



# CaffeineFix

Version 1.0 [201002]

NextMove Software Inc  
1303 Bartlet Court,  
Santa Fe,  
New Mexico, USA.

Tel: +1 505-216-9871  
Fax: +1 505-216-9778

[www.nextmovesoftware.com](http://www.nextmovesoftware.com)

## Licensing

CaffeineFix is licensed per site, as a toolkit with open source components, for \$10K USD. Academic and government discounts available. End-user node-locked licenses per-seat available upon request.

## Support

Support is offered at 15% per year and includes software updates and technical assistance.

## Languages

Support for multiple programming and scripting languages including C, C++, Java, Python, C#, VBA, VBScript and FORTRAN.

## Platforms

Available on multiple operating systems and architectures, including Microsoft Windows, Linux, UNIX and Mac OS X.

## ***Chemical Nomenclature Spelling Correction***

The complex technical nomenclature used by chemists to describe molecular structures presents novel challenges to regular natural language processing (NLP) tools. Software for handling English text often can't handle the non-standard use of whitespace, hyphenation, punctuation, Greek characters, italics and even superscripts found in chemical names. Likewise, the unusual letter combinations found in IUPAC and traditional names can trip up the trigram-analysis frequently used in spell checking software.

**CaffeineFix** overcomes the limitations of existing solutions by developing novel algorithms purely for handling IUPAC-like organic chemistry nomenclature. Unlike dictionary-based approaches, **CaffeineFix's** novel "*push-down automaton*" technology allows it to correct against an infinite number of words/chemical names. "*Levenshtein distance*" can be used to identify corrections and automatically correct unambiguous errors. User parameterizable substitution matrices and *indel* penalties can be used to optimize for a particular application, for example reducing the cost of substituting "rn" with "m", or "l" with "1" when working with scanned text, or the relative cost of deleting hyphens when processing paginated text.

**12-dichlorobenzne?** Did you mean **1,2-dichlorobenzene?**  
**didec-2-ene?** Did you mean **dodec-2-ene?**

- Correct typos in documents within word processors, spreadsheets, presentation and other office software.
- Enhance text-based searches of registration systems or chemical supplier catalogues, with Google-like "did you mean ...?" functionality.
- Fix OCR errors in scanned documents.
- Improve the performance of name-to-structure conversion in text-mining and chemical entity extraction applications.