FAIR_bioinfo for bioinformaticians

Introduction to the tools of reproducibility in bioinformatics

C. Hernandez[1] T. Denecker[1] J.Sellier[2] C. Toffano-Nioche[1]

[1] Institute for Integrative Biology of the Cell (I2BC)
UMR 9198, Université Paris-Sud, CNRS, CEA
91190 - Gif-sur-Yvette, France

[2] Institut Français de Bioinformatique
à compléter

Sept. 2020

Literate programming

Introduction

What is literate programming ?

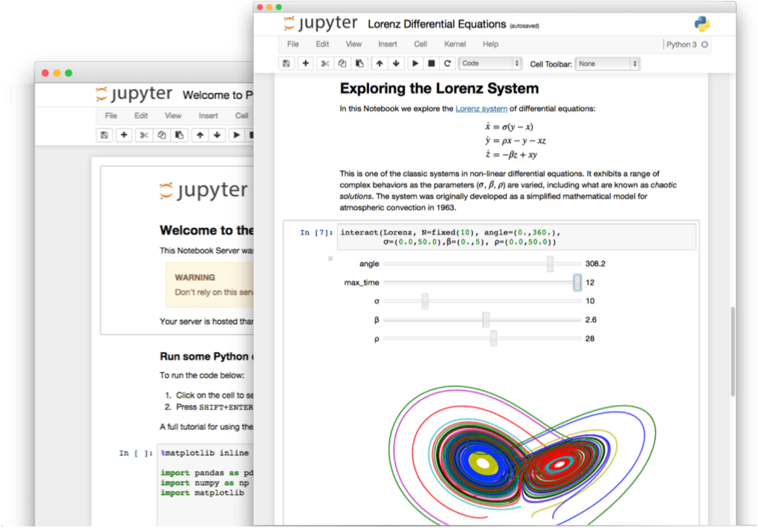
Definition

"Literate programming is a programming paradigm introduced by Donald Knuth in which a computer program is given an explanation of its logic in a natural language, such as English, interspersed with snippets of macros and traditional source code, from which compilable source code can be generated."

Wikipedia, 18/08/2020 (https://en.wikipedia.org/wiki/Literate_programming#Workflow)

Introduction

What does it look like ?



jupyter Welcome to P
File Edit View Insert Cell
jupyter

Welcome to the

This Notebook Server wa

WARNING

Don't rely on this serv

Your server is hosted tha

Run some Python

To run the code below:

1. Click on the cell to se
2. Press SHIFT+ENTER

A full tutorial for using the

```
In [ ]: %matplotlib inline

        import pandas as pd
        import numpy as np
        import matplotlib
```

jupyter Lorenz Differential Equations (autosaved)
File Edit View Insert Cell Kernel Help
Python 3
Code
Cell Toolbar: None

Exploring the Lorenz System

In this Notebook we explore the Lorenz system of differential equations:

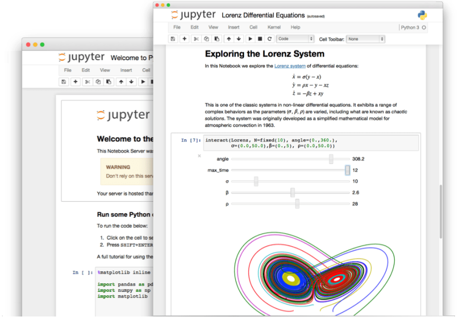
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

```
In [7]: interact(Lorenz, N=fixed(10), angle=(0.,360.),
                 σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0))
```

```
   angle  308.2
max_time  12
       σ  10
       β  2.6
       ρ  28
```

Introduction


jupyter Welcome to P
File Edit View Insert Cell
jupyter
Welcome to the
This Notebook Server wa
WARNING
Don't rely on this ser
Your server is hosted tha
Run some Python
To run the code below:
1. Click on the cell to se
2. Press SHIFT+ENTER
A full tutorial for using the
jupyter Lorenz Differential Equations (autosaved)
File Edit View Insert Cell Kernel Help
Python 3
Code
Cell Toolbar: None

```
In [ ]: %matplotlib inline

        import pandas as pd
        import numpy as np
        import matplotlib
```

Exploring the Lorenz System

In this Notebook we explore the Lorenz system of differential equations:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

```
In [7]: interact(Lorenz, N=fixed(10), angle=(0.,360.),
                 σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0))
```

Interactive programming interface allowing to combine both natural and computer languages.

