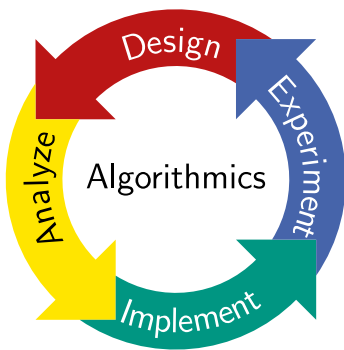


Algorithm Engineering

Peter Sanders · Dorothea Wagner



Einleitung

Effiziente Algorithmen und Datenstrukturen sind eine Grundvoraussetzung für anspruchsvolle Computeranwendungen. Algorithmik – die systematische Entwicklung effizienter Algorithmen – ist deshalb in etlichen Bereichen entscheidend für die Umsetzung technologischer Möglichkeiten und somit von großer Bedeutung für Technik, Wirtschaft, Wissenschaft und unser tägliches Leben.

Traditionell hat sich die Algorithmik der Methodik der *Algorithmentheorie* bedient, die aus der Mathematik stammt: Algorithmen werden für einfache und abstrakte Problem- und Maschinenmodelle entworfen. Hauptergebnis sind dann beweisbare Leistungsgarantien für beliebige Eingaben. Dieser Ansatz führt in vielen Fällen zu eleganten, zeitlosen Lösungen, die sich an viele Anwendungen anpassen lassen. Die harten Leistungsgarantien ergeben zuverlässig eine hohe Effizienz und gute Lösungsqualität auch für zur Implementierungszeit unbekannte Typen von Eingaben. Auswahl und Implementation eines Algorithmus ist aus Sicht der Algorithmentheorie Teil der Anwendungsentwick-

lung. Nach allgemeiner Beobachtung ist diese Art des Ergebnistransfers aber ein sehr langsamer Prozess. Bei wachsenden Anforderungen an innovative Algorithmen ergeben sich daraus immer größere Lücken zwischen Theorie und Praxis: Reale Hardware entwickelt sich durch Parallelismus, Pipelining, Speicherhierarchien u.s.w. immer weiter weg von einfachen Maschinenmodellen. Anwendungen werden immer komplexer. Gleichzeitig entwickelt die Algorithmentheorie immer ausgeklügeltere Algorithmen, die zwar wichtige Ideen enthalten, aber oft kaum implementierbar sind. Außerdem haben reale Eingaben oft wenig mit den Worst-Case-Szenarien der theoretischen Analyse zu tun. Im Extremfall werden vielversprechende algorithmische Ansätze vernachlässigt, weil eine vollständige Analyse mathematisch zu schwierig wäre.

Algorithm Engineering ist eine Methodik für die Entwicklung und Erforschung von Algorithmen, die dazu beiträgt, die oben beschriebenen Lücken zwischen Theorie und Praxis zu überwinden. Algorithm Engineering vereint theoretische und experimentelle Ansätze und erlaubt eine enge Kopplung an Anwendungen. Kern des Algorithm Engineering ist ein Kreislauf aus Entwurf,

DOI 10.1007/s00287-013-0684-1
© Springer-Verlag Berlin Heidelberg 2013

Peter Sanders · Dorothea Wagner
Karlsruher Institut für Technologie,
Karlsruhe
E-Mail: {sanders, dorothea.wagner}@kit.edu

*Vorschläge an Prof. Dr. Frank Puppe
<puppe@informatik.uni-wuerzburg.de> oder
Prof. Dr. Dieter Steinbauer <dieter.steinbauer@schufa.de>

Alle „Aktuellen Schlagwörter“ seit 1988 finden Sie unter:
www.ai-wuerzburg.de/as

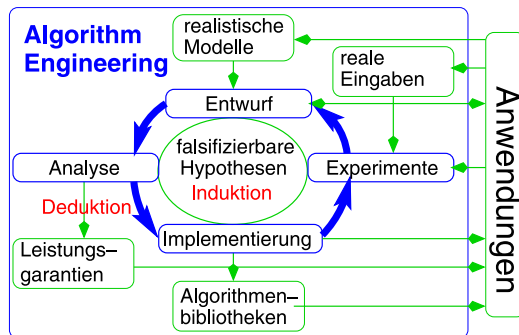


Abb. 1 Algorithm Engineering als Kreislauf aus Entwurf, Analyse, Implementierung und experimenteller Analyse

Analyse, Implementierung und experimenteller Bewertung von Algorithmen. Hinzu kommt eine stärkere Einbindung von Anwendungen und realistischen Maschinenmodellen. Implementierungen können direkt oder indirekt (über wiederverwendbare Algorithmenbibliotheken) in Anwendungen einfließen. Umgekehrt liefern Anwendungen verfeinerte Modelle für den Algorithmenentwurf sowie Szenarien für realistische Experimente. Abbildung 1 zeigt ein Schema dieses Ansatzes. Die einzelnen Aspekte werden in den folgenden Abschnitten diskutiert.

