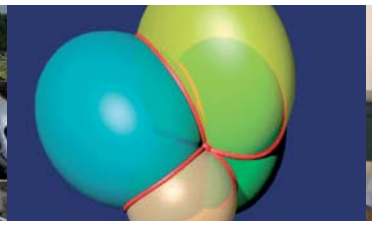


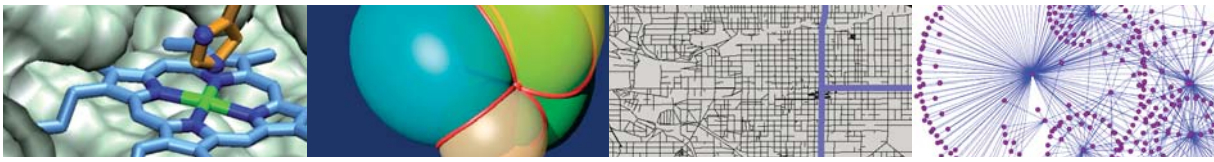


max planck institut
informatik

Bericht 2005/2006

05 06 07 08 09

max planck institut
informatik



Bericht 2005/2006

Direktorium

Prof. Dr. Gerhard Weikum
Prof. Dr. Kurt Mehlhorn
Prof. Dr. Thomas Lengauer, Ph.D.
Prof. Dr. Hans-Peter Seidel

Fachbeirat

Prof. Dr. Pankajj Kumar Agarwal, Duke University, USA
Prof. Dr. Douglas L. Brutlag, Stanford University, School of Medicine, USA
Prof. Dr. Joseph M. Hellerstein, University of California, Berkeley, USA
Prof. Dr. Yannis E. Ioannidis, University of Athens, Greece
Prof. Dr. Friedhelm Meyer auf der Heide, Heinz Nixdorf Institut, Universität Gesamthochschule Paderborn, Germany
Prof. Dr. Eugene Myers, Howard Hughes Medical Institute, USA
Prof. Dr. Frank Pfenning, Computer Science Department, Carnegie Mellon University, USA
Prof. Dr. Claude Puech, Laboratoire GRAVIR-INRIA, France
Prof. Dr. Éva Tardos, Cornell University, USA
Prof. Dr. Demetri Terzopoulos, Department of Computer Science, New York University, USA

Kuratorium

Dr. Hanspeter Georgi, Minister für Wirtschaft, Saarland
Peter Stefan Herbst, Chefredakteur Saarbrücker Zeitung
Prof. Dr. Joachim Hertel, Infor Business Solutions AG
Prof. Dr. Matthias Jarke, RWTH Aachen
Ministerialdirigent Dr. Wolf-Dieter Lukas, Bundesministerium für Bildung und Forschung
Fritz Raff, Intendant des Saarländischen Rundfunks
Dr. Hartmut Raffler, Siemens AG
Jürgen Schreier, Minister für Bildung, Kultur und Wissenschaft des Saarlandes
Prof. Dr. Wolffried Stucky, Institute of Applied Informatics and Formal Description Methods, Universität Karlsruhe
Dr. Richard Weber, Präsident IHK des Saarlandes
Prof. Dr. Margret Wintermantel, Präsidentin der Hochschulrektorenkonferenz
Prof. Dr. Volker Linneweber, Präsident der Universität des Saarlandes

INHALTE

VORWORT

VORWORT

Dies ist der zweite Jahresbericht, den wir, das Max-Planck-Institut für Informatik, einer breiteren Öffentlichkeit vorlegen. Wir wollen damit allen Wissenschaftsinteressierten Themen, Ziele und Methoden der modernen Informatik und die Arbeiten unseres Instituts vorstellen. Insbesondere hoffen wir, Ihnen, liebe Leser, die Faszination unserer Wissenschaft näherzubringen.

Unser erster Jahresbericht 2003/2004 liegt zwei Jahre zurück, und seitdem hat sich an unserem Institut und in unserem Umfeld einiges getan. Die inhaltlichen Neuerungen – brandaktuelle Forschungsthemen und unsere neuesten Resultate – finden Sie in den einzelnen Abschnitten des Berichts selbst. An diesen Themen werden wir auch in den nächsten beiden Jahren weiter forschen. Im letzten Teil des Berichts finden Sie eine Auswahl von wissenschaftlichen Publikationen für die Jahre 2005/2006 sowie einige Kennzahlen unseres Instituts für diesen Zeitraum.

Organisatorisch spiegelt sich der ständige Wandel unserer Disziplin in der Einrichtung zweier Forschungsgruppen wider, einer Gruppe über die Automatisierung von Logik und einer Gruppe über Maschinelles Lernen. Im externen Umfeld unseres Instituts ist die Gründung des neuen Max-Planck-Instituts für Softwaresysteme an den beiden Standorten Saarbrücken und Kaiserslautern ein herausragendes Ereignis. Gründungsdirektor des neuen Instituts ist Prof. Dr. Peter Druschel. Während unser Institut primär an Algorithmen, den fundamentalen Bausteinen der Informatik, arbeitet, widmet sich das neue Institut vor allem dem Zusammenspiel solcher Komponenten in großen Systemen und dem grundsätzlichen Verständnis der Prinzipien komplexer Softwaresysteme. Zwischen beiden Instituten wird es intensive Zusammenarbeit geben.

2006 war das Jahr der Informatik, und unser Institut war in diesem Kontext an zahlreichen Veranstaltungen beteiligt: Ausstellungen, Vortragsreihen, Podiumsdiskussionen und vielem mehr. Höhepunkte waren die lange Nacht der Informatik in der Saarbrücker Innenstadt im Juli 2006 sowie die Festveranstaltung anlässlich des 15-jährigen Institutsjubiläums im November 2006, zu der mehr als 150 Alumni und externe Gäste ans Institut kamen.