In one file:

- Explanations
- Code
- Results
- Graphs and plots

Introduction

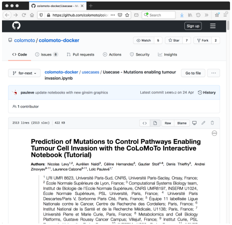
Why using literate programming frameworks ?

Use cases:

- Day to day analyses
- Analysis reports
- Writing scientific articles

Example of an article entirely written using a notebook

File (on a repository)



colomoto-docker / usecases / Usecase - Mutations enabling tumour invasion.ipynb

pauleve update notebooks with new ginsim graphics Latest commit 1498c2 on 24 Apr History

1 contributor

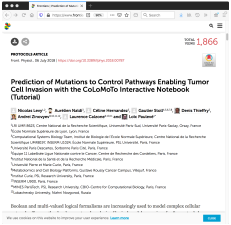
2513 lines (2513 sloc) 422 KB

Prediction of Mutations to Control Pathways Enabling Tumour Cell Invasion with the CoLoMoTo Interactive Notebook (Tutorial)

Authors: Nicolas Levy[1,2], Aurélien Naldi[3], Céline Hernandez[3], Gautier Stoll[4-8], Denis Thieffry[3], Andrei Zinovyev[9-11], Laurence Calzone[9-11], Loïc Paulevé[1,*]

> [1] LRI UMR 8623, Université Paris-Sud, CNRS, Université Paris-Saclay, Orsay, France; [2] École Normale Supérieure de Lyon, France; [3] Computational Systems Biology team, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, École Normale Supérieure, PSL Université, Paris, France; [4] Université Paris Descartes/Paris V, Sorbonne Paris Cité, Paris, France; [5] Équipe 11 labellisée Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers; Paris, France; [6] Institut National de la Santé et de la Recherche Médicale, U1138; Paris, France; [7] Université Pierre et Marie Curie, Paris, France; [8] Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus; Villejuif, France; [9] Institut Curie, PSL

Published article



PROTOCOLS ARTICLE

Front. Physiol., 06 July 2018 | https://doi.org/10.3389/fphys.2018.00787

Prediction of Mutations to Control Pathways Enabling Tumor Cell Invasion with the CoLoMoTo Interactive Notebook (Tutorial)

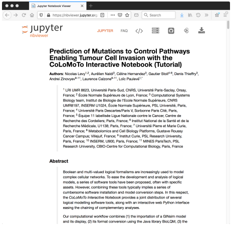
Nicolas Levy[1,2], Aurélien Naldi[3], Céline Hernandez[3], Gautier Stoll[4,5,6,7,8], Denis Thieffry[3], Andrei Zinovyev[9,10,11,12], Laurence Calzone[9,10,11] and Loïc Paulevé[1*]

[1]LRI UMR 8623, Centre National de la Recherche Scientifique, Université Paris-Sud, Université Paris-Saclay, Orsay, France
[2]École Normale Supérieure de Lyon, Lyon, France
[3]Computational Systems Biology Team, Institut de Biologie de l'École Normale Supérieure, Centre National de la Recherche Scientifique UMR8197, INSERM U1024, École Normale Supérieure, PSL Université, Paris, France
[4]Université Paris Descartes, Sorbonne Paris Cité, Paris, France
[5]Équipe 11 Labellisée Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers, Paris, France
[6]Institut National de la Santé et de la Recherche Médicale, Paris, France
[7]Université Pierre et Marie Curie, Paris, France
[8]Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus, Villejuif, France
[9]Institut Curie, PSL Research University, Paris, France
[10]INSERM U900, Paris, France
[11]MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France
[12]Lobachevsky University, Nizhni Novgorod, Russia

Boolean and multi-valued logical formalisms are increasingly used to model complex cellular

We use cookies on this website to improve your user experience. Learn more CLOSE

Executable file



Prediction of Mutations to Control Pathways Enabling Tumour Cell Invasion with the CoLoMoTo Interactive Notebook (Tutorial)

Authors: Nicolas Levy[1,2], Aurélien Naldi[3], Céline Hernandez[3], Gautier Stoll[4-8], Denis Thieffry[3], Andrei Zinovyev[9-11], Laurence Calzone[9-11], Loïc Paulevé[1,*]

