

Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

bartek@mimuw.edu.pl

<http://regulomics.mimuw.edu.pl>

Wykład 2. - Projekty BioX

4. III. 2020

Plan na dziś

- Bioperl i trochę historii
- Open Bioinformatics Foundation
- Software in the Public Interest
- Biopython, Biojava, Bioruby
- BioSQL, BioDAS, EMBOSS
- Bioinformatics Open Source Conference
- Google Summer of Code
- Przegląd aktywności projektów
- Biopython trochę bardziej z bliska

Ciekawe daty z historii oprogramowania bioinformatycznego

- 1970 – słowo bioinformatyka (Hogeweg)
- 1978 – macierze PAM
- 1981 - algorytm Smitha-Watermana
- 1982 - Genbank
- 1987 – serwer bioinfo w EMBL (początki EMBOSS)
- 1988 - Algorytm Clustal
- 1991 - BLAST
- 1996 – BioPerl (release 0.7 w 2000)
- 1999 – BioPython
- 2001 – Bioconductor
- 2005 – Galaxy project

BioPerl

- Założony około roku 1996
- Pierwsze “wydanie” 0.7 – 2000
- Wersja 1.0 – 2002
- 2002 - The Bioperl Toolkit: Perl Modules for the Life Sciences JE. Stajich et al. *Genome Res*
- 2002 – powstaje Open Bioinformatics Foundation
- Do 2010 niesłabnąca popularność

Open Bioinformatics Foundation

- Zarządzanie projektami, hosting, storage, backups, etc.
- Uformowana w 2001/2002 na potrzeby projektów bioperl/piopython/biojava
- W 2005 zyskała formalną “osobowość” jako fundacja not-for-profit
- W 2012 przekazano prawa do zarządzania fundacją na rzecz Software in the public interest

Software in the public interest

- Organizacja Not for profit
- Zajmuje się projektami open source, nie tylko bioinformatycznymi (Haskell, ffmpeg, Debian,...)
- Zbiera wpłaty, podpisuje umowy, oferuje pomoc prawną
- Demokratyczne zarządzanie
- Pomaga organizować sprzęt i usługi dla utrzymania projektów
- “Freedom to leave”

