Name to structure, Structure to name, chemicalize.org

Daniel Bonniot de Ruisselet

ChemAxon

Solutions for Cheminformatics

Motivations

Why use chemical names?

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

09 15-16S
M
Meetingium
32°47.367' -117°15.159'

Overview

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

09 15-16S
M
Meetingium
32°47.367' -117°15.159'

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

09 15-16S
M
Meetingium
32°47.367' -117°15.159'

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

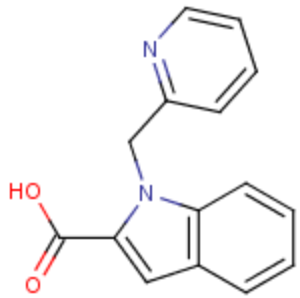
- Implemented: text, html, xml
- Planned: PDF, Doc, ...

OCR error fixing (new)

Scanned texts contain numerous recognition errors

Examples:

- L (small L) instead of 1
- I instead of l (Il, iL)



l-(pyridin-2-ylmethy l)- 1H- indole-2-carboxylic acid

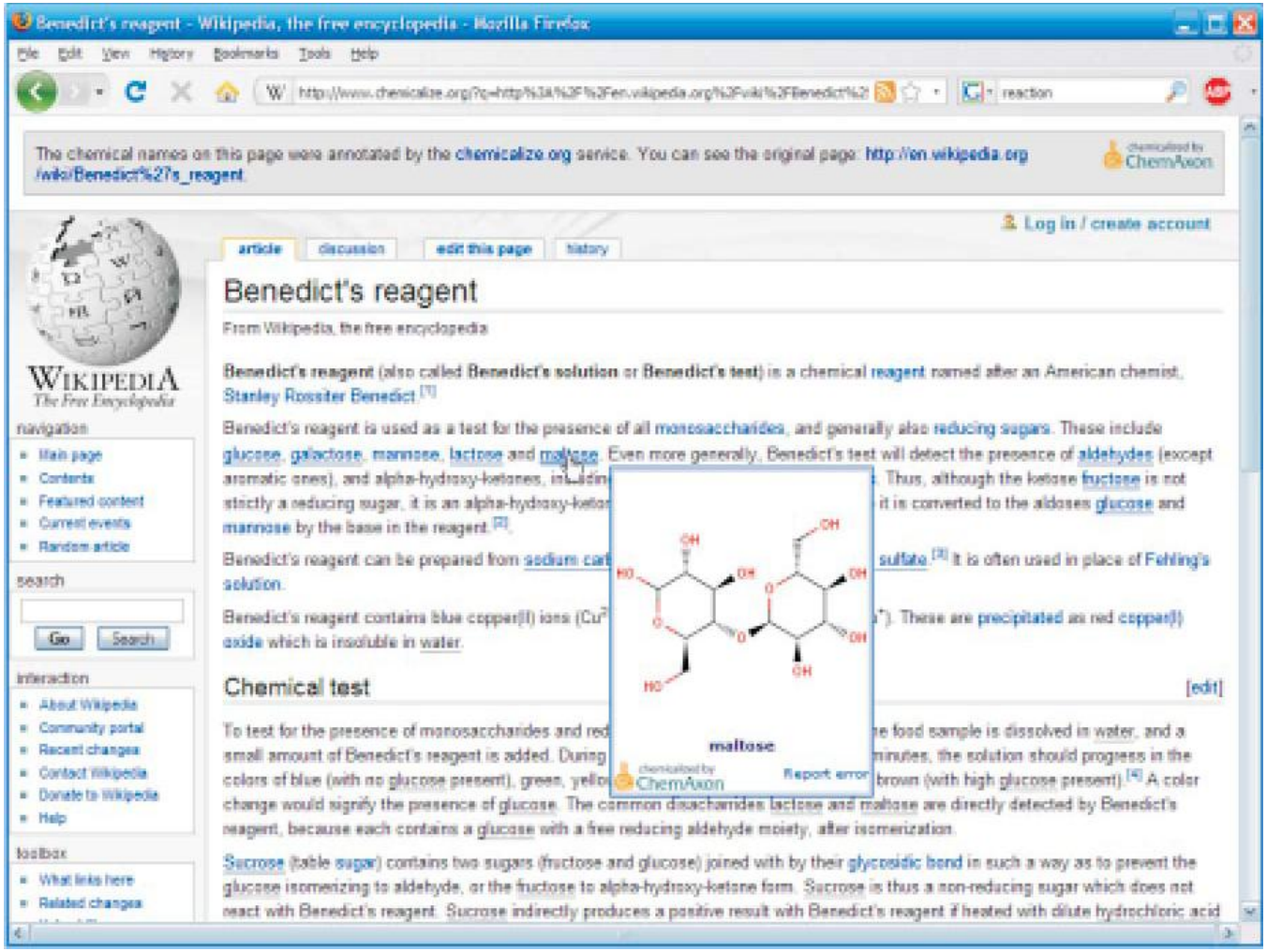
Uses:

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

Chemicalize.org

- 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

Chemicalize.org

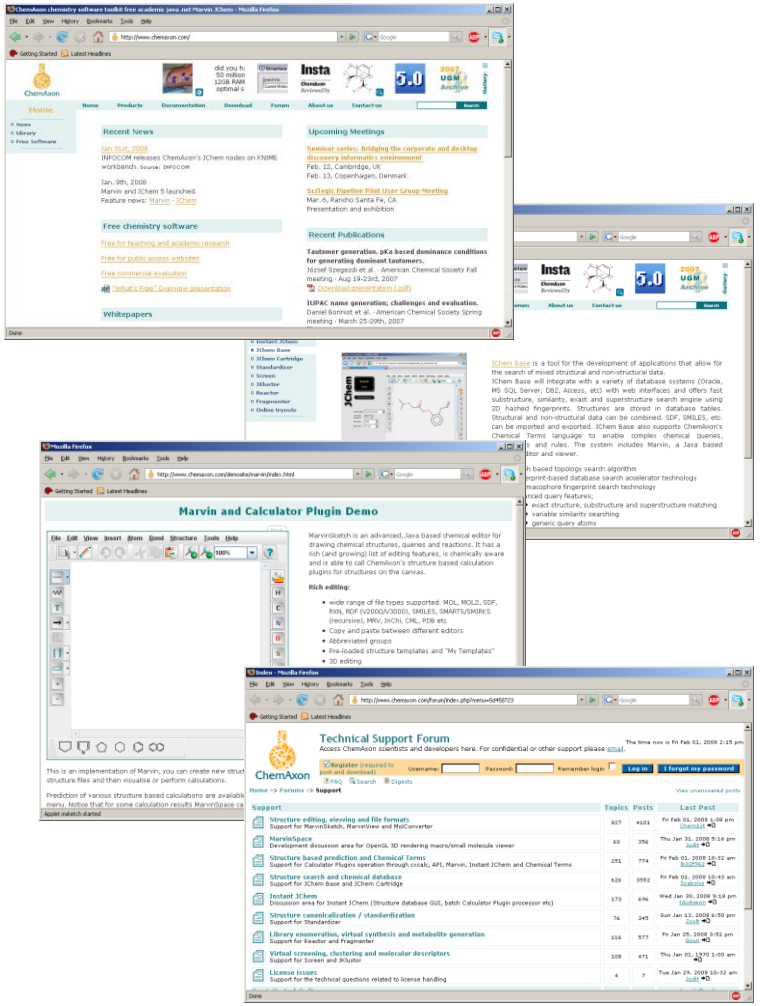


Recap

- 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



09 15-16S
M
Meetingium
32°47.367' -117°15.159'

ChemAxon