Auch im Jahr 2007, dem Jahr der Geisteswissenschaften, gibt es bereits Pläne für Auftritte unseres Instituts in der breiteren Öffentlichkeit. Informatik ist ein Grundlagenfach, das nicht nur in den Naturwissenschaften und technischen Disziplin Anwendung findet, sondern auch für sprach- und kognitionswissenschaftliche Themen zunehmend an Bedeutung gewinnt.

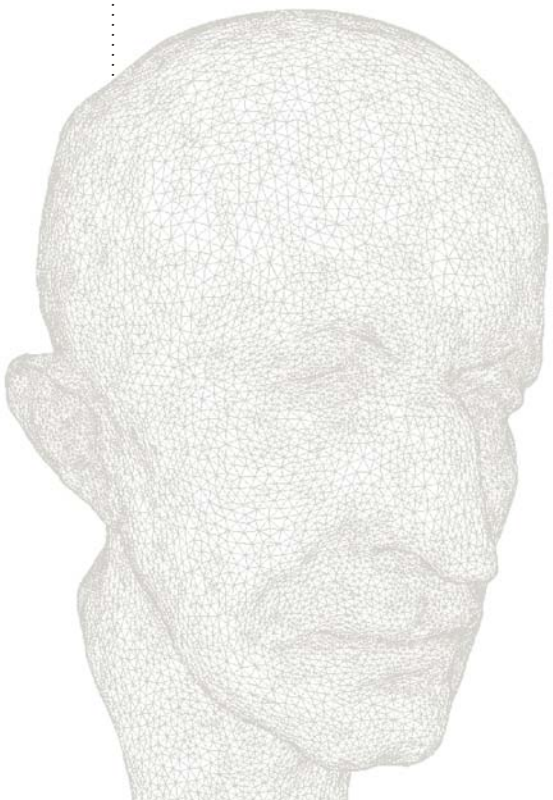
Ich wünsche Ihnen viel Spaß bei der Lektüre dieses Berichts.

Gerhard Weikum *Geschäftsführender Direktor*

Das Max-Planck-Institut für Informatik
Ein Überblick

Computersysteme beeinflussen in steigendem Maße unser Leben. Sie bilden die Grundlage nicht nur für praktisch alle geschäftlichen Prozesse, sondern haben schon seit längerem auch in Wissenschaft und Technik und im letzten Jahrzehnt auch in beeindruckendem Maße in unseren Alltag und in unsere Unterhaltung dominant Einzug gehalten. Heute ist die digitale Informationsverarbeitung aus praktisch keinem Bereich des Lebens mehr wegzudenken. Damit ist sie ein gesellschaftlich bestimmender Faktor.

Zusätzlich sind Computer sowie die auf ihnen laufende Software und die aus ihnen gebildeten Netzwerke – allen voran das weltumspannende Internet – wohl die komplexesten Strukturen, die je von Menschenhand geschaffen wurden. In der Tat sind Hardware und in noch weit größerem Maße Software so komplex, dass sie nicht mehr in allen ihren Einzelheiten verstanden werden können. Das macht Computersysteme zu einem sowohl machtvollen als auch mysteriösen Werkzeug. Sowohl das Leistungsvermögen als auch die Geheimnisse von Computersystemen verlangen nach ihrer wissenschaftlichen Erforschung.



ÜBERBLICK

Der wissenschaftliche Umgang mit Computersystemen ist Grundlagenforschung, die jedoch in vielen Fällen dramatische Fortschritte in der Anwendung nach sich zieht. Der Einzug von Arbeitsplatzrechnern mit Betriebssystemen mit Windows-Oberfläche (Xerox PARC), elektronischer Dokumentenverarbeitung (Stanford, Bell Labs), relationalen Datenbanken (Berkeley, IBM Research), sicherem Electronic Banking mittels kryptographischer Methoden (MIT), Kompressionsverfahren für Video und Musik (Fraunhofer), oft wenig mehr als eine Dekade nach den ausnahmslos in Grundlagenlabors erreichten Forschungsdurchbrüchen legt davon beredt Zeugnis ab.

Diese Überzeugung hat sich die Max-Planck-Gesellschaft vor etwa 16 Jahren zueigen gemacht, als sie das erste Max-Planck-Institut gründete, das sich ausschließlich mit Informatik – der Wissenschaft von Computersystemen *(engl. Computer Science)* – beschäftigt. Der Erfolg des Max-Planck-Institut für Informatik hat die Max-Planck-Gesellschaft überzeugt ein zweites Informatik-Institut, das neue Max-Planck-Institut für Softwaresysteme an den beiden Standorten Saarbrücken und Kaiserslautern, ins Leben zu rufen. Gründungsdirektor des neuen Instituts ist Prof. Dr. Peter Druschel.
Während das Max-Planck-Institut für Informatik primär an Algorithmen, den fundamentalen Bausteinen der Informatik, arbeitet, widmet sich das neue Institut vor allem dem Zusammenspiel solcher Komponenten in großen Systemen und dem grundsätzlichen Verständnis der Prinzipien komplexer Softwaresysteme.

