

# Visualisierung als Gegenstand der Ausbildung von Bioinformatikern

Ralf Dörner, Christian Seiler, Jens Barthelmes

Detlef Krömker

Fraunhofer AGC  
Varrentrappstraße 40-42  
60486 Frankfurt am Main  
{doerner,cseiler,barthelmes}@agc.fhg.de

Goethe-Universität Frankfurt a.M.  
Varrentrappstraße 40-42  
60486 Frankfurt am Main  
kroemker@cs.uni-frankfurt.de

**Abstract:** Die Visualisierung ist in der Ausbildung von Bioinformatikern häufig unterrepräsentiert, obwohl das Thema durch die zunehmende Datenflut und den hohen Komplexitätsgrad in der Biologie immer mehr an Bedeutung gewinnt. An der Johann Wolfgang Goethe-Universität Frankfurt am Main wurde im Wintersemester 02/03 erstmals eine kombinierte Lehrveranstaltung (Vorlesung, Übung und Praktikum) zur Visualisierung durchgeführt für Studenten der Informatik und Bioinformatik im Hauptstudium. Anhand von Beispielen zeigen wir, was bei der Konzeption und der Durchführung einer solchen Lehrveranstaltung zu beachten ist und stellen unsere Erfahrungen dar.

## 1. Konzeption

Die Ausbildung in der Visualisierung soll auf der einen Seite den theoretischen Hintergrund (z.B. über die menschliche Wahrnehmung, Farblehre, etc.) vermitteln. Auf der anderen Seite soll der Bioinformatiker in der Lage sein, vorhandene APIs und Visualisierungssysteme (z.B. das Visualization Toolkit) zu nutzen und erlernte Methoden auf konkrete Fragestellungen anzuwenden. Da die Güte einer Visualisierung immer sehr stark von den Anforderungen des Anwenders abhängt, ist die Wichtigkeit von Usability-Tests besonders hervorzuheben.

Analysiert man die genannten Lernziele, ist die Veranstaltung eines Praktikums unumgänglich. Unser Konzept für das Praktikum besteht darin, in einer kleineren initialen Aufgabe (ca. 20% der Praktikumszeit), Grundkenntnisse einzuüben in der Erstellung von einfachen computergraphischen Anwendungen inklusive des Aufbaus einer geeigneten Arbeitsumgebung ebenso wie in der Benutzung von Visualisierungssystemen. Als Beispiel einer derartigen Aufgabe sei die Visualisierung eines zweidimensionalen Vektorfeldes mittels Partikelsimulation genannt. Diese initiale Aufgabe dient auch der Homogenisierung der Kenntnisse der Lerngruppe – nach unserer Beobachtung sind gerade Programmierfertigkeiten häufig sehr unterschiedlich entwickelt. Die initiale Aufgabe ist so zu konzipieren, dass sie dem Lerner eine direkte graphische Rückkopplung seiner Aktivitäten gibt. Nach der initialen Aufgabe werden die

Ergebnisse der Bearbeitung verglichen, was insbesondere jedem Lerner die Möglichkeit der Selbsteinschätzung seiner Fertigkeiten gibt. Die sich daran anschließende weiterführende Aufgabe besteht in der Lösung eines realen Visualisierungsproblems aus der Bioinformatik. Ein integraler Bestandteil ist die Durchführung einer Usability Studie – zu deren Durchführung werden dem Lerner zeitnah Lernmaterialien zur Verfügung gestellt. Im Gegensatz zur initialen Aufgabe werden unterschiedliche Aufgabenstellungen an die Lerner verteilt. In einem abschließenden Workshop stellen die Lerner ihre Visualisierungslösungen vor und berichten über ihre Erfahrung bei der Umsetzung als auch bei der Durchführung der Usability Studie. Neben dem Kennenlernen von Fallbeispielen des breiten Spektrums von Visualisierungsaufgaben und –lösungen sollen die Lerner auch ihre Fähigkeiten im Peer-Review und die dazu benötigte Anwendung der theoretisch vermittelten Kriterien einüben. Das Praktikum wurde als Gruppenaufgabe konzipiert. Hierzu wurden die Studenten in Gruppen von 2 bis 4 Lernenden gefasst, die die Aufgaben im folgenden gemeinsam bearbeiteten. Die Zeiteinteilung lag in der Verantwortung der Studenten, wobei darauf geachtet wurde, dass während des Tages immer ein Betreuer als Ansprechpartner zu Verfügung stand. Fundiert wurde das Praktikum durch die Veranstaltung einer Vorlesung mit angeschlossener Übung. Die Vorlesung haben wir als wöchentlichen zweistündigen Fachvortrag organisiert. Den Studenten wurden drei Lehrbücher zur Verfügung gestellt [SM00, Wa00, Sp00].

## 2. Beispielaufgaben des Praktikums

**Proteinnetzwerk:** In dieser Aufgabe wurden Protein-Interaktionsdaten aus DIP (Database of Interacting Proteins, ca. 7 000 Proteine und 15 000 Interaktionen) sowohl als ungerichteter Graph als auch unter Anwendung der Hyperbolic Viewer-Technik interaktiv visualisiert. Zusätzlich werden – soweit vorhanden – Annotationsdaten zu den Proteinen eingeblendet. Dem Benutzer wurde eine möglichst leichte Navigation durch die Daten ermöglicht. Am Ende wurde ein Usability-Test durchgeführt, um zu prüfen, ob die Darstellung im Hyperbolic Viewer Vorteile brachte.

