

Name to structure, Structure to name, chemicalize.org

Daniel Bonniot de Ruisselet



ChemAxon
Solutions for Cheminformatics

Why use chemical names?

- Easier than drawing
- Familiar
- Used in patents, articles, ...

We will talk about:

- Name generation (structure to name, s2n)
- Name import (name to structure, n2s)
- Name extraction (document to structure, d2s)
- Name correction (OCR-error fixing)
- Name highlighting (chemicalize.org)

Structure to name

Usage:

- Plugin in MarvinView and MarvinSketch
- Label updated in real-time in MarvinSketch
- Batch: Save As: IUPAC Name in MarvinView
- Batch: command line (molconvert, cxcalc)
- Instant JChem

Options:

- Strict IUPAC or traditional
- Timeout

Structure to name

Already stable (focus on n2s)

Comparison between 5.1.0 and 5.2.5 on NCI database (260K structures)

- Both over 99.9% named, 4% changed
- More fused names supported (60% to 66%):
 - 5-methyl-6-azatricyclo[8.4.0.0^{2,7}]tetradeca-1(10),2(7),3,5,8,11,13-heptaene
 - 3-methylbenzo[f]quinoline
- Better support for ions (e.g. -olate)
- Stricter IUPAC numbering and priorities
- Overspecific E/Z labels removed

Name to structure

Usage

- Edit/Import Name... in MarvinSketch
- Automatic format recognition
- Paste a name from the clipboard
- Open IUPAC Name file in MarvinView or MarvinSketch (.name extension)
- Batch from command line (molconvert)

Name to structure: evaluation

Molecule->Name->Molecule on NCI

- 5.1.0: 90.0% names imported, 68.7% identical
- 5.2.2: 97.6% names imported, 94.1% identical
- 5.2.5: 97.8% names imported, 95.9% identical

Pubchem data

- 5.1.0: 88.8% names imported, 94.0% identical
- 5.2.2: 98.3% names imported, 95.6% identical
- 5.2.5: 98.3% names imported, 96.1% identical

Name to structure: +33% speed

Customization (new)

Extend name-to-structure conversion using your in-house data

- Dictionary file
- Simple API:
 - Database lookup
 - Webservice lookup
 - ...
- Fully flexible

Document to structure (new)

Goal

- Process documents containing text
- Recognize chemical names
- Convert them to structures
- Return locations, names and structures

Formats

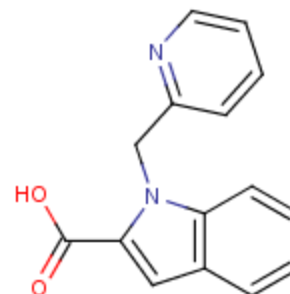
- Implemented: text, html, xml
- Planned: PDF, Doc, ...

OCR error fixing (new)

Scanned texts contain numerous recognition errors

Examples:

- L (small L) instead of 1
- I instead of I (II, iL)



1-(pyridin-2-ylmethyl)-1H-indole-2-carboxylic acid

Uses:

- By default in d2s
- Option in n2s (upcoming)

- Adds structural information to existing public webpages
- Popup window with structure image
- Link to structure predictions (logP, pKa, ...)
- Searchable structure->webpage index
- Could be installed natively on custom website, with custom features

Recap

- Name-to-structure and structure-to-name available and improving fast
- Document-to-structure just released
- Extend using your in-house dictionaries and databases
- Try it, send feedback, we're listening!

Find out more

- **Product descriptions & links**

www.chemaxon.com/products.html

- **Forum**

www.chemaxon.com/forum

- **Presentations and posters**

www.chemaxon.com/conf

- **Download**

www.chemaxon.com/download.html

The collage consists of four screenshots from the ChemAxon website:

- Top-left:** The main website page with navigation tabs (Home, Products, Documentation, Download, Forum, About us, Contact us) and a 'Recent News' section listing updates from 2007.
- Top-right:** The 'Upcoming Meetings' section, listing events such as 'Seminars series: Bridging the corporate and desktop' and 'Software Pipeline Pilot User Group Meeting'.
- Bottom-left:** The 'Marvin and Calculator Plugin Demo' page, featuring a rich editing list with features like 'wide range of file types supported', 'copy and paste between different editors', and '3D editing'.
- Bottom-right:** The 'Technical Support Forum' page, displaying a table of support topics and posts.

| Support | Topics | Posts | Last Post |
|--|--------|-------|-------------------------------------|
| Structure editing, viewing and file formats | 827 | 4081 | Feb 01, 2008 1:00 pm Closed |
| Support for MarvinSketch, MarvinView and MarvinCenter | | | |
| MarvinSpace | 60 | 356 | The Jan 31, 2008 5:16 pm Closed |
| Development discussion area for OpenGL 3D rendering main/mml molecule viewer | | | |
| Structure based prediction and Chemical Terms | 351 | 174 | Feb 01, 2008 10:02 am Closed |
| Support for Calculator Plugins operation through cracks, API, Marvin, Instant XChem and Chemical Terms | | | |
| Structure search and chemical database | 426 | 1982 | Feb 01, 2008 10:43 am Closed |
| Support for XChem Base and XChem Cartridge | | | |
| Instant XChem | | | |
| Discussion area for Instant XChem (Structure database GUI, both Calculator Plugin processor etc) | 173 | 416 | Wed Jan 30, 2008 9:10 pm Closed |
| Structure conversion/alignment / standardization | 76 | 245 | Wed Jan 30, 2008 9:30 pm Closed |
| Support for Standardizer | | | |
| Library enumeration, virtual synthesis and metabolite generation | 116 | 517 | Feb 28, 2008 3:02 am Closed |
| Support for Reactor and Fragmenter | | | |
| Virtual screening, clustering and molecular descriptors | 308 | 471 | The Jan 29, 2008 1:00 am Closed |
| Support for Screen and Shaker | | | |
| License Issues | 4 | 7 | The Jan 29, 2008 10:32 am Closed |
| Support for the technical questions related to license handling | | | |