Historie und Institutsstruktur

Das Max-Planck-Institut für Informatik wurde im Jahre 1990 gegründet. Prof. Kurt Mehlhorn war der Gründungsdirektor und leitet seitdem am Institut die Abteilung *„Algorithmen und Komplexität“*. Prof. Harald Ganzinger war von Anfang an mit dabei und leitete bis zu seinem Tod im Jahr 2004 die Abteilung *„Logik der Programmierung“*, die seitdem kommissarisch von Prof. Thomas Lengauer fortgeführt wird. Im Jahre 1999 folgte der Aufbau einer dritten Abteilung *„Computergraphik“* unter der Leitung von Prof. Hans-Peter Seidel.
Prof. Thomas Lengauer kam im Jahre 2001 an das Institut und leitet dort seitdem die Abteilung *„Bioinformatik und Angewandte Algorithmik“*. Seit 2003 leitet Prof. Gerhard Weikum am Institut die Abteilung *„Datenbanken und Informationssysteme“*. Im Vollausbau ist das Institut auf fünf Abteilungen konzipiert.

Neben den Abteilungen beherbergt das Institut selbständig arbeitende Forschungsgruppen. Derzeit sind die Forschungsgruppen *„Automatisierung der Logik“*, geleitet von PD Dr. Christoph Weidenbach, und *„Maschinelles Lernen“*, geleitet von Prof. Tobias Scheffer, am Institut tätig.

Forschungsthemen

Der zentrale Forschungsgegenstand des Instituts ist der *Algorithmus*. Ein Algorithmus ist eine Rechenvorschrift – eine genaue Anweisungsfolge an den Computer, wie er etwas zu berechnen hat. Unsere Arbeitshypothese ist der schnellere Fortschritt in der Informatik durch neue Algorithmen. In den letzten Jahrzehnten stellte die Entwicklung immer schnellerer Rechner einen Meilenstein beim Fortschritt in der Computertechnologie dar. Allerdings wird die dadurch erzielte Beschleunigung der Berechnungen von der Zunahme an Geschwindigkeit, Leistung und Robustheit in den Schatten gestellt, die durch neue Algorithmen erzielt werden. Um ein typisches Beispiel zu nennen: Der Stand der Hardware und Algorithmen im Jahr 1970 ermöglichte die Berechnung einer optimalen Reiseroute für einen Handelsreisenden (ein klassisches Optimierungsproblem und anerkannter Benchmark für die Rechenleistung) durch 120 Städte. Die Erhöhung der Anzahl an Städten von n auf n+1 führt zu einem multiplikativen Anstieg der Anzahl an möglichen Routen um einen Faktor n. Legen wir nun also die durch die heutige Technologie höhere Hardware-Geschwindigkeit und die Algorithmen von 1970 zugrunde, so könnten wir lediglich optimale Routen zwischen 135 Städten ermitteln. Es ist der Fortschritt bei den Algorithmen, der es heute ermöglicht, optimale Routen zwischen Tausenden von Städten zu finden. Würden wir uns hier nur auf den Fortschritt bei der Hardware verlassen, wäre eine solche Leistung in Hunderten von Jahren nicht möglich.

Die Aufgabe, die von eigener Hand erstellten Algorithmen und deren Realisierung in Computerprogrammen zu verstehen, ist eine wissenschaftliche und hat zwei wichtige Aspekte. Zum einen die Frage ob das Programm auch das berechnet was beabsichtigt war und auch nicht „abstürzt“, „einfriert“ oder alle Ressourcen des Computers blockiert. Zum anderen die Frage, ob das Programm auch „effizient“ ist und der beste mögliche Algorithmus gefunden wurde. In der Abteilung *„Algorithmen und Komplexität“* werden die Ressourcen untersucht, die ein Algorithmus für seine Berechnung braucht. Die wichtigsten Ressourcen sind Rechenzeit (Wie lange muss ich auf das Berechnungsergebnis warten?) und Speicherplatz (Reicht mein Speicher für meine Berechnung?). Dabei werden nicht

nur neue Algorithmen entwickelt, die den Bedarf an Rechenzeit und Speicherplatz minimieren und somit eine direkte hohe praktische Relevanz haben, sondern es werden auch die grundsätzlichen Grenzen dieser Vorgehensweise beleuchtet: Wieviel Rechenzeit/Speicherplatz ist grundsätzlich für eine Berechnung notwendig? Die Abteilung *„Logik der Programmierung“* und die Forschungsgruppe *„Automatisierung der Logik“* beschäftigen sich mit Bausteinen, mit deren Hilfe das korrekte Funktionieren eines Programms automatisch nachgewiesen werden kann.

Die Abteilung *„Computergraphik“* widmet sich dem Rechner als Instrument zur Darstellung von Bildern und Filmen. Sie trägt damit der Tatsache Rechnung, dass der Computer zunehmend nicht als Vermittler von Zahlen und Texten sondern vor allem von Bildern und multimedialen Daten in Erscheinung tritt. Auch hier geht es um die Grundfragen: Was ist grundsätzlich machbar? Und: Wieviele Ressourcen werden dafür benötigt? Anstelle des Korrektheitsbegriffs tritt hier der Begriff der naturgetreuen Wiedergabe, ein Konzept, das tiefgehende physikalische Aspekte beinhaltet. Ferner wird der Rechner nicht nur als Bildproduzent eingesetzt, sondern er soll (mithilfe geeigneter Algorithmen) auch Bilder „verstehen“, eine Aufgabe, die ebenfalls eine große wissenschaftliche Herausforderung darstellt. Die Abteilung entwickelt auf der anwendungsnahen Seite eine Vielzahl von Verfahren die zur schnellen Erstellung von besseren Bildern und Filmen führen.

Die Abteilung *„Bioinformatik und angewandte Algorithmik“* trägt der Tatsache Rechnung, dass der Computer in den letzten Jahren besonders im Bereich der Lebenswissenschaften eine zentrale Bedeutung erlangt hat, und hier insbesondere bei der Interpretation von biologischen Daten. Der Rechner ist ein wesentliches Instrument der modernen Biologie und Medizin. Das Verständnis biologischer Vorgänge auf molekularer Ebene ist ohne den Rechner nicht möglich, zum einen, weil es in der modernen Biologie immense Datenmengen zu verarbeiten gilt und zum anderen, weil die Komplexität der biochemischen Interaktionen in einem lebenden Organismus das Studium dieser Kreisläufe ohne Zuhilfenahme des Rechners aussichtslos macht. Bioinformatische Methoden sind somit ein Grundbestandteil für die moderne Forschung zur Diagnose und Therapie von Krankheiten.