> [1] LRI UMR 8623, Université Paris-Sud, CNRS, Université Paris-Saclay, Orsay, France; [2] École Normale Supérieure de Lyon, France; [3] Computational Systems Biology team, Institut de Biologie de l'Ecole Normale Supérieure, CNRS UMR8197, INSERM U1024, École Normale Supérieure, PSL Université, Paris, France; [4] Université Paris Descartes/Paris V, Sorbonne Paris Cité, Paris, France; [5] Équipe 11 labellisée Ligue Nationale contre le Cancer, Centre de Recherche des Cordeliers; Paris, France; [6] Institut National de la Santé et de la Recherche Médicale, U1138; Paris, France; [7] Université Pierre et Marie Curie, Paris, France; [8] Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus; Villejuif, France; [9] Institut Curie, PSL Research University, Paris, France; [10] INSERM, U900, Paris, France; [11] MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France

Abstract

> Boolean and multi-valued logical formalisms are increasingly used to model complex cellular networks. To ease the development and analysis of logical models, a series of software tools have been proposed, often with specific assets. However, combining these tools typically implies a series of cumbersome software installation and model conversion steps. In this respect, the *CoLoMoTo Interactive Notebook* provides a joint distribution of several logical modelling software tools, along with an interactive web Python interface easing the chaining of complementary analyses.
>
> Our computational workflow combines (1) the importation of a GINsim model and its display, (2) its format conversion using the Java library BioLQM, (3) the

Literate programming

This session :

- Markdown
- Rmarkdown and RStudio
- Jupyter

Markup and markdown

Definition

A markup language uses tags to define elements within a document.

Three different types and usage :

- Presentational (used by traditional word-processing systems)
  - Markup is invisible
- Procedural, provides instructions to process the text (e.g. TeX, PostScript)
  - Markup is visible and can be directly manipulated by the author.
- Descriptive, to label documents parts (e.g. LaTeX, HTML, XML...)
  - Emphasizes the document structure.

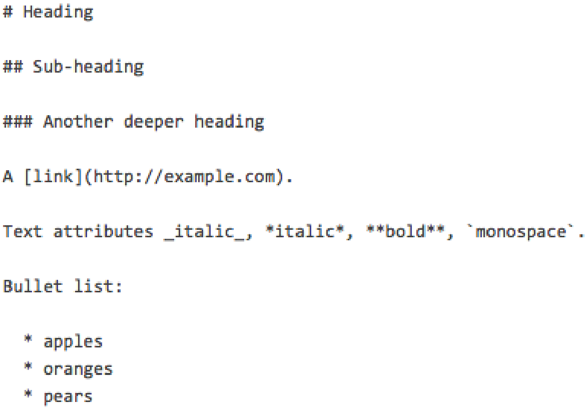
Markdown language

Markdown is a Lightweight markup language.
Designed to be :

- easy to write using any generic text editor (plain-text-formatting syntax)
- easy to read in its raw form

Markdown language

You've probably see it already on GitHub (README), Wikipedia...

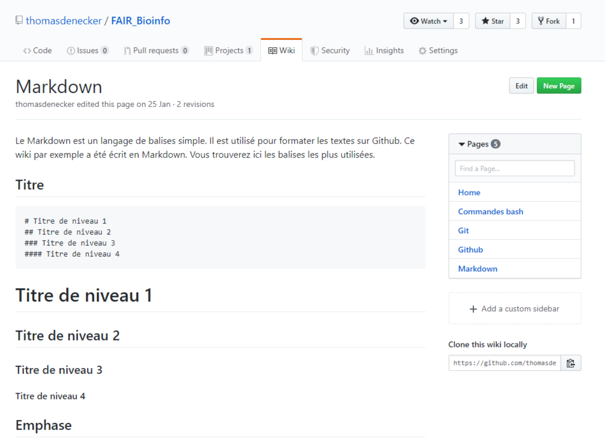


```
# Heading

## Sub-heading

### Another deeper heading

A [link](http://example.com).

Text attributes _italic_, *italic*, **bold**, `monospace`.

Bullet list:

  * apples
  * oranges
  * pears
```

Github guides : https://guides.github.com/features/mastering-markdown/

Markdown language

Une page de wiki est disponible sur FAIR_bioinfo
https://github.com/thomasdenecker/FAIR_Bioinfo/wiki/Markdown



Markdown

thomasdenecker edited this page on 25 Jan · 2 revisions

Le Markdown est un langage de balises simple. Il est utilisé pour formater les textes sur Github. Ce wiki par exemple a été écrit en Markdown. Vous trouverez ici les balises les plus utilisées.

Titre

```
# Titre de niveau 1
## Titre de niveau 2
### Titre de niveau 3
#### Titre de niveau 4
```

Titre de niveau 1

Titre de niveau 2

Titre de niveau 3

Titre de niveau 4

Emphase