



Wissensmanagement in der Bioinformatik

Exposé zur Studienarbeit „Visualisierung GRIPP-indexierter Graphen“

Betreuer: Silke Trißl, Prof. Ulf Leser

Johannes Vogt

vogt@informatik.hu-berlin.de

Zeitraum: 01.04.2006 – 01.07.2006

1. Motivation

In der Bioinformatik und der Biologie werden verschiedenste Strukturen in Form von teilweise sehr großen Bäumen oder Graphen dargestellt. Um Anfragen an diese Graphen effizient durchführen zu können, werden Indizierungsmethoden benötigt, die sowohl speicherplatzeffizient sollen als auch schnell berechenbar sein müssen. Eine dieser Indizierungsmethoden ist GRIPP (graph indexing based on pre-/postorder numbering). GRIPP ist eine Erweiterung der pre-/postorder Indizierung von Bäumen auf allgemeine Graphen. Um die Entwicklung von GRIPP zu unterstützen und seine Funktionsweise zu verdeutlichen, wird ein Werkzeug benötigt, das die Indexstruktur des Graphen oder Teilgraphen visualisiert, die Suchanfragen entgegennimmt, und den Ablauf der Suche darstellt.

2. Zielsetzung

Ziel der Studienarbeit ist die Implementierung eines Java-Programms, das einerseits Graphen und deren GRIPP-Indexbäume visualisiert, und andererseits die Bearbeitung von Anfragen auf diesen Graphen schrittweise zeigt.

3. Hintergrund

Bei GRIPP – Indexbäumen handelt es sich um Bäume mit zwei unterschiedlichen Arten von Knoten, Baum-Instanzen und Nichtbaum-

Instanzen. Die Nichtbaum-Instanzen, die immer Blattknoten des Indexbaums sind, dienen als Verweise zu anderen Knoten. Hierdurch ist es möglich, allgemeine (auch zyklische) Graphen auf GRIPP-Indexbäumen abzubilden. Sowohl die Graphen, d.h. die Knoten und Kanten, als auch die GRIPP Indexstrukturen sind in einem relationalen Datenbanksystem gespeichert. Stop-Knoten sind Wurzelknoten von in sich abgeschlossenen Teilbäumen, das heisst keine Nichtbaum-Instanzen von ausserhalb des Teilbaumes zeigen in den Teilbaum und umgekehrt. In der Visualisierung, sollen die Stop-Knoten farblich hervorgehoben sein. Wenn der Indexbaum bei einer Suchanfrage durchlaufen wird, werden sogenannte „hops“ durchgeführt. Bei einem hop wird von einer Nichtbauminstanz zu dem Teilbaum der entsprechenden Bauminstanz gesprungen, und in diesem weitergesucht. Durch die skip-Strategie werden bei der Suche nur Teilbäume betrachtet, die keine schon geprüften Teilbäume enthalten. Sowohl die hop-, als auch die skip-Ereignisse sollen visuell nachvollzogen werden können. Im Fall der hops, indem der Zielteilbaum farblich markiert wird, und im Fall der skips, indem nur der tatsächlich durchsuchte Teilbaum stark markiert wird, der zunächst weggelassene Teilbaum dagegen schwach.

4. Vorgehensweise

Als erstes erstelle ich mit Visualisierungsbibliothek *prefuse* ein Modul zur Visualisierung von nichtindizierten Graphen. Im zweiten Schritt erweitere ich dieses Modul um die Möglichkeit GRIPP-indizierte Graphen als Indexbaum einzulesen. Bei der Visualisierung dieser Indexbäume werden die verschiedenen Knotentypen farblich unterschiedlich dargestellt. Ausserdem wird beim Anklicken eines Knotens dessen Bauminstanz fokussiert, und alle zugehörigen Nichtbaum-Instanzen und der Teilbaum der Bauminstanz farblich markiert. Die Graphdaten kommen jeweils aus einer Oracle-Datenbank. Über einen „Einlesen“-Dialog kann die Datenbank und die Tabellen, aus denen gelesen werden soll, ausgewählt werden. Anschliessend füge ich Eingabemöglichkeiten für Reachability-Anfragen hinzu und realisiere deren Funktionalität. Diese Suchanfragen protokollieren den gegangenen Suchpfad in einer Tabelle, die Schrittweise vom Nutzer in der Visualisierung nachvollzogen werden kann. Hierfür wird ein Menu zur Steuerung benötigt, sowie ein Modul, das den aktuellen Suchzustand definiert, wonach entschieden wird, welche Knoten graphisch markiert werden müssen, und welcher Teilbaum gerade angezeigt werden soll.

Literatur

Technical Report, in preparation.