Die Abteilung *„Datenbanken und Informationssysteme“* schließlich widmet sich besonders der Thematik der Verteilung, Organisation und Suche von Daten in großen Computernetzen wie dem Internet. Dabei stehen Aspekte der effektiven Suche nach Information in Netzen (Suchmaschinen wie Google sind entsprechende Instrumente), der Ausfallsicherheit von Methoden im Falle, dass Teile des Netzes nicht zugänglich sind, sowie der effektiven Verteilung von Rechenaufgaben auf im Netz zur Verfügung stehender Rechenleistung (z.B. in Peer-to-Peer-Systemen) im Vordergrund. Der praktische Nutzen dieser Forschung drängt sich auf: Wer hat sich nicht schon einmal gewünscht, mit graphischer statt textueller Information nach Bildern suchen zu können oder selbst bei schwierigen Anfragen von der Suchmaschine auch tatsächlich die relevanten Hits als erste präsentiert zu bekommen? Eine Grundlage dieser Vision ist das Arbeitsgebiet der Forschungsgruppe *„Maschinelles Lernen“*, die sich damit beschäftigt, wie aus der Flut verfügbarer Informationen kompakte Modelle zur Analyse und Vorhersage von Mustern und Trends

ÜBERBLICK

automatisch generiert werden können. Lernverfahren kommen z.B. bei der Diagnose von E-Mail-Spam und dubiosen Webseiten zum Einsatz.

Zielsetzung

Ziel des Max-Planck-Institut für Informatik ist es, gleichermaßen durch wissenschaftliche Publikationen, Software und Ausbildung des akademischen Nachwuchses Wirkung zu erzielen.

Publikationen

Wir betrachten die Informatik als einen Bereich, der danach strebt, den Einsatz von Rechnern auf ein tiefgehendes Verständnis der zugrunde liegenden Prinzipien der Algorithmen zu fundieren. Das umfasst die mathematische und/oder formale Untersuchung von Algorithmen, wo immer dies möglich ist. Folglich ist die Grundlagenforschung in diesem Bereich teilweise stark mathematisch geprägt. Ein Teil dieser Forschung gehört zu den etablierten Feldern der theoretischen Informatik und erzielt mathematisch präzise Resultate. Der andere Teil beschäftigt sich mit den praktischen Auswirkungen der Algorithmen und untersucht diese sowohl durch die Entwicklung und Evaluierung von Software als auch durch ihr Verhalten auf Modellen.

Wir sind darum bemüht, unsere Forschungsresultate zu verbreiten. Indem wir unsere Ideen den verschiedensten Forschungs- und Entwicklungsgemeinschaften vorstellen und durch sie testen lassen, gelangen wir zu einem besseren Verständnis. Wir veröffentlichen unsere Resultate und Ergebnisse in begutachteten, einschlägigen Zeitschriften und auf Konferenzen, und wir nutzen unsere Internet-Seiten, um sie so weit wie möglich für die Gemeinschaft verfügbar zu machen.

Software

In konkreten Anwendungsbereichen sind viele rechentechnische Probleme so komplex, dass eine tiefgehende formale Behandlung nicht durchführbar ist. Deshalb ist unsere Analyse der beteiligten Algorithmen in diesen Fällen eher experimentell. Diese stützt sich in der Regel auf eine systematische Bewertung auf der Grundlage von sorgfältig von Hand gepflegten Anwendungsdaten und speziell entwickelten statistischen Modellen und nicht zuletzt auf den Einsatz praktischer Systeme im Anwendungsgebiet. Tatsächlich sind viele Probleme in komplexen Anwendungsgebieten zunächst unscharf oder nur teilweise spezifiziert, so dass die Modellierung, d.h. die formale Definition eines Problems, ein wichtiger Aspekt der Forschung ist.

Die Tragfähigkeit neuer Ideen wird durch die Integration neuer Algorithmen in Software und die Bewertung ihrer Anwendbarkeit unter realistischen Bedingungen getestet. Kurzfristig ist diese Erfahrung für die Verfeinerung der Entwürfe nützlich. Langfristig ist sie unschätzbar für den Fortschritt des Wissens. Die meisten bedeutsamen Forschungsfortschritte im Bereich der Informatik und der Algorithmen entstanden durch diese Verbindung von neuen theoretischen Erkenntnissen und experimenteller Bewertung.

Wir suchen nach Benutzern unserer Prototyp-Software unter denjenigen, die gemeinsame Interessen mit uns haben, und wir fördern die Zusammenarbeit mit Forschern sowohl aus akademischen als auch aus industriellen Bereichen.

Nachwuchsförderung

Das dritte Ziel des Instituts ist die Schaffung eines stimulierenden Klimas für Nachwuchsforscher, damit diese die Möglichkeit haben, ihre eigenen Ideen zu entwickeln und eigene Gruppen aufzubauen. Das Max-Planck-Institut für Informatik betreibt ein aktives Förderprogramm für Doktoranden und Postdoktoranden. Dieses beginnt bis zur Promotion mit dem Doktorandenprogramm der *„International Max Planck Research School for Computer Science" (IMPRS-CS)* und erlaubt nach der Promotion über internationale Kooperationsabkommen wie dem *„Max Planck Center for Visual Computing"* im Bereich der Computergraphik und die Beteiligung an internationalen Forschungsprojekten den Austausch mit Spitzeninstitutionen in der ganzen Welt. Wir ermutigen damit unsere Nachwuchsforscher, ihre eigenen Forschungsprogramme zu etablieren und zu anderen Einrichtungen zu wechseln. Seit Gründung des Instituts gingen zahlreiche Forscher vom Saarbrücker Max-Planck-Institut für Informatik zu anderen Forschungseinrichtungen und viele von ihnen nahmen eine Professur an.

