Python in Bioinformatics

Gang Chen
chengang@bgitechsolutions.com

November 8, 2014

华大科技 BGI·tech
香港中文大學 The Chinese University of Hong Kong

Outline

Next

Python Package

Python Package Development

Example

see hello directory

Next

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

scipy project

Installation and Example

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

BioPython

Installation and Example

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

Python for Machine Learning

- scikit-learn
- pyml

scikit-learn

PyML is an interactive object oriented framework for machine learning written in Python. PyML focuses on SVMs and other kernel methods.

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

igraph for network visualization

Next

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

Clouding Computing

- Amazon Web Service: aws.amazon.com
- Aliyun: aliyun.com
- Google Compute Engine: cloud.google.com
- Microsoft Azure: azure.microsoft.com
- ...

Bioinformatics in the Cloud

- DNANexus: DNANexus.com
- SBGenomics: SBGenomics.com
  rabix: rabix.org
- GeneDock: GeneDock.com
- L3-Bioinformatics: l3-bioinfo.com
- tute genomics, Variant Analysis from Qiagen, ...

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

Overview

Installation

Next

1. Python Package
2. Python based Bioinformatics Projects
   - Scipy
   - BioPython
   - Machine Learning
   - Network Visualization and Analysis
3. Bioinformatics in the Cloud using Python
   - Clouding Computing and Bioinformatics
   - Python SDK of DNANexus
   - Rabix from SBGenomics

Reproducible Research

Reproducible Research

The goal of reproducible research is to tie specific instructions to data analysis and experimental data so that scholarship can be recreated, better understood and verified.

References

- https://www.coursera.org/course/repdata
- http://cran.r-project.org/web/views/ReproducibleResearch.html

Rabix Project

see rabix_ismb.pdf