Różne projekty BioX

- BioPerl
- BioPython
- BioJava
- BioSQL
- BioRuby
- BioHaskell
- BioJS
-

Rozwój BioPerl'a

GitHub

This repository Search

Explore Features Enterprise Blog

Sign up

Sign In

bioperl / bioperl-live

Watch 54

Star 223

Fork 157

Code

Issues 37

Pull requests 2

Actions

Projects 0

Security

Insights

Pulse

Contributors

Commits

Code frequency

Dependency graph

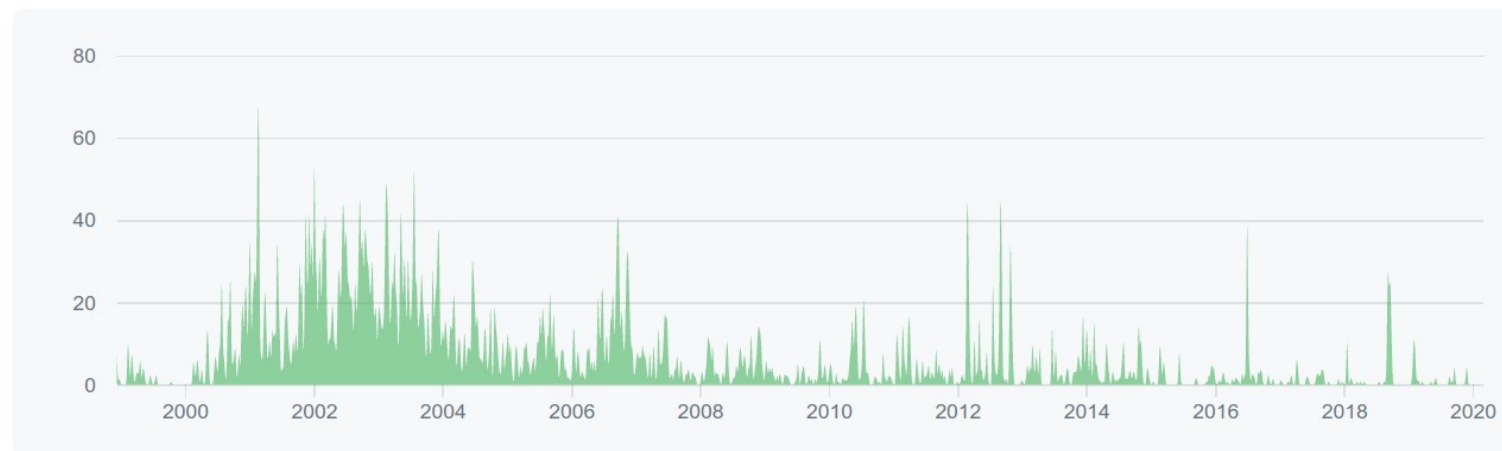
Network

Forks

Dec 13, 1998 – Mar 4, 2020

Contributions: **Commits**

Contributions to master, excluding merge commits



Użycie Biojava

biojava / biojava

Watch 68

Star 349

Fork 269

Code

Issues 61

Pull requests 1

Actions

Security

Insights

Pulse

Contributors

Commits

Code frequency

Dependency graph

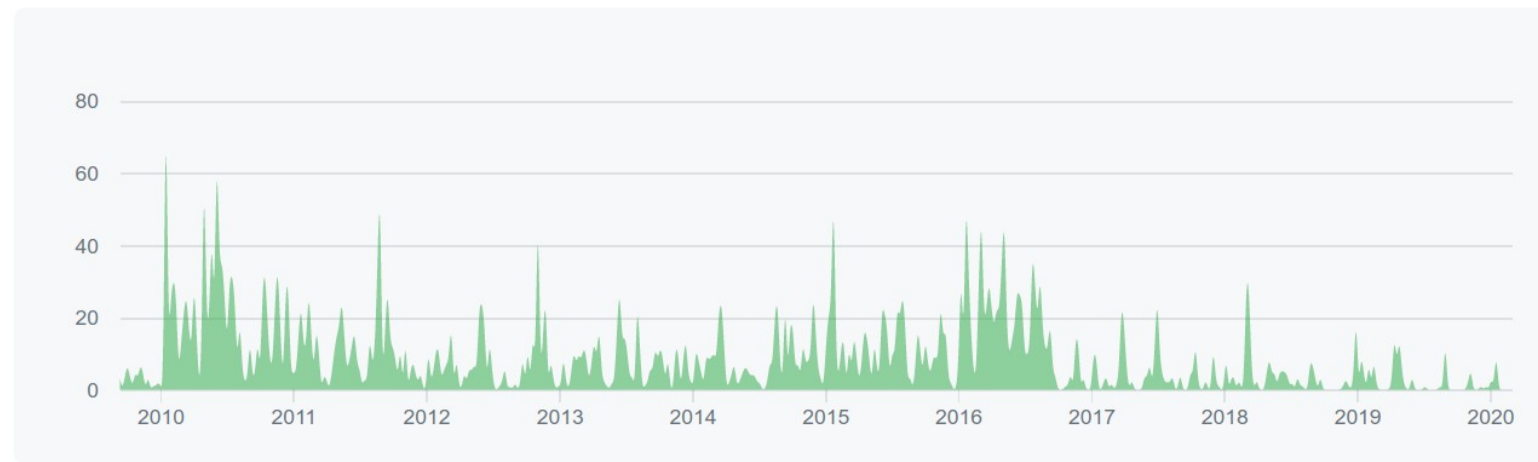
Network

Forks

Sep 13, 2009 – Mar 4, 2020

Contributions: **Commits**

Contributions to master, excluding merge commits



Bioruby...