Gliederung des Berichts

Nach einer Kurzvorstellung der Abteilungen und Forschungsgruppen des Instituts gibt dieser Bericht einen Überblick über die Institutsarbeit, der nach Themenbereichen gegliedert ist. Der Bericht endet mit der Vorstellung der IMPRS-CS, einer Darstellung des Instituts in Zahlen, infrastruktureller Aspekte des Instituts sowie der tabellarischen Auflistung von Kooperationen und Publikationen. Wir wünschen Ihnen viel Freude bei der Lektüre und sind gerne bereit, weiterführende Fragen zu beantworten. Ansprechpartner werden für jedes Thema separat genannt. :::

Algorithmen und Komplexität

PROF. DR. KURT MEHLHORN

ABT. 1

Die Arbeitsgruppe existiert seit Gründung des Instituts und umfasst derzeit etwa 35 Mitarbeiter und Doktoranden. Unsere Ziele sind:

- **herausragende Grundlagenforschung im Bereich Algorithmen,**
- **Umsetzung unserer Grundlagenarbeiten in Demonstratoren und allgemein nützliche Softwarebibliotheken,**
- **Förderung des wissenschaftlichen Nachwuchses in einer stimulierenden Arbeitsgruppe.**

Wir sind in allen drei Aspekten erfolgreich und wirken durch Veröffentlichungen, Software und Personen. Wir publizieren reichlich und in den besten Zeitschriften und Tagungen des Gebiets, unsere Softwarebibliotheken LEDA und CGAL werden weltweit genutzt, die CompleteSearch Suchmaschine eröffnet neue Möglichkeiten für effiziente und intelligente Suche und viele ehemalige Mitglieder der Gruppe haben inzwischen gehobene Stellen in Forschung und Industrie im In- und Ausland.

KONTAKT

Algorithmen und Komplexität
Sekretariat
Ingrid Finkler-Paul | Petra Mayer
Telefon +49 681 9325-100
Email infi@mpi-inf.mpg.de
mayer@mpi-inf.mpg.de

ABT. 1

Algorithmen sind das Herz aller Softwaresysteme. Wir bearbeiten den Entwurf und die Analyse von Algorithmen in vielen Facetten: kombinatorische, geometrische und algebraische Algorithmen, Datenstrukturen und Suchverfahren, verschiedenste Rechnermodelle (sequentiell, parallel, verteilt, flacher Speicher oder Speicherhierarchie), exakte und approximative Lösungen, deterministische und randomisierte Lösungen, obere und untere Schranken, Analyse im schlechtesten Fall und im Mittel. Dabei geht es um die Entwicklung effizienter Algorithmen sowohl für Modellprobleme (= die abstrahierte Version von Anwendungsproblemen) als auch für konkrete Anwendungen, z.B. intelligente Suche oder Bioinformatik oder Computer-Aided Design. Einen Teil unserer theoretischen Einsichten setzen wir um in Software-Demonstratoren und Softwarebibliotheken.

Herausragende theoretische Ergebnisse der letzten beiden Jahre sind neue Algorithmen zur exakten Nullstellenbestimmung von Polynomen mit reellen Koeffizienten, zur Bestimmung der Topologie und Geometrie von algebraischen Kurven, zur Bestimmung der optimalen Zyklenbasis von Graphen, Approximationsalgorithmen für Scheduling Probleme und Distanzprobleme, exakte Algorithmen für Steinerbaumprobleme, Algorithmen zur Analyse der Topologie und Geometrie von ad-hoc Netzwerken, neue Methoden für Randomisiertes Runden, für die Analyse von evolutionären Algorithmen und für exaktes geometrisches Rechnen mit approximativer Arithmetik.

Herausragende praktische Ergebnisse der letzten beiden Jahre sind die Softwarebibliothek EXACUS zum exakten Rechnen mit Kurven und Flächen, die CompleteSearch Suchmaschine und höchst effiziente Verfahren zur Bestimmung schnellster Wege in Straßengraphen.

Unsere theoretischen und praktischen Arbeiten befruchten sich gegenseitig. Unsere theoretischen Arbeiten sind die Grundlage für die Demonstratoren und Bibliotheken. So beruht die CompleteSearch Engine etwa auf einer neuen Indexstruktur, die mächtiger ist als bekannte Strukturen aber dennoch nicht mehr Platz benötigt. So beruht EXACUS ganz wesentlich auf den neuen Algorithmen zur Isolation von Nullstellen und zur Bestimmung der Topologie von algebraischen Kurven, und so ist das Verfahren zur Wegesuche inspiriert von der Theorie der Graphspanner. Umgekehrt zeigen die experimentellen Arbeiten die Grenzen der Theorie auf und führen zu neuen Fragestellungen. Außerdem befriedigt es enorm, wenn unsere Systeme von anderen Forschern genutzt werden und zu Anwendungen in der Industrie führen.

Die Kombination von theoretischer und experimenteller Forschung in der Algorithmik hat sich inzwischen breiter durchgesetzt. Die DFG hat gerade ein Schwerpunktprogramm Algorithm Engineering eingerichtet.