**Visueller Fingerabdruck von Proteinen:** Diese Aufgabe bestand darin, eine kompakte aber gleichzeitig intuitive visuelle Repräsentation (Glyphen) für Proteindaten (Molekulargewicht, Klasse, Unterklasse(n), Lokalisation, Bindungspartner, Modifikationen etc.) zu entwickeln. Ziel war es, durch diese Visualisierung einerseits eine bessere Wiedererkennung der Proteine (z.B. in Pathway-Darstellungen) zu ermöglichen und andererseits mehrere Eigenschaften in einer kompakten Darstellung z.B. zum schnellen Vergleich mehrerer Proteine zu vereinen. Die Qualität der Visualisierung wurde anschließend in einem Usability-Test verifiziert.

**Cluster-Visualisierung:** In dieser weiterführenden Aufgabe wurden drei Datenbasen aus der Biologie mit unterschiedlicher Anzahl von Datensätzen (zwischen 20 und 200.000) vorgegeben, als auch drei Clusteringverfahren. Neben zwei konventionellen Verfahren wurde das ASRA-Verfahren [DW02] gewählt. Die Grundidee des Verfahrens besteht darin, dass jeder Datensatz durch einen Avatar repräsentiert wird und sich diese Avatare anhand eines Bewegungsmodells in einer abgegrenzten Fläche bewegen:

Avatare, die ähnliche Datensätze repräsentieren, bewegen sich aufeinander zu bzw. unähnliche stossen einander ab. Der Benutzer kann interaktiv das Bewegungsmodell verändern und mit den Avataren selbst interagieren. Die Identifizierung von Clustern erfolgt also explorativ durch visuelle Inspektion der Avatare und ihres Verhaltens in Reaktion auf Benutzerinteraktionen. Die Datensätze wurden auf Cluster untersucht und visualisiert. Darauf aufbauend wurden die Verfahren mittels Anwendertest bewertet.

### 3. Ergebnisse

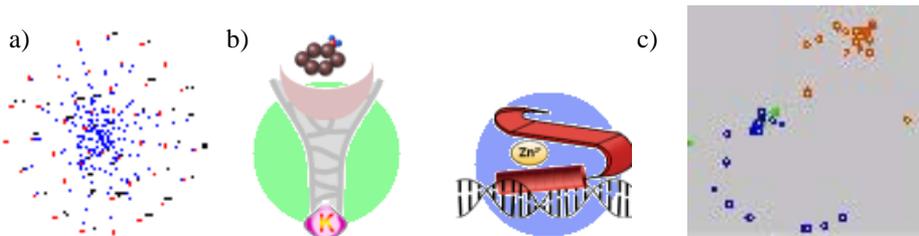


Abb.1: Aufgaben a) „Proteinnetzwerk“ b) „Protein-Fingerabdruck“ c) „Clustering“

Abbildung 1 zeigt exemplarisch Ergebnisse aus den studentischen Arbeiten. Eine Umfrage bei den 18 Teilnehmern am Ende der Veranstaltung ergab, dass die Herausforderung vor allem in der Programmiermethodik und –technik lag, während der Schwierigkeitsgrad als angemessen bewertet wurde. Dies lag aufgrund der meist ungenügenden Vorerfahrungen der Lernenden mit der Programmierung innerhalb der Erwartung. Auch der Arbeitsaufwand lag mit Durchschnittlich etwa 15 Arbeitsstunden im erwarteten Bereich. Allein die Bewertung der Teamarbeit schnitt in der Bewertung der Lernenden leicht unterdurchschnittlich ab. Aus unserer Sicht aber gelang ein guter Austausch zwischen den verschiedenen Gruppen. Das Prinzip der Aufgabenteilung im Praktikum in eine initiale und eine weiterführende Aufgabe hat sich aus unserer Sicht sehr bewährt. Nicht nur wurden so technische Probleme bereits in einer frühen Phase der Veranstaltung behandelt und so dass es zu einer geringeren Belastung der Endphase kam, sondern es wurden auch grundlegende Programmierfertigkeiten eingeübt, die für die weitere Bearbeitung unerlässlich waren. Betrachtet man den gesamten Komplex aus Vorlesung, Übung und Praktikum, so erkennt man, dass Lehrveranstaltung zur Visualisierung in der Bioinformatik weder Anhängsel an „klassische“ Bioinformatikthemen wie etwa Datenbanksysteme sind, noch auf Vorlesungen zur allgemeinen Computergraphik aufbauen.

- [SM00] Schumann, H., Müller, W.: Visualisierung. Grundlagen und allgemeine Methoden. Springer Verlag, Berlin, 2000
- [Wa00] Ware, C.: Information Visualization. Perception for Design. Morgan Kaufmann Publishers, 2000.
- [Sp00] Spence, R.: Information Visualization. Addison-Wesley, 2000.
- [DW02] Dörner, R., Ware, C.: Interactive techniques for exploring data using behavioral animation. Proceedings of The Eighth International Conference on Distributed Multimedia Systems 2002. Skokie, IL : Knowledge Systems Institute, 2002. S. S.622-629.