 bioruby / **bioruby**

 Watch 32

 Star 348

 Fork 114

 Code

 Issues **10**

 Pull requests **3**

 Actions

 Projects **0**

 Wiki

 Security

 Insights

[Pulse](#)

Contributors

[Commits](#)

[Code frequency](#)

[Dependency graph](#)

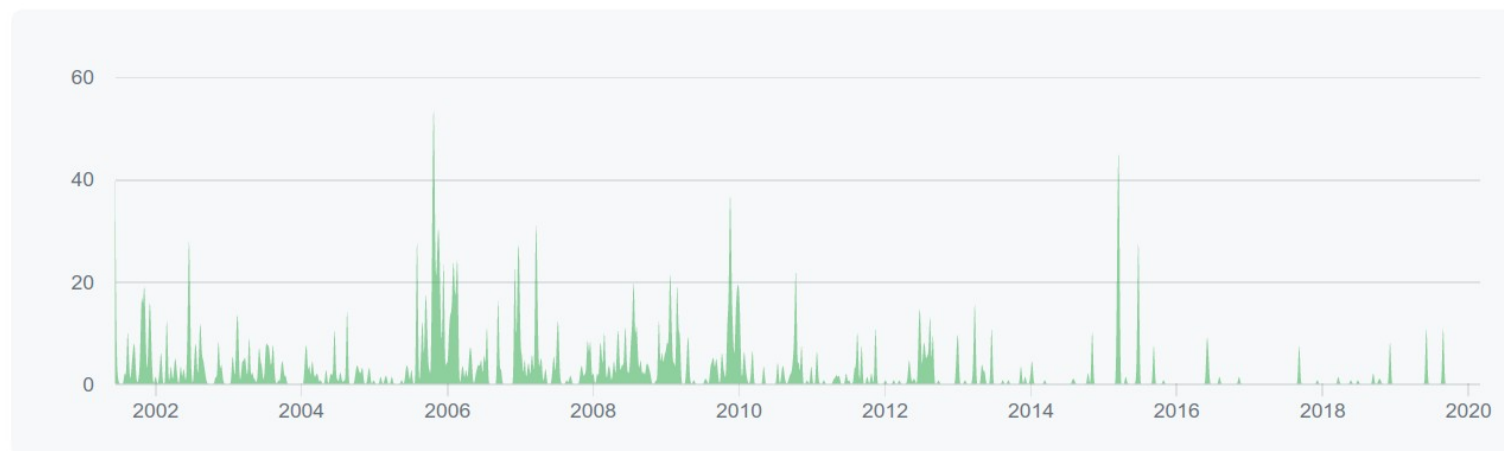
[Network](#)

[Forks](#)