Die Gruppe ist in mehrere internationale Projekte eingebunden: die europäischen Projekte ACS (Algorithms for Complex Shapes) und DELIS (Dynamically Evolving Large Information Systems) und das GIF-Projekt Graphenalgorithmen (mit der Universität Tel Aviv). In Deutschland nehmen wir an dem Schwerpunktprogramm Algorithm Engineering mit drei Teilprojekten teil.

Viele ehemalige Mitarbeiter der Gruppe sind nun an Universitäten, Forschungseinrichtungen und in der Industrie im In- und Ausland tätig. :::

Logik der Programmierung

KOMMISSARISCH PROF. DR. THOMAS LENGAUER, PH.D.

ABT. 2

„Logik der Programmierung“ ist eine der zwei Abteilungen, die seit der Gründung des Max-Planck-Instituts für Informatik 1991 bestehen. Bis zu seinem Tod im Juni 2004 wurde die Abteilung von Prof. Dr. Harald Ganzinger geleitet, seitdem kommissarisch durch Prof. Dr. Thomas Lengauer, Ph.D. Schwerpunktthema der Abteilung ist gegenwärtig die deduktionsbasierte Programm- und Systemanalyse.

KONTAKT

Logik der Programmierung
Sekretariat
Ellen Fries
Telefon +49 681 9325-502
Email fries@mpi-inf.mpg.de

ABT. 2

Forschungsziel der deduktiven Programm- und Systemanalyse sind Verfahren, die automatisch Eigenschaften von Programmen und Systemen berechnen. Im Falle von Programmen sind die konkreten Ziele dieser Verfahren die Überprüfung der Abwesenheit von Laufzeitfehlern bei der Ausführung des analysierten Programms und die Abwesenheit von Inaktivität. Ein typischer Laufzeitfehler ist ein *buffer overflow* (der vorhandene Speicher z.B. für die Darstellung der Werte eines Zählers wird überschritten), ein typischer Inaktivitätszustand ein *deadlock* (z.B. wartet bei der Übertragung von Daten zwischen dem Betriebssystem und einem Gerätetreiber der eine auf den anderen).

Im Falle von Systemen werden aktuell Stabilitätseigenschaften Hybrider Systeme sowie Sicherheits- und Prozesseigenschaften von Systemen in der öffentlichen Verwaltung untersucht. In allen möglichen Anwendungen ist die Garantie der Abwesenheit von Fehlern ein wertvolles Qualitätsmerkmal des Produktes. In Zusammenarbeit mit Forschergruppen aus der ganzen Welt werden neuartige Verfahren entwickelt, die die oben genannten Verfahren in klassische Methoden der Programm- und Systemanalyse einbinden, wie sie heute jeder Computer zur Optimierung benutzt.

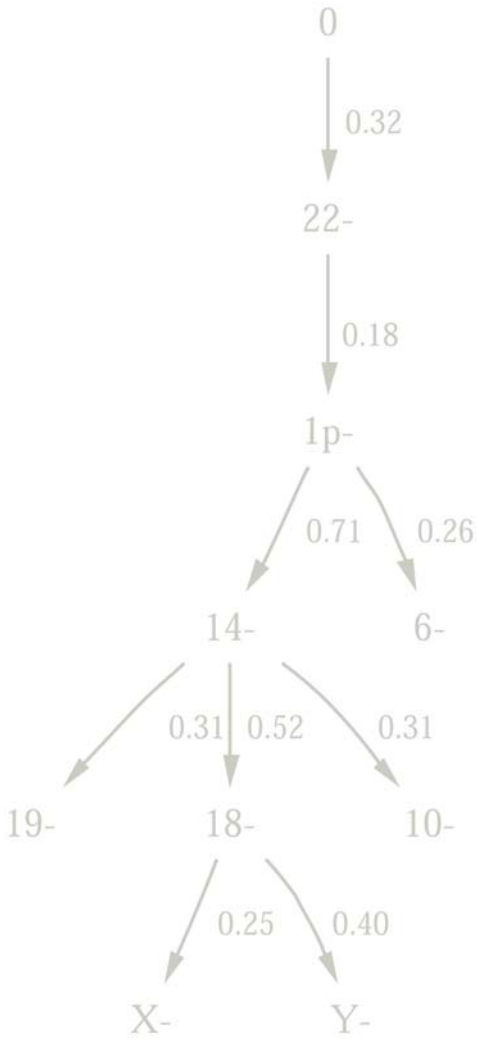
Die Abteilung ist Mitglied des Transregio-Sonderforschungsbereichs „AVACS“ der Deutschen Forschungsgemeinschaft. Das europäische Forschungsprojekt „R4eGov“, an dem in Saarbrücken außer dem Max-Planck-Institut für Informatik noch das Institut für Wirtschaftsinformatik am Deutschen Forschungszentrum für Künstliche Intelligenz beteiligt ist, beschäftigt sich mit der Digitalisierung europäischer Verwaltungen und insbesondere mit der Analyse von Geschäftsprozessen. :::

Bioinformatik und Angewandte Algorithmik

PROF. DR. THOMAS LENGAUER, PH.D.

ABT. 3

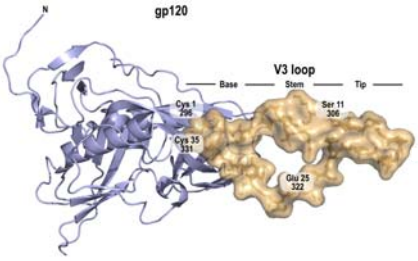
Diese Abteilung existiert seit Oktober 2001 und wird von Prof. Dr. Thomas Lengauer, Ph.D. geleitet. Die Abteilung hat etwa 25 Wissenschaftler. Sie forscht derzeit ausschließlich in den Gebieten Bioinformatik und Chemie-Informatik.

