

Groovy Cheminformatics...

Egon Willighagen 

Published February 6, 2011

Citation

Willighagen, E. (2011). Groovy Cheminformatics... In *chem-bla-ics*. chem-bla-ics. <https://doi.org/10.59350/8my4k-rfz51>

Keywords

Cdk, Java, Cheminf, Cdkbook

Copyright

Copyright © Egon Willighagen 2011. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Update: the [fourth edition](#) is out.

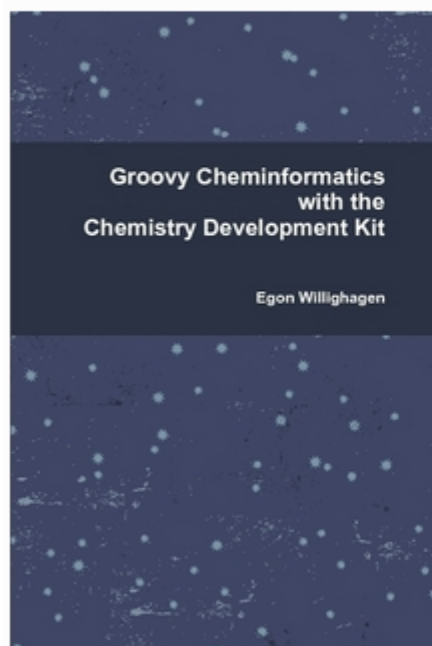
Some project are never finished. Neither is this one, but it is never too late to change how things work, so, taking advantage of publishing-on-demand, here I introduce the release-soon, release-often equivalent of cheminformatics books, my [Groovy Cheminformatics with the Chemistry Development Kit](#) book:

Groovy Cheminformatics with the Chemistry Development Kit

By Egon Willighagen

Paperback, 72 pages ★★★★★

This item has not been rated yet



List Price: ~~\$50.13~~

Price: **\$30.08**

You Save: **\$20.05 (40%)**

Ships in 3–5 business days

This book introduces the reader to the wide variety of functionality available in the Chemistry Development Kit (CDK) library. It will discuss parts of the data model, basic cheminformatics algorithms, chemical file formats, etc. It discusses bits of chemical graph theory, computer representation, etc. But the goal of this book is not to provide an introduction into cheminformatics. However, this book does require a basic chemical education. It assumes that you know what atoms are, how they are connected by chemical bonds, and it assumes some basic computer knowledge. This book is about learning how to perform cheminformatics tasks using the CDK. But to keep the required knowledge to a minimum, the examples will be pretty verbose. The book details the CDK data model, discusses input and output, atom types, graph properties, missing information, and substructure searching. It demonstrates the CDK library with 75 code examples.

With a serious discount for just being the first edition (1.3.8-0), but still counting at 72 pages with 75 code examples, this edition marks a personal milestone (and probably not much more than that). There remains much to do, but I promised a release by tomorrow, so here it is. Next releases will contain more code examples, more functionality descriptions, and more literature reviewing where such code is used in science. The plan is to make new editions with each new [CDK](#) release, as well as new editions when I added a new chapter, section, or just paragraph. But, there will not be a Nightly build service anytime soon.

The current table of content is as follows:

1	Introduction	1
	Bibliography	1
2	Atoms, Bonds and Molecules	3
2.1	Atoms	3
2.1.1	IElement	3
2.1.2	IIsotope	4
2.1.3	IAtomType	5
2.2	Bonds	5
2.3	Molecules	6
2.3.1	Iterating of atoms and bonds	7
2.3.2	Neighboring atoms and bonds	8
2.4	Implicit and Explicit Hydrogens	9
2.5	Chemical Objects	10
3	Salts and other disconnected structures	13
3.1	Salts	13
3.2	Crystals	13
	Bibliography	15
4	Paired and unpaired electrons	17
4.1	Lone Pairs	17
4.2	Unpaired electrons	18
5	Input/Output	19
5.1	File Format Detection	19
5.2	Example: Downloading Domoic Acid from PubChem	20
5.3	Gzipped files	21
5.4	Iterating Readers	22
5.4.1	MDL SD files	22
5.4.2	PubChem Compounds XML files	23
5.5	Customizing the Output	23
5.5.1	Setting Properties	25
5.6	Input Validation	26
5.6.1	Reading modes	26

Now, the book content is **not** open content. However, it contains nothing that is not available in other means. It's just the compilation that makes this book interesting, as well as that I put effort in ensuring the code examples remain working. For that, I ask a minor financial contribution.