Jun 17, 2001 – Mar 4, 2020

Contributions: **Commits** ▾

Contributions to master, excluding merge commits



Użycie BioPython'a

GitHub

This repository Search

Explore Features Enterprise Blog

Sign up

Sign in

biopython / biopython

Watch 149

★ Star 2.1k

Fork 1.1k

Code

Issues 325

Pull requests 93

Actions

Security

Insights

Pulse

Contributors

Commits

Code frequency

Dependency graph

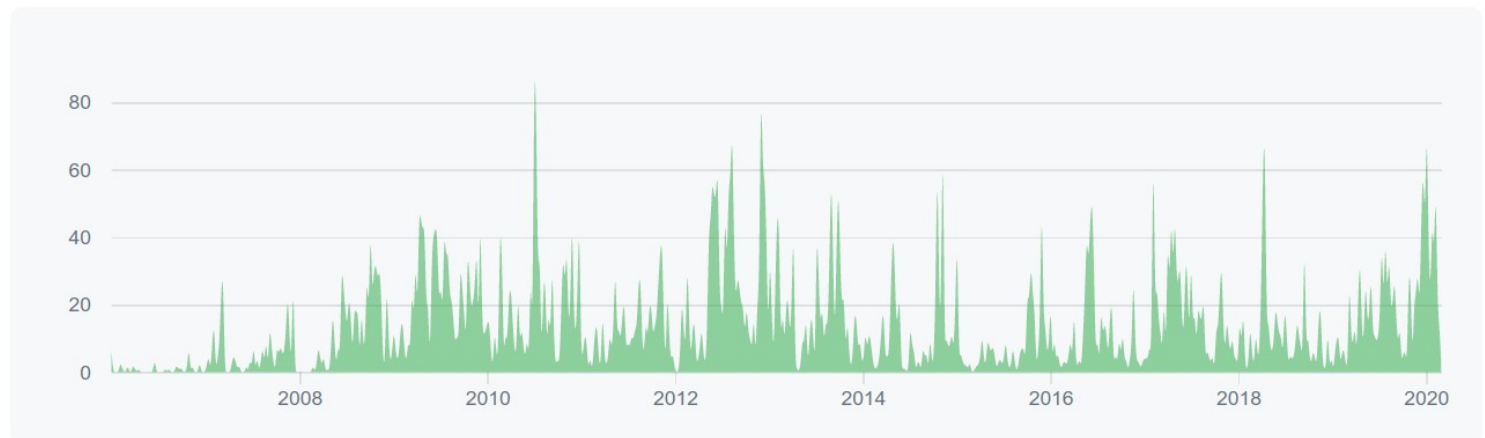
Network

Forks

Jan 1, 2006 – Mar 4, 2020

Contributions: Commits

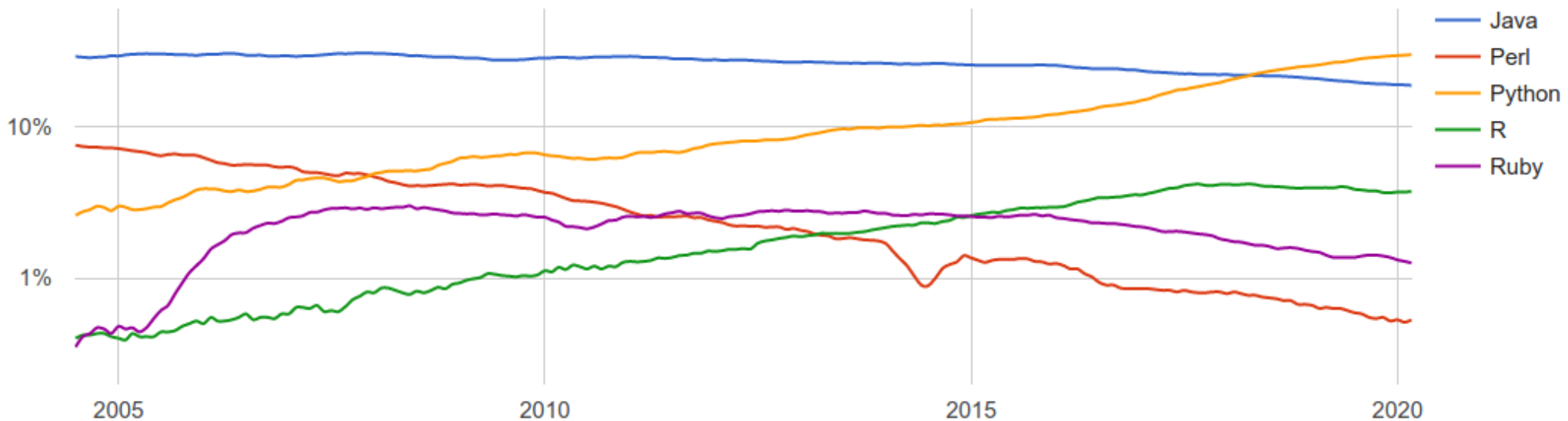
Contributions to master, excluding merge commits



