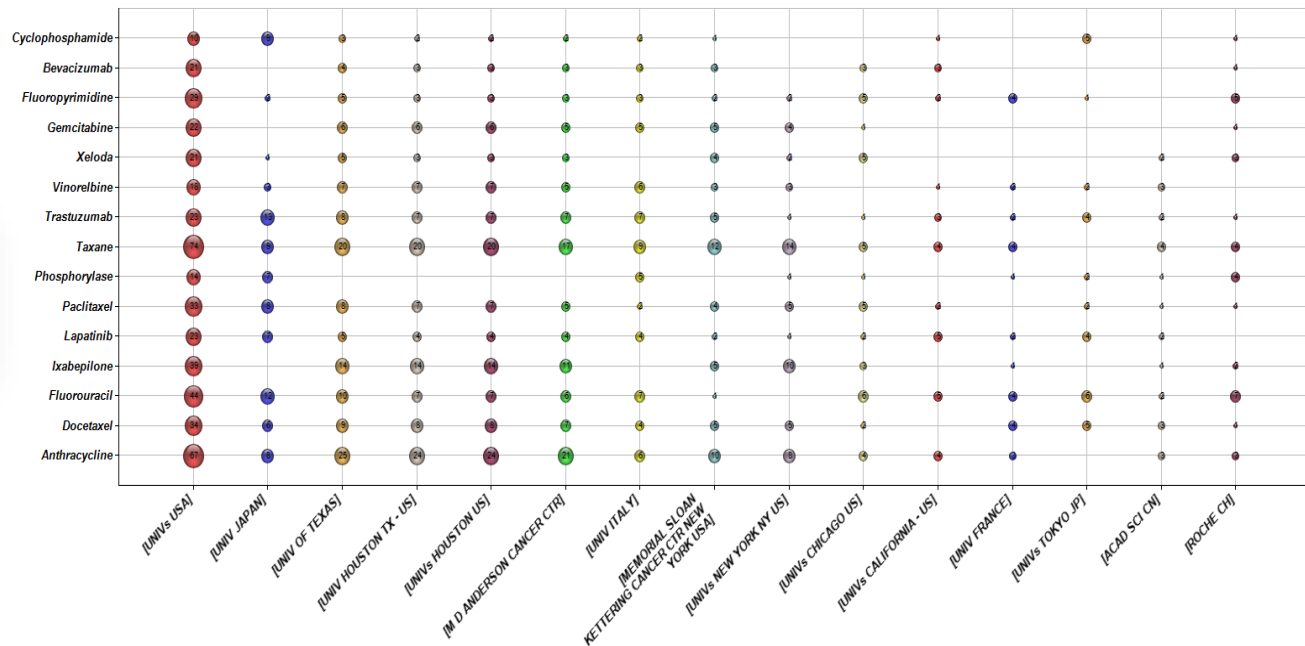
- Chemical Structures detected in patents, scientific pub...
- Records categorization based on Chemical Structures & Sub-Structures

### Key benefits

- Process all sources including internal documents
- Produce landscapes based on Chemical Structures & Sub-Structures

# Questel-ChemAxon Partnership

## Chemical Indexing & Structure Searching



# Questel-ChemAxon Partnership

## Chemical Indexing & Structure Searching

### Features

- Chemical Structures detected in Patents from 100+ PTO
- Normalized chemical names included in patents & used in similarity search

### Key benefits

- Detect chemical products among 90 million patents
- Search efficiently thanks to chemical name normalization
- Fast track updating