



# BCB Lecture

**Zeit:** Montag, 7. März 2005, 16:00 Uhr

**Place:** Humboldt Universität zu Berlin, Institut für Informatik  
(Humboldt-Kabinett), Rudower Chaussee 25, 12489 Berlin

Gabriele Witterstein

*Zentrum für Mathematik, Technische Universität München*

## **Konstruktionsalgorithmus einer persistenten Indexstruktur für Genome**

Infolge weltweiter Genom-Sequenzierungsprojekte entstehen grosse Mengen an DNA-Rohdaten, welche in Datenbanken gehalten werden. Eine grundlegende Fragestellung ist deren textuelle Analyse innerhalb dieser Datenbanken. Es wird eine Variante eines persistenten Suffix Trees vorgestellt, welche sich platzsparend auf die Festplatte abbilden lässt, bei möglichst effizientem Zugriff. Die hauptsächliche wissenschaftliche Fragestellung ist die Konstruktion einer solchen persistenten Datenstruktur unter einem 'Two-Level-Memory Model' mit geringem Hauptspeicher, das heisst, die Überwindung des 'Memory Bottleneck'. Es wird gezeigt, dass bisher bekannte Ansätze zur Lösung dieses Problems für sehr lange Sequenzen nur eingeschränkt anwendbar sind. Des weiteren wird ein neuer Konstruktionsalgorithmus vorgestellt, welcher auf der Partitionierung der Datenmenge basiert und eine durchschnittliche Komplexität von der Ordnung  $O(n \log n)$  hat. Testläufe mit realen Sequenzen belegen die Anwendbarkeit des Algorithmus in der Bioinformatik. Vergleiche mit ähnlichen Algorithmen auf diesem Gebiet dokumentieren einen Performancegewinn.

**Gäste sind herzlich willkommen!**

Weitere Informationen unter <http://www.bcbio.de>