Globalne trendy wg PYPL

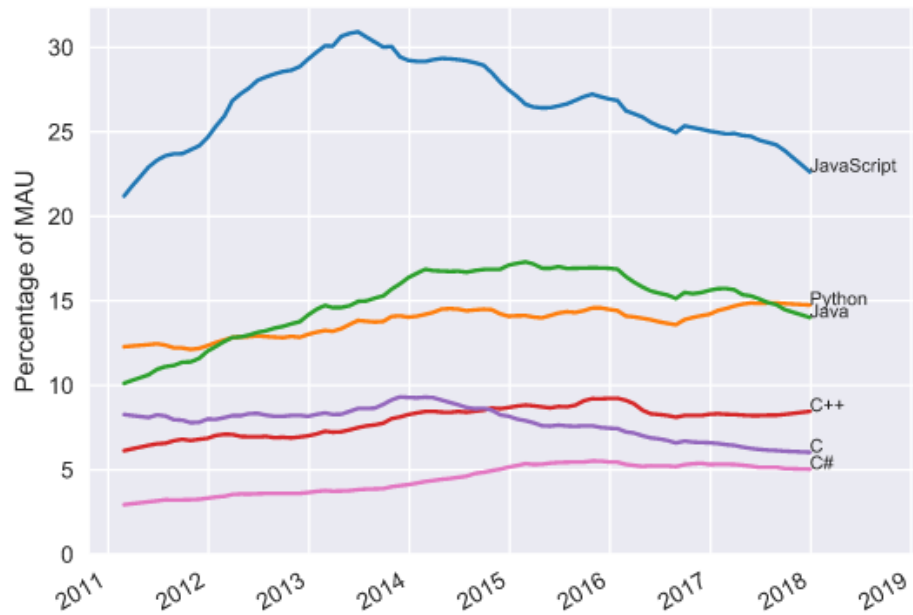
Worldwide, Python is the most popular language, Python grew the most in the last 5 years (19.0%) and Java lost the most (-6.8%)

PYPL PopularitY of Programming Language



This chart uses a [logarithmic scale](#). It can show your favorite languages in a country

Nieco inne wg github'a



Rank	Language	Monthly Active Users	Trend
1	JavaScript	22.63%	
2	Python	14.75%	
3	Java	14.01%	
4	C++	8.45%	
5	C	6.03%	
6	PHP	5.85%	
7	C#	5.03%	
8	Shell	4.85%	
9	Go	4.10%	
10	TypeScript	3.89%	
11	Ruby	3.27%	
12	Jupyter Notebook	2.74%	
13	Objective-C	1.99%	
14	Swift	1.89%	
15	Kotlin	1.28%	
16	R	0.81%	

Inne projekty BioX

- Bio-DAS – Distributed Annotation System
- Bio-SQL – serialization in SQL
- Bio-MOBY – index of webservices
- GMOD – generic model organism database

Bioinformatics Open Source Convention

← → ↻ open-bio.org/bosc2000/



Bioinformatics Open Source Conference

The Bioinformatics Open Source Conference (BOSC) is the successor to the successful [bioperl-99](#) conference. Like bioperl-99, BOSC is a satellite conference of [ISMB](#), allowing people who are making the trip to San Diego to extend it for a couple of days and talk real code.

BOSC is designed to be open to all the open source efforts in bioinformatics, including Perl, Java, Python, C and C++ - even Fortran would be fun. We **don't** expect all attendees to be participating in a particular open source project (like, say, [biojava](#)), but we do expect that a lot of the people involved in the open source projects in bioinformatics will attend.

BOSC will be organized around presentations selected from submitted abstracts. Confirmed speakers include:

- Tim O'Reilly on Open Source
- Lincoln Stein on Genome Annotation
- Ewan Birney on Bioperl

Key dates:

- April 15** Registration opens
- June 30** Registration closes
- June 30** Abstract deadline
- August 17** BOSC starts, 1pm
- August 18** BOSC ends, 7pm
- August 19** Tutorial day at ISMB
- August 20** ISMB conference presentations

check out the [ISMB2000 pages](#) for information on hotels, travel to and from San Diego as well as other relevant information!

