

Architektura dużych projektów bioinformatycznych

Bartek Wilczyński

bartek@mimuw.edu.pl

<http://www.mimuw.edu.pl/~bartek>

Wykład 1. - Wprowadzenie

5. X. 2015

Cele do wykonania

- Zapoznanie się z narzędziami do współpracy nad tworzeniem oprogramowania (git, github itp.)
- Przegląd powszechnie dostępnych projektów implementujących istotne rozwiązania dla bioinformatyki
- Rola otwartego kodu dla wiarygodności i reprodukowalności badań naukowych
- Praktyczne zajęcia pozwalające “oswoić” się z wybranym dużym projektem bioinformatycznym
- Wykonanie projektu zespołowego

Plan wykładów

- 1.Wprowadzenie
- 2.Projekty BioX (Perl, Python, Java, itp...)
- 3.Oprogramowania typu LIMS (Galaxy, Base itp)
- 4.Workflows (Taverna, Galaxy, itp)
- 5.Przeglądarki genomowe (UCSC, Ensembl, itp)
- 6.Wizualizacja (Cytoscape, Circos, itp)
- 7.Oprogramowanie statystyczne (Bioconductor itp)
- 8.-15. Referaty studentów

Plan laboratoriów

Zagadnienia:

- systemy kontroli wersji (Subversion i GIT)
- Python (modyfikacje istniejących projektów/tworzenie własnych, automatyzacja testów, współpraca online)
- Galaxy (tworzenie workflow'ów w Galaxy)
- Circos (Wizualizacja w środowisku skryptowym)

Zaliczenie: cztery zadania zaliczeniowe
(w tym jedno duże zadanie zespołowe)

- Kontakt: **Paweł Bednarz**
p.bednarz@mimuw.edu.pl

Kryteria zaliczenia

- Zaliczenie laboratorium
 - Suma punktów z zadań daje proponowaną ocenę z ćwiczeń (musi być powyżej 2, żeby móc zaliczyć przedmiot)
- Zaliczenie wykładu
 - Referat na ocenę
- Średnia ocena z ćwiczeń i referatu daje punkt wyjścia do egzaminu ustnego, na którym można ocenę zmienić.

Proponowane tematy referatów

- Przegląd narzędzi open-source do:
 - Biologii strukturalnej (białka, RNA, leki)
 - Filogenetyki
 - Chemoinformatyki
 - Hurtownie danych (Biomart itp)
 - Genomiki funkcjonalnej (chip-seq itp)
 - Machine learning (klasowanie, klasyfikacja itp)
 - Przetwarzanie obrazów (z mikroskopów, skanerów itp)
 - Własne sugestie...?

Schemat dobrego referatu

- Elementarne wprowadzenie do tematyki i roli w niej oprogramowania (co badamy, po co badamy, dlaczego potrzebujemy do tego niebanalnego oprogramowania)
- Przegląd typowych problemów w danej dziedzinie i przykłady popularnego oprogramowania do ich rozwiązywania, z uwzględnieniem:
 - Licencji (open-source, komercyjne)
 - Sposobu zarządzania kodem
 - Krótkiej historii projektu

Parę słów o OpenSource

- Free software to nie tylko Open Source
- Projekt GNU
- Licencje GNU – GPL w różnych wersjach
- Czy Free Software to “Public Domain”?
- Czy można zarobić na Open Source?
- A co z patentami?
- I co to ma wspólnego z bioinformatyką?

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