Algorithm-Engineering-Kreislauf

Realistische Modelle

Basis jedes systematischen Algorithmenentwurfs sind mathematische Modelle, die das zu lösende Problem und die ausführende Maschine beschreiben. Entscheidend ist hier die richtige Balance zwischen Einfachheit und Genauigkeit. Besonders deutlich wird das bei den Maschinenmodellen. Während die Algorithmentheorie weitgehend mit dem von-Neumann-Modell arbeitet, bei dem ein einfach aufgebauter Prozessor wahlfreien Zugriff auf einen homogen strukturierten Speicher hat, bestehen moderne Computer aus vielen Prozessorkernen und Speichermodulen (Caches, Hauptspeicher, Platten, ...), die ein komplexes hierarchisches Netzwerk bilden. Selbst die einzelnen Kerne haben durch verschiedene Formen von Befehlsparallelismus (Pipelining, Superskalarität, Vektorbefehle, spekulative Ausführung, ...) ein sehr komplexes Verhalten. Eine genaue Modellierung der Ausführungszeit eines Programms ist deshalb kaum möglich und wäre auch nicht zielführend, weil das Ergebnis we-

der verständlich noch auf andere Architekturen übertragbar wäre. Das Algorithm Engineering setzt deshalb auf Abstraktionen, die einzelne besonders wichtige Aspekte des Maschinenmodells erfassen. Erfolgreich ist zum Beispiel das I/O-Modell, das sich auf zwei Speicherhierarchieebenen beschränkt – einen schnellen Speicher mit beschränkter Kapazität M und einen großen Speicher, auf den in Blöcken der Größe B zugegriffen wird [6]. Ein nützliches Modell für Parallelverarbeitung betrachtet p gleichartige Rechnerknoten, die über ein Netzwerk verbunden sind. Knoten interagieren durch Nachrichtenaustausch. Jeder Prozessor kann zu einem Zeitpunkt nur eine Nachricht senden und/oder empfangen und der Zeitbedarf für eine Nachricht der Länge n ist $\alpha n + \beta$ für geeignete Parameter α und β . Dieses Modell passt sehr gut zu Cluster-Computern und führt auch bei Many-Core-Rechnern mit gemeinsamem (Non-Uniform Memory Access) Speicher zu effizienten Algorithmen.

Das Abstraktionsniveau für Anwendungsprobleme ist naturgemäß sehr problemabhängig. Bei einem relationalen Datenbanksystem können wir zum Beispiel sehr genau spezifizieren, was das Ergebnis einer SQL-Anfrage sein soll. Bei einem Routenplaner für den Straßenverkehr ist das richtige Modell dagegen viel weniger klar. Eine einfache Abstraktion ist die Modellierung des Straßennetzes durch einen Graphen, bei dem Kanten Straßenabschnitte darstellen und Kantengewichte die Fahrzeit oder ein anderes Kostenmaß angeben [3]. Die genaue Modellierung fortgeschrittener Aspekte wie tageszeitabhängige Verkehrsdichte, Stauumfahrungen oder Ampelwartezeiten ist dagegen extrem schwierig und scheitert schon an der Verfügbarkeit geeigneter Daten. Trotzdem können wir hoffen, durch geeignete Modelle verbesserte Ergebnisse zu erzielen. Zum Beispiel lassen sich schnelle Routenplanungstechniken für das einfache Modell auf ein zeitabhängiges Modell mit stückweise linearen Fahrzeitfunktionen übertragen oder an komplexere Metriken als die Fahrzeit anpassen.

Entwurf

Beim Algorithmenentwurf sucht die Algorithmentheorie meistens nach asymptotisch effizienten Algorithmen für den schlechtesten Fall. Im Algorithm Engineering interessieren wir uns zusätzlich für Implementierbarkeit. Die Algorithmen sollten

also einfach sein und wenn möglich bereits verfügbare Komponenten wiederverwenden. Bei der Effizienz und Lösungsqualität interessiert uns nicht nur der schlechtesten Fall, sondern mehr noch das Verhalten für „typische“ Eingaben, das deutlich anders sein kann. Natürlich sind konstante Faktoren in der Ausführungszeit keineswegs zu vernachlässigen. Zum Beispiel ist unser Algorithmus zur Bestimmung minimaler Spannbäume im I/O-Modell theoretisch um einen logarithmischen Faktor schlechter als das beste bekannte Verfahren. Für realistische Hardware wird aus diesem Faktor aber eine Konstante, und verglichen zu den theoretisch besten Verfahren ergibt sich eine mindestens dreimal schnellere und sehr einfach zu implementierende Methode [4].