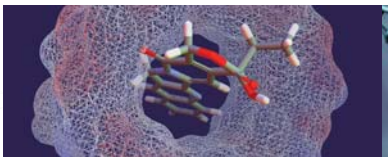


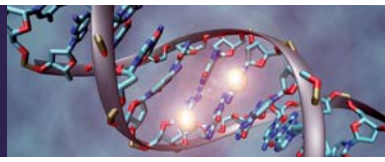
KONTAKT

Bioinformatik und Angewandte Algorithmik
Sekretariat
Ruth Schneppen-Christmann
Telefon +49 681 9325-300
Email ruth@mpi-inf.mpg.de

ABT. 3







Im Bereich der Bioinformatik forscht die Abteilung vornehmlich an Themen, die im engeren oder weiteren Sinne für die Diagnose und Therapie von Krankheiten von Belang sind. Auf molekularer Ebene können Krankheitsprozesse auf Anomalien in der „biochemischen Verschaltung" eines Organismus zurückgeführt werden. Die Bausteine solcher biochemischer Netzwerke sind in der Regel Proteine, die aneinander, an Nukleinsäuren oder an kleine organische Moleküle binden und auf diese Weise Reaktionen katalysieren, das Ablesen von Genen steuern oder Signale weiterleiten. Die Aufklärung dieser Funktionsweisen erfordert die Bestimmung der dreidimensionalen Strukturen der beteiligten Proteine, die Analyse von Beziehungen zwischen Proteinstruktur und Proteinfunktion (*„Struktur-Funktionsbeziehungen bei Proteinen"*, Seite 32), die Modellierung von Bindungsereignissen zwischen Biomolekülen (*„Docking und Wirkstoffdesign"*, Seite 30) sowie die Analyse von komplexen Wechselwirkungs-Netzwerken zwischen Proteinen (*„Analyse von Proteinnetzwerken"*, Seite 35). Am Max-Planck-Institut für Informatik wird in allen diesen Bereichen Methodenentwicklung betrieben (*„Algorithmen in der Bioinformatik"*, Seite 28 und *„Geometrische Probleme in der Bioinformatik"*, Seite 50). Die Methoden werden darüber hinaus in konkreten Fallbeispielen auf Infektionskrankheiten wie AIDS und auf altersbedingte Krankheiten wie Krebs und neurodegenerativen Krankheiten angewandt. Während bei Krebs die Früherkennung anhand genetischer und so genannter epigenetischer Veränderungen im Vordergrund steht (*„Statistische Analyse bei medizinischer Diagnose und Prognose"*, Seite 79, *„Bioinformatische Epigenetik: Bioinformatik für neue Wege in der Krebsforschung"*, Seite 78), liegt der Fokus bei neurodegenerativen Krankheiten auf der Identifizierung und Charakterisierung krankheitsinduzierender Proteine (*„Funktionsanalyse medizinisch relevanter Proteine"*, Seite 33). Bei der Suche nach optimierten Therapien für Infektionskrankheiten spielt AIDS eine herausragende Rolle. Für diese Krankheit gehen wir am Max-Planck-Institut für Informatik sogar noch einen Schritt weiter und analysieren Resistenzen des HI-Virus gegen verabreichte Wirkstofftherapien (*„Analyse von HIV-Resistenz"*, Seite 34).

Ein Großteil der Methodenentwicklung führt zu Softwaresystemen, die weltweit von zahlreichen akademischen und oft auch industriellen Nutzern angewandt werden. Beispiele hierfür sind das Programm BiQ Analyzer zur Qualitätssicherung epigenetischer Daten (*„Bioinformatische Epigenetik: Bioinformatik für neue Wege in der Krebsforschung"*, Seite 78), sowie der Server geno2pheno zur Analyse von HIV-Resistenzen (*„Analyse von HIV-Resistenz"*, Seite 34).

Die Forschung über molekulare Interaktionen findet auch ihre Anwendung in nichtbiologischen Bereichen der Katalyse und Materialforschung (*„Chemieinformatik"*, Seite 31).

Die Abteilung ist einer der tragenden Säulen des Zentrums für Bioinformatik Saar, einer wissenschaftlichen Einrichtung an der Universität des Saarlandes, die Lehre und Forschung im Bereich der Bioinformatik zum Gegenstand hat. Die Abteilung ist Mitglied des Network of Excellence „Biosapiens", sowie der Strategischen Forschungsinitiative „Euresist" der Europäischen Union, der Klinischen Forschergruppe 129 der Deutschen Forschungsgemeinschaft zur Aufklärung der Funktion des Erregers HCV der Hepatitis C, sowie des vom Bundesforschungsministeriums geförderten Nationalen Genomforschungsnetzes.

Einen umfassenden Überblick über das weitere Forschungsgebiet der Abteilung 3 gibt das von Prof. Lengauer neu herausgegebene Buch *„Bioinformatics – From Genomes to Therapies"*, das bei Wiley-VCH erschienen ist (Seite 29). :::

Computergraphik

PROF. DR. HANS-PETER SEIDEL

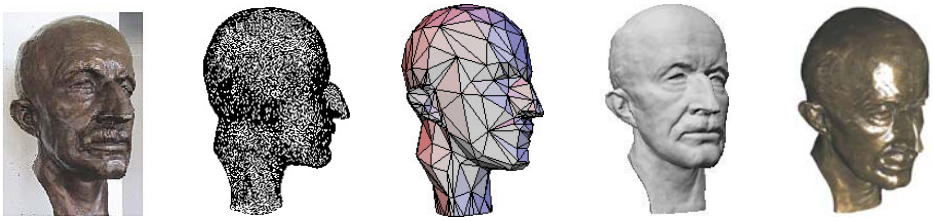
ABT. 4

Die Arbeitsgruppe Computergraphik wurde 1999 gegründet und umfasst derzeit knapp 40 Wissenschaftler. Ein wichtiges Charakteristikum der Arbeiten ist die durchgängige Betrachtung der gesamten Verarbeitungskette von der Datenakquisition über die Modellierung bis zur Bildsynthese (3D Bildanalyse und -synthese). Typisch für das Gebiet ist das Zusammentreffen sehr großer Datensätze mit der Forderung nach schneller, wenn möglich interaktiver, Darstellung.

