

Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

bartek@mimuw.edu.pl

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

Wykład 2. - Projekty BioX

6. III. 2019

Plan na dziś

- Bioperl i trochę historii
- Open Bioinformatics Foundation
- Software for 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

- 1965 – macierze PAM
- 1970 – słowo bioinformatyka (Hogeweg)
- 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

★ Star 131

Fork 81

Contributors

Commits

Code frequency

Punch card

Network

Members

Dec 6, 1998 – Oct 13, 2014

Contributions to master, excluding merge commits

Contributions: Commits



Użycie BioPython'a

GitHub

This repository Search

Explore Features Enterprise Blog

Sign up

Sign in

biopython / biopython

★ Star 505

🍴 Fork 300

Contributors

Commits

Code frequency

Punch card

Network

Members

Dec 5, 1999 – Oct 13, 2014

Contributions to master, excluding merge commits

Contributions: Commits



Użycie Biojava

GitHub

This repository Search

Explore Features Enterprise Blog

Sign up

Sign in

biojava / biojava

★ Star 46

Fork 68

Contributors

Commits

Code frequency

Punch card

Network

Members

Sep 13, 2009 – Oct 11, 2014

Contributions: **Commits** ▾

Contributions to master, excluding merge commits



Bioruby...

GitHub

This repository Search

Explore Features Enterprise Blog

Sign up

Sign In



bioruby / bioruby

★ Star

248

Fork

100

Contributors

Commits

Code frequency

Punch card

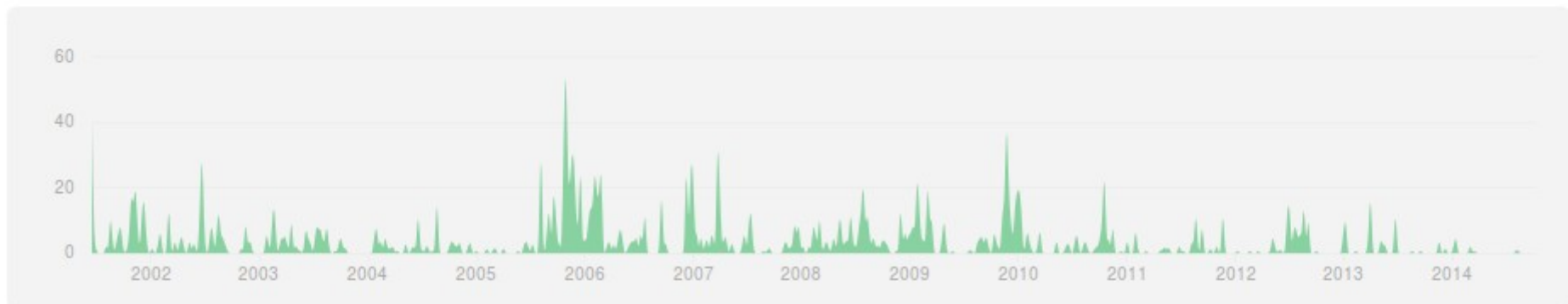
Network

Members

Jun 17, 2001 – Oct 13, 2014

Contributions: **Commits** ▾

Contributions to master, excluding merge commits



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

BOSC 2017 Schedule - Open Bioinformatics Foundation - Chromium

BOSC 2017 Schedul x Bartek

Secure | https://www.open-bio.org/wiki/BOSC_2017_Schedule

Create account Log in

O|B|F


Page Discussion Read View source View history Search Open Bioinformatics Foundation

BOSC 2017 Schedule

Complete Program

This page has links to all the publicly available talk slides, videos and posters. Most of the slides and posters are on our [BOSC F1000 Research Collection](#) (others are on SlideShare and FigShare), while the videos are on the [BOSC 2017 YouTube playlist](#).

- [Download the complete program \(including all talk and poster abstracts\) as a PDF](#)
- [ISMB/ECCB schedule for BOSC, including brief descriptions of each talk topic](#)



Pre-BOSC: Codefest 2017

[Codefest 2017](#) (a pre-BOSC hackathon) will take place July 20-21 at [Brmlab, Prague](#) (a non-profit, community-run hackerspace). Codefest is FREE and open to all (even if you're not attending BOSC). See [Codefest 2017](#) for details and to sign up.


Schedule at a glance

Day 1 (Saturday, July 22)		Day 2 (Sunday, July 23)	
Time	Session	Time	Session
8:30-9:30	(ISMB announcements & ISMB keynote)	8:30-9:30	(ISMB announcements & ISMB keynote)
10:00-10:10	BOSC announcements	10:00-10:05	BOSC announcements
10:10-10:45	Introduction to the Open Bioinformatics Foundation; OBF in the Google Summer of Code; Codefest 2017 Report	10:05-11:30	Session: Community Building and Citizen Science
10:45		11:30-12:20	Late-Breaking Lightning Talks

https://www.open-bio.org/wiki/BOSC_2017 s

BOSC 2018 – wspólnie z Galaxy Community conference



 GCC BOSC 2018 Attending this event? [SIGN UP](#) [LOG IN](#)

 OHSU	 GCC2018 Galaxy Community Conference		 BOSC Portland 2018		 BOSC 2018 Bioinformatics Open Source Conference	 REED COLLEGE
A bioinformatics			community conference			

The 2018 Galaxy Community Conference (GCC2018) and Bioinformatics Open Source Conference 2018 (BOSC2018) are joining together for a unified event to be held in [Portland, Oregon](#), United States, June 25-30, 2018. The event includes [two days of training](#), a [two day meeting](#), and finishes with [two or four days of intense collaboration at the GCC OBF CollaborationFest Core and Encore](#) events. The meeting will feature joint and parallel sessions, [shared keynotes](#), poster and demo sessions, and social events. Colocating GCC and BOSC will bring together the widest possible community of bioinformatics developers and practitioners into a single place. At this event the whole will be greater than the sum of the parts.

The event is organized by [Oregon Health & Science University](#) and will be held at [Reed College](#), a small liberal arts college in Portland. Conference housing is available in [two Reed College residence hall complexes](#). [Portland](#) is a vibrant city in the American Northwest, close to the Cascade Mountains, the Oregon Coast, and the Columbia River.

- [Registration](#) and [lodging](#) information
- [Submit an abstract](#) - oral presentations due March 16

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 13000 uczestników, dziesiątki milionów linii kodu co roku
- OBF czynnym uczestnikiem od 2009

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, biopython-dev)
- Documentation guidelines
- licenses

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