

Seminar Theoretische Informatik

Analyse von Algorithmen

Wir wollen uns in einem zweiteiligen Blockseminar mit der *Analyse von Algorithmen* beschäftigen. Dabei wollen wir uns in erster Linie mit der durchschnittlichen Laufzeit auseinandersetzen, was bedeutet, dass wir über alle möglichen Inputdaten mitteln. Da die Inputdaten meist eine kombinatorische Struktur aufweisen (etwa Bäume oder Permutationen), bietet das Seminar auch Einblick in die Eigenschaften dieser Objekte. Hauptwerkzeug werden dabei *Erzeugende Funktionen* (engl.: *generating functions*) sein, die einen Zusammenhang zwischen der diskreten Mathematik und der Analysis herstellen.

Die Hauptliteratur ist dabei das Buch *An Introduction to the Analysis of Algorithms* [SF96] von R. Sedgwick und P. Flajolet. Sekundär werden wir *Analytic Combinatorics* [FS09] derselben Autoren verwenden (im Netz frei verfügbar), außerdem *Average Case Analysis of Algorithms on Sequences* von W. Szpankowski [Szp01] sowie die Buchserie *The Art of Computer Programming* von Knuth.

Im ersten Block wird eine Einführung in die nötigen mathematischen Grundlagen gegeben. Dabei planen wir etwa 4 – 6 Vorlesungen am Anfang des Semesters (Anwesenheit wird erwartet). Bei Wunsch können auch im ersten Block Vorträge gehalten werden.

Im zweiten Block, wahrscheinlich am Anfang der Semesterferien, sollen die Studenten ein Thema vorstellen.

Zielgruppe. Das Seminar richtet sich an StudentInnen der Informatik und Mathematik im Bachelor- und Masterstudium. Das Seminar wird mit fünf Punkten bewertet. Die erwartete Vortragslänge beträgt ca. 60 Minuten.

Voraussetzungen. Die mathematischen Grundlagen werden im ersten Block geboten, somit sind kaum Voraussetzungen nötig. Lediglich komplexe Zahlen und die Landau-Notation (auch "groß-O Notation") sollten Ihnen vertraut sein. Hilfreich sind Vorlesungen der Kombinatorik und der Analysis.

Themen. Vorlesungen: Mathematische Grundlagen.

1. Einführung und Rekurrenzen

2. Erzeugende Funktionen
3. Asymptotische Approximation
4. Analytische Kombinatorik

Wenn gewünscht, können hier auch Themen übernommen werden.
Vorträge:

1. Themenkomplex *Bäume* (Kapitel 6 in [SF96])
 - (a) *Bäume, Catalan-Zahlen und Binäre Suchbäume*: Es sollen die Grundeigenschaften von Bäumen vorgestellt werden. Des Weiteren sollen noch einige andere Interpretationen der Catalanzahlen und Eigenschaften binärer Suchbäume behandelt werden.
 - (b) *Die durchschnittliche Höhe von Bäumen*: Dieses Thema ist mathematisch anspruchsvoll. Hier soll die durchschnittliche Höhe von geordneten Bäumen vorgerechnet werden. Literatur hierzu ist das klassische Paper [DBRK71].
 - (c) *Anderer Baumklassen*: Hier sollen kombinatorische Eigenschaften von ungewurzelten, gelabelten und anderer Baumklassen vorgestellt werden.

2. Themenkomplex *Permutationen*

Uniform verteilte Permutationen sind für viele Algorithmen ein natürliches Inputmodell. Die Laufzeit eines Algorithmus hängt dann beispielsweise von der durchschnittlichen Anzahl der Zyklen in einer Permutation ab. Literatur für beide Vorträge ist [SF96, Kapitel 7].

- (a) *Grundeigenschaften von Permutationen*: In diesem Vortrag sollen verschiedene Parameter vorgestellt werden, die für Permutationen wichtig sind.
- (b) *Average-Case Verhalten diverser Suchalgorithmen*: Aufbauend auf den vorherigen Vortrag, soll die durchschnittliche Komplexität von *Insertion Sort*, *Shellsort*, und *Selection Sort* vorgestellt werden.

3. Themenkomplex *Strings*:

In diesem Themenkomplex sollen Algorithmen über Strings, also Sequenzen von Symbolen eines festen Alphabets, untersucht werden. Zentral ist die Suche nach einem String u in einem String v .

- (a) *Grundeigenschaften von Strings*: Es sollen verschiedene simple Fragen beantwortet werden, beispielsweise wie oft ein String der Länge p in einem String der Länge n durchschnittlich vorkommt, oder wie lang die längste Sequenz von 0 in einem Bitstring durchschnittlich ist.
- (b) *Der Knuth-Morris-Pratt Algorithmus*: Der KMP-Algorithmus [KMP77] ist ein bekannter Algorithmus, der Vorkommen eines Strings u in v sucht. In diesem Vortrag soll der Algorithmus vorgestellt und analysiert werden.
- (c) *Tries und LZW-Kompression*: *Tries* sind eine beliebte Datenstruktur zur Indexierung von Strings. Es sollen verschiedene *Tries* vorgestellt und analysiert werden.

4. Themenkomplex *Abbildungen*:

- (a) *Das Geburtstagsparadoxon & das Sammelbilderproblem*: Hier sollen zwei bekannte Probleme der Kombinatorik vorgestellt werden.
- (b) *Hashing-Algorithmen*: Aufbauend auf den vorherigen Vortrag sollen verschiedene Hashing-Algorithmen vorgestellt und analysiert werden.
- (c) *Zufällige Abbildungen und Faktorisieren via Pollard-Rho*: Zur Zufallszahlengenerierung könnten prinzipiell zufällige Methoden herangezogen werden. Hier soll dargestellt werden, wo dabei Probleme auftauchen können (Stichwort: Erwartete Zyklenlänge). Aufbauend darauf soll die Laufzeit eines Faktorisierungsalgorithmus analysiert werden.

Es wäre wünschenswert, wenn sich Gruppen für die einzelnen Themenkomplexe bilden könnten (beispielsweise gemeinsame Vorträge für den Themenkomplex *Permutationen*).

Andere Themen sind möglich. Interessenten melden sich bitte unter eric.noeth@eti.uni-siegen.de.

Literatur

- [DBRK71] N.G. De Bruijn, S.O. Rice, and Donald Ervin Knuth. Average height of planted plane trees. Technical Report STAN-CS-71-218, Stanford University (Stanford, CA US), 1971.

- [FS09] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [KMP77] Donald E. Knuth, J.H. Morris, and Vaughan R. Pratt. Fast pattern matching in strings. *SIAM Journal of Computing*, 6(2):323–350, 1977.
- [SF96] Robert Sedgewick and Philippe Flajolet. *An introduction to the analysis of algorithms*. Addison-Wesley-Longman, 1996.
- [Szp01] Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences*. John Wiley & Sons, Inc., New York, NY, USA, 2001.