[REGISTRATION \(Abstract Submission is closed!\)](#)

[CONFERENCE SCHEDULE & ABSTRACTS](#)

The [BOSC organizing committee](#) is:

- Ewan Birney (EBI)
- Chris Dagdigan (The Genetics Institute)
- Nomi Harris (UC Berkeley)
- Helge Weissig (UCSD/San Diego Supercomputer Center)

Bioinformatics Open Source Convention 2020

[Home](#) > [Events](#) > **BOSC 2020**

BOSC 2020

BCC2020, the Bioinformatics Community Conference, will bring together the Galaxy Community Conference (GCC) and the Bioinformatics Open Source Conference (BOSC) to meet jointly at Victoria University in Toronto, Canada.

The main BCC conference will take place **July 19-21, 2020**, with an optional pre-conference training day July 18 and CoFest and CoFest Encore days July 22-25.

The Bioinformatics Open Source Conference (BOSC) has been held annually since 2000. BOSC 2019 was part of ISMB/ECCB in Basel.



Key Dates

- March, 2020: Abstract submission opens
- April 1: Deadline to apply for OBF Travel Fellowship
- April 17: Abstract submission closes
- May 12: Authors notified
- May 22: Early registration ends
- May 29: Late poster/late-breaking lightning talk submission closes
- June 12: Advance registration ends
- July 18: BCC2020 optional tutorials, Toronto
- **July 19-21, 2020: BOSC 2020**, part of BCC2020
- July 22-23 and 24-25: CoFest and CoFest Encore

Session Topics

BOSC covers all aspects of bioinformatics software and open science, including (but not limited to) these topics:

- Open Science and Reproducible Research
- Open Biomedical Data
- Citizen/Participatory Science
- Standards and Interoperability
- Data Science
- Workflows
- Open Approaches to Translational Bioinformatics
- Developer Tools and Libraries
- Inclusion, Outreach and Training
- Bioinformatics Open Source Project Reports (about new or existing projects)

Overview

BOSC is organized by the [Open Bioinformatics Foundation \(OBF\)](#), a non-profit group dedicated to promoting the practice and philosophy of open source software development and open science within the biological research community.

Google summer of Code

- Organizowany przez google od 2005 roku
- Zbiera potencjalne projekty, mentorów i oferuje studentom niewielkie wynagrodzenie za zaangażowanie w projekt open Source
- W sumie ponad 15000 studentów, dziesiątki milionów linii kodu co roku
- OBF czynnym uczestnikiem od 2009
- Oprócz tego, m.in. Biogears, Canadian Centre for Computational Genomics, National Resource for Network Biology, Open Chemistry, Open EMR, OpenMRS, Open Genome Informatics

Open Bioinformatics Foundation (OBF)

<https://www.open-bio.org>


Promoting practice & philosophy of OSS & Open Science in biological research.

The Open Bioinformatics Foundation (OBF) is a non-profit, volunteer-run group dedicated to promoting the practice and philosophy of Open Source software development and Open Science within the biological research community.

Our main activities are:

- Running and supporting the BOSC conferences.
- Organizing and supporting developer-centric "Hackathon" events.
- Participating in the Google Summer of Code program on behalf of our member projects as an umbrella mentoring organization.
- Managing servers, colocation facilities, bank account, domain names, and other assets for the benefit of our member projects.



