

Bioinformatics Databases: State of the Art and Research Perspectives

François Bry, Peer Kröger

Institute for Computer Science

University of Munich, Germany

{bry,kroegerp}@informatik.uni-muenchen.de

Abstract

Bioinformatics or computational biology, i.e. the application of mathematical and computer science methods to solving problems in molecular biology that require large scale data, computation, and analysis, is a research area currently receiving a considerable attention. Databases play an essential role in molecular biology and consequently in bioinformatics. molecular biology data are often relatively cheap to produce, leading to a proliferation of databases: the number of bioinformatics databases accessible worldwide probably lies between 500 and 1.000. Not only molecular biology data, but also molecular biology literature and literature references are stored in databases. Bioinformatics databases are often very large (e.g. the sequence database GenBank contains more than 4×10^6 nucleotide sequences) and in general grows rapidly (e.g. about 8000 abstracts are added every month to the literature database PubMed). Bioinformatics databases are heterogeneous in their data, in their data modeling paradigms, in their management systems, and in the data analysis tools they supports. Furthermore, bioinformatics databases are often implemented, queried, updated, and managed using methods rarely applied for other databases. This presentation aims at introducing in current bioinformatics databases, stressing their aspects departing from conventional databases. A more detailed survey can be found in [1] upon which this presentation is built on.

References

- [1] François Bry and Peer Kröger. *A Computational Biology Database Digest: Data, Data Analysis, and Data Management*. Distributed and Parallel Databases, Vol. 13, pp. 7-42, Kluwer Academic Press, 2003.