KONTAKT

Computergraphik
Sekretariat
Sabine Budde
Telefon +49 681 9325-400
Email budde@mpi-inf.mpg.de

ABT. 4



Computer werden heute vielfach dazu benutzt, um Ausschnitte der realen oder einer virtuellen Welt auf dem Rechner nachzubilden, zu simulieren und darzustellen. Aufgrund der Bedeutung visueller Information für den Menschen hat sich die Computergraphik deshalb im vergangenen Jahrzehnt zu einer Schlüsseltechnologie der modernen Informations- und Kommunikationsgesellschaft entwickelt, deren zukünftiges Anwendungspotential durch Schlagworte wie Multimedia, digitales Fernsehen, Telekommunikation, virtuelle Realität oder 3D-Internet lediglich angedeutet ist. Typisch für das Gebiet ist das Zusammentreffen sehr großer Datensätze mit der Forderung nach schneller (wenn möglich interaktiver) visueller Darstellung der Ergebnisse mit hoher Bildqualität. Außerdem soll der Benutzer in die Lage versetzt werden, auf möglichst intuitive Art und Weise mit seiner Umgebung zu interagieren. Hierbei werden verteilte Anwendungen immer wichtiger.

Die genannten Herausforderungen erfordern auch in wissenschaftlicher Hinsicht neue Ansätze. Ein wichtiges Charakteristikum der Arbeitsgruppe ist deshalb die durchgängige Betrachtung der gesamten Verarbeitungskette von der Datenakquisition über die Modellierung (Erzeugung einer geeigneten rechnerinternen Szenebeschreibung) bis zur Bildsynthese (Erzeugung von beliebigen Ansichten). Diese integrierte Sichtweise ist notwendig, um die Leistungsfähigkeit moderner Hardware sowohl bei der Eingabe (bildgebende Verfahren) wie auch bei der Ausgabe (Graphikhardware) adäquat auszunutzen. Inzwischen wurde für diese integrierte Sichtweise der Begriff der 3D-Bildanalyse und -synthese geprägt. Als zentrale wissenschaftliche Herausforderungen ergeben sich hieraus insbesondere die Entwicklung geeigneter Modellierungswerkzeuge zur effizienten Handhabung und Weiterverarbeitung der Datenflut auf der Eingabeseite sowie die Entwicklung neuer Algorithmen zur schnellen und dabei qualitativ hochwertigen Darstellung unter enger Verzahnung mit den Möglichkeiten und Perspektiven moderner Graphikhardware auf der Ausgabeseite.

Die wissenschaftlichen Aktivitäten der Arbeitsgruppe Computergraphik sind in eine Reihe von Projektaktivitäten auf nationaler, europäischer und internationaler Ebene eingebettet.

Von besonderer Bedeutung ist das von der Max-Planck-Gesellschaft und Stanford University mit Unterstützung des BMBF im Oktober 2003 gemeinsam eingerichtete *„Max Planck Center for Visual Computing and Communication“*. Ziel dieses Brückenschlags zwischen den beiden herausragenden Standorten in Deutschland und in den USA ist es, die Forschungsanstrengungen auf diesem Schlüsselgebiet der modernen Informations- und Kommunikationstechnologie zu stärken und zu bündeln, und durch die Etablierung neuer Austauschmechanismen mit attraktiven Rückkehrmöglichkeiten einen wesentlichen Beitrag zur Herausbildung und Rückgewinnung hervorragender Nachwuchswissenschaftler zu leisten. Die gemeinsame Leitung des Zentrums liegt in den Händen von Professor Bernd Girod (Stanford University) und Professor Hans-Peter Seidel (Max-Planck-Institut für Informatik).

Mehrere Wissenschaftler der Gruppe erhielten Rufe auf Professuren im In- und Ausland. Die Gruppe hat eine Reihe von Preisen angezogen, darunter neben Nachwuchspreisen für die Wissenschaftler auch den Leibniz-Preis der Deutschen Forschungsgemeinschaft für Professor Seidel. :::

Datenbanken und Informationssysteme

PROF. DR.-ING. GERHARD WEIKUM

ABT. 5

Die von Gerhard Weikum geleitete Abteilung wurde im Oktober 2003 am Max-Planck-Institut für Informatik etabliert. Die Abteilung forscht in zwei großen Schwerpunktbereichen:

1. **der intelligenten Informationsorganisation und -suche, insbesondere in Intranets, digitalen Bibliotheken und im Web, sowie**
2. **der Architektur und Beherrschung selbst organisierender verteilter Informationssysteme, insbesondere so genannter Peer-to-Peer-Systeme.**

KONTAKT

Datenbanken und Informationssysteme
Sekretariat
Petra Schaaf
Telefon +49 681 9325-500
Email schaaf@mpi-inf.mpg.de

ABT. 5

Internet-Suchmaschinen liefern für einfache Anfragen sehr gute Suchresultate, doch gibt es auch viele Situationen, in denen sie versagen. Zwei schwierige Beispiele, bei denen heutige Methoden scheitern, sind: *„Welche Professoren, die in Saarbrücken arbeiten, halten Vorlesungen über Information Retrieval und sind an EU-Projekten beteiligt?“* und *„Wie