 [VIEW IDEAS LIST](#)

 Proposals accepted starting on March 16, 2020

Technologies

python

javascript

c/c++

Topics

Science and Medicine

bioinformatics

computational biology

 [MAILING LIST](#)



Pomysły na projekty w ramach OBF

[About](#)[Events](#)[Projects](#)[Travel Awards](#)[News](#)[Home](#) > [Events](#) > [GSoC](#) > **Project Ideas**

Google Summer of Code 2020 Project ideas

Quick links

- [Cross-Project Ideas](#)
- [Protein Database Suitability by DeNovo Sequencing](#)
- [OpenMS R-package](#)
- [NGLess: Expanding data types](#)
- [NGLess: Integration with nixpkgs](#)
- [NGLess: \(JIT\) Compilation of NGLess](#)
- [NGLess: Improved reporting of results](#)
- [SIMDe: Add new implementations of ISA extensions](#)
- [SIMDe: Add new ISA extensions](#)
- [pysradb – Enhancing search for next-generation sequencing datasets](#)
- [The fastest GPU Variation Graph explorer \(VG\)](#)
- [Generative biolearn project](#)
- [Generate CWL CommandLineTool descriptor from Rust clap CLI description](#)
- [Fuzz the CWL Reference Runner](#)

Cross-Project Ideas

OBF is an umbrella organization which represents many different programming languages used in bioinformatics. In addition to working with each of the "Bio*" projects (listed below) we also accept "cross-project" ideas that cover multiple programming languages or projects. These collaborative ideas are broadly defined and can be thought of as "unfinished" – interested students should adapt the ideas to their own strengths and goals, and are responsible for the quality of the final proposed idea in their application.

Feel free to propose your own entirely new idea.

Biopython

- Biopython.org
- Źródła na githubie
- Narzędzia do:
 - analizy sekwencji
 - Serializacji,
 - PDB,
 - text-mining,
 - Grafika,
 - ...

Ważne kwestie dot. Biopythona

- Jak kod projektu jest zarządzany na githubie
- Contributing howto
- Git-usage
- Testing framework
- Mailing lists (biopython,)
- Documentation guidelines
- licenses



Python Tools for
Computational
Molecular Biology

Documentation

Download

Mailing lists

News

Biopython Contributors

Scriptcentral

Source Code

GitHub project

Biopython version 1.76

© 2019. All rights
reserved.

Mailing lists

Discussion and questions

We actively discuss Biopython and Python usage in Bioinformatics on a few dedicated mailing lists. Please join us; to avoid spam, you can post a message only if you are subscribed to the mailing list.

A few other useful places for questions and answers - but please don't duplicate your questions here and on the mailing list:

- [BioStar](#) – A dedicated Bioinformatics question and answer forum. Please tag your post with: Biopython.
- [StackOverflow](#) – The main Stack Overflow programming question and answer forum. Please tag your posts with: Biopython.

Searching

Google is a practical way to search the mailing list archives. A number of questions have been discussed and answered; before posting please search for relevant discussions which might help.

Discussion List

biopython@biopython.org

Unmoderated list for use, comments, help, and other things related to Biopython. Since January 2018, this includes any development discussions not held on GitHub.

Only subscribers can send to the mailing list. Sadly due to the server load from spam, any emails from a non-subscriber are silently ignored. If your message fails to appear in the archive, double check you are subscribed with the exact address your sent emails from.

[List Website](#) | [Web Archive](#) | [Example Google Search](#)

Development List (closed since January 2018)

biopython-dev@biopython.org

This was an unmoderated list for Biopython developers and interested people. Since we moved to GitHub development discussion here dropped, and as of January 2018 we have closed this list - all subscribers were added to the main discussion list (see above).

Procedura dodania kodu

- Komunikacja z core teamem
- Rozwój rozproszony (na osobnej gałęzi)
 - Zgodność kodu z konwencjami
 - Dokumentacja
 - Testy
- Pull request
- Integracja z główną gałęzią i zarządzanie kodem
- Wejście do oficjalnych wydań

Pierwsze zadanie (indywidualne)

- Każdy uczestnik kursu powinien wybrać sobie modyfikację do kodu biopythona (lub innego projektu open source dostępnego na githubie)
- Ustalić ze mną czy jest ona sensowna (nie musi to być nic wielkiego)
- Wykonać fork w serwisie github i dokonać zmiany
- Dać pull request do mnie