Analyse

Die Algorithmenanalyse bleibt im Algorithm Engineering wichtig, weil sie in kompakter Form Voraussagen über das Verhalten von Algorithmen in einer Vielzahl von Situationen erlaubt. Hier entstehen auch neue theoretische Herausforderungen, etwa bei der Entwicklung von Analysetechniken jenseits der Worst-Case-Analyse. Smoothed Analysis [7] oder die Betrachtung parametrisierter Komplexität [5] liefern neue Leistungsgarantien von Algorithmen, die schon lange im praktischen Gebrauch sind, sich aber einer einfachen Analyse entziehen.

Implementierung

Die effiziente Implementierung von Algorithmen ist eine große Herausforderung, weil es große semantische Lücken zwischen der abstrakten Beschreibung eines Algorithmus, höheren Programmiersprachen und der ausführenden Hardware gibt. Diese Lücken werden durch die immer komplexere Hardware eher größer als kleiner.

Experimente

Aussagekräftige Experimente sind der Schlüssel zum Schließen des Algorithm-Engineering-Entwicklungszyklus und erfordern eine sorgfältige Planung, oft hohen Rechenaufwand und eine noch sorgfältigere Interpretation der Ergebnisse. Die experimentelle Methodik kann von Poppers wissenschaftlicher Methode lernen: Experimente sind getrieben von falsifizierbaren Hypothesen über das Verhalten der untersuchten Algorithmen.

Die Hypothesen stammen aus Entwurf, Analyse, Implementierung oder aus vorhergehenden Experimenten. Die Ergebnisse bestätigen, widerlegen oder verfeinern die Hypothesen. Dies liefert dann Ideen für den Entwurf verbesserter Algorithmen, eine genauere Analyse oder effizientere Implementierung.

Wichtig ist die Reproduzierbarkeit von Experimenten. Das bedeutet, dass die involvierten Programme, Eingabedaten und sonstigen Ablaufparameter sorgfältig dokumentiert werden. Diese Informationen müssen nach allgemein anerkannten Richtlinien mindestens zehn Jahre aufgehoben werden. Wissenschaftler, die Experimente überprüfen oder unter veränderten Bedingungen wiederholen möchten, sollten Zugang zu diesen Informationen bekommen. Eine exakte Wiederholbarkeit von Experimenten nach mehreren Jahren käme allerdings dem exakten Wiederaufbau einer Versuchsanlage in den Natur- oder Lebenswissenschaften gleich, und ist oft nicht möglich, da die benutzten Rechner nicht mehr existieren und alte Softwareversionen auf neuer Hardware vielfach nicht mehr laufen. Deshalb ist es ratsam, nicht nur Laufzeiten zu messen, sondern auch weniger hardwareabhängige Größen. Zum Beispiel kann man bei einem Sortierverfahren Vergleichsoperationen zählen oder das I/O-Volumen eines Sekundärspeicheralgorithmus bestimmen.

Instanzen und Benchmarks

Sammlungen realistischer Eingaben für ein algorithmisches Problem sind entscheidend für Fortschritte beim Algorithm Engineering, weil sie einen einfachen und objektiven Vergleich verschiedener Ansätze erlauben. Zum Beispiel hat das Verfügbaren kontinentgroßer Straßengraphen um 2005 zu einer regelrechten Explosion der Leistung von Routenplanungsalgorithmen geführt [3]. Die Vereinbarung von Benchmark-Instanzen ist ein wichtiger Forschungsaspekt, zu dem die *DIMACS Implementation Challenges*¹ einen bedeutenden Beitrag leisten. So sind für die Graphpartitionierung und -clustering bei der *10th DIMACS Implementation Challenge - Graph Partitioning and Graph Clustering*² Standards für die experimentelle Evaluation gesetzt worden [1].

¹ <http://dimacs.rutgers.edu/Challenges/>

² <http://www.cc.gatech.edu/dimacs10/>

Obwohl synthetische Instanzen weniger realistisch sind, haben auch diese Vorteile. Da sie sich in beliebiger Größe erzeugen lassen, erleichtern sie Skalierbarkeitsexperimente und können als Stresstests dienen. Zudem bieten synthetische Instanzen bessere Handhabe, im größerem Umfang und unter genauerer Kontrolle relevanter Parameter Experimente durchzuführen.

