

Scientific Computing

- Publications
- Center for Biomedical Computing
- Projects
- Available Master's topics
- Intranet
- People

CBC Seminar on Python for Interactive, Large-scale Biomedical Computing - September 23, 2009

Center for Biomedical Computing (CBC) has a decade of experience with Python for scientific computing, and in particular the techniques that are needed to speed up Python for extremely computationally extensive large-scale problems. A Norwegian Center for Research Based Innovation (SFI), Statistics for Innovation (sfi)², with researchers from the Norwegian Computing Centre, Department of Informatics, and others, works on a comprehensive system for genome analysis, implemented in Python. The experience with Python for large-scale biomedical computing from this project and the unsolved efficiency issues are of interest to the scientific Python programmers at CBC. A one-day seminar was set up to exchange ideas between the two groups and discuss possible ways of collaborating on biomedical computations.

Total number of participants: 7

Total number of guests outside of CBC: 3

Number of different nationalities represented: 2

Total number of talks: 1

Plan for the workshop:

We start with a presentation of the Hyperbrowser system, mentioning the background and usages of the system, while focusing on issues related to efficiency and Python. We expect the presentation, including questions, to take about an hour. We then compile a list of generic topics related to this that we have a shared interest in, and continue by discussing these issues for about an hour. Finally, we discuss possibilities for future cooperation.

Abstract:

DNA is basically a sequence of molecules that can be represented as a three billion long sequence of the letters A,C,G and T. Of course it is also much more. The DNA is a double helix with physical characteristics varying from position to position, and the DNA hosts a range of functional elements such as genes. Thousands of such annotations along the DNA is available, and a current challenge is to increase understanding of these annotations and their interactions.

Researchers at (sfi)² have developed a web-based system that allows biologist to perform complex analyses on selected annotation data. The main characteristics of this system are its simplicity of use, the system intelligence, and the maintainability, robustness and extendability inherent in the system architecture. We need a high-level language like Python for its flexibility and efficient development, while we need computational efficiency in order to support large analyses in near real time.

Our basic approach to efficiency is a flexible data representation, which allows each annotation type to be represented efficiently, and a flexible architecture of analysis classes that splits large computations and memoizes intermediate results for later use. Still, we also need a low run time factor per operation to reach our desired efficiency. To achieve this, we use numpy-vectors to represent chunks of annotation data, and rely on (mainly standard) numpy-operations at the core of analysis computations. Our system thus combines efficient operations on numpy-objects for the bulk of computations, with a flexible system in pure python for handling the complex application logic involving such numpy-objects.

While the combination of standard numpy vector-operations and pure python application logic in general allows for both rapid development and execution, we also face some challenges. One challenge is that certain analyses are difficult to express as a series of standard vector-operations. As our system should consist of hundreds of different analyses, we don't want too many individual analyses to require the development of custom c-extensions. Another challenge is that in certain situations the running time is dominated by the application logic, instead of numpy operations, meaning that the performance is effectively at a pure python level.

What	▪ Workshop
When	Sep 23, 2009 from 01:00 PM to 03:00 PM
Where	Hjørnehiet @ Simula
Contact Name	Hans Petter Langtangen
Attendees	Xing Cai Ola Skavhaug Hans Petter Langtangen Hans E. Plesser (UMB) Sveinung Gundersen (URR) Øyvind I Øvergaard (UiO) Geir Kjetil Sandve (UiO)
Add event to calendar	 vCal  iCal