Algorithmenbibliotheken

Algorithmenbibliotheken sind die Schnittstelle zwischen Algorithm Engineering und Software Engineering. Die Bibliotheken sollten effizient, leicht benutzbar, portabel und gut dokumentiert sein. Algorithmenbibliotheken erleichtern den Transfer von Know-how aus dem Algorithm Engineering in Anwendungen. Innerhalb des Algorithm Engineering vereinfachen sie den Vergleich zwischen Algorithmen und die Konstruktion von darauf aufbauender Funktionalität (siehe auch das Beispiel in Abschn. „Entwurf“). Softwareengineering für Algorithmenbibliotheken ist eine besondere Herausforderung, weil es oft einen natürlichen Konflikt zwischen einfacher Benutzbarkeit, allgemeiner Einsetzbarkeit einerseits und Effizienz der Implementierungen andererseits gibt. Dazu kommt, dass zur Erstellungszeit der Bibliothek gar nicht alle Anwendungen bekannt sind. Besonders hoch sind auch die Anforderungen an die Zuverlässigkeit, weil ein Benutzer einer Algorithmenbibliothek schnell entnervt aufgibt wenn eine Bibliothek – wirklich oder scheinbar – nicht funktioniert (bekanntlich schreiben viele Programmierer/Arbeitsgruppen ihre Codes lieber selbst).

All diese Schwierigkeiten machen deutlich, dass eine gute Algorithmenbibliothek viel schwieriger zu erstellen ist als die prototypischen Implementierungen, die sonst den Algorithm-Engineering-Zyklus vorantreiben. Nur ein Bruchteil der implementierten Verfahren wird also seinen Weg in Bibliotheken finden.

Zusammenfassung und Ausblick

Die Forschung im Algorithm Engineering hat in einigen Bereichen bereits erfolgreich dazu beigetragen, die Kluft zwischen Theorie und Praxis zu überbrücken. Die rasanten Fortschritte, die durch neue Algorithmen für die Routenplanung oder die Suche in Texten erzielt wurden, sind Paradebeispiele für das Algorithm Engineering. Besonders faszinierend sind Forschungsergebnisse, die auf dem vollständigen Kreislauf des Algorithm Engineering basieren und aus Experimenten und Anwendungen gewonnene theoretische Erkenntnisse wieder in den Entwurf neuer Algorithmen einbeziehen [2].

Mit der Weiterentwicklung der Hardware oder der zunehmenden Verfügbarkeit riesiger Datensätze und der wachsenden Bedeutung von Anwendungen, die auf riesigen Datensätzen basieren, stellen sich immer wieder neue Herausforderungen an die Forschung im Algorithm Engineering. Aber auch in der Lehre an Universitäten sollte heutzutage Algorithm Engineering in der Ausbildung von Informatikern einen festen Platz haben.

Danksagung

Teilweise unterstützt durch die DFG im Rahmen des Schwerpunktprogramms SPP 1307 Algorithm Engineering unter SA 933/4-3 und WA 654/15-3.

Literatur

1. Bader DA, Meyerhenke H, Sanders P, Schulz C, Schumm A, Wagner D (2012) A benchmarking set for graph clustering and partitioning. In: Rokne J, Alhajj R (eds) Encyclopedia of Social Network Analysis and Mining. To appear
2. Delling D, Goldberg AV and Werneck RF (2011) Shortest paths in road networks: from practice to theory and back. In: Sanders P, Wagner D (eds), it – Information Technology, Algorithm Engineering, vol 53, Oldenburg-Verlag, pp 294–301
3. Delling D, Sanders P, Schultes D, Wagner D (2009) Engineering route planning algorithms. In: Algorithmics of Large and Complex Networks, LNCS State-of-the-Art Survey, vol 5515, Springer, pp 117–139
4. Dementiev R, Sanders P, Schultes D, Sibeyn J (2004) Engineering an external memory minimum spanning tree algorithm. In: IFIP TCS, Toulouse, pp 195–208
5. Downey RG, Fellows MR (1999) Parameterized Complexity. Springer
6. Meyer U, Sanders P, Sibeyn J (eds) (2003) Algorithms for Memory Hierarchies. LNCS Tutorial, vol 2625, Springer
7. Spielman DA, Teng S (2001) Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In: Proceedings on 33rd Annual ACM Symposium on Theory of Computing (STOC), Heraklion, Crete, Greece, 6–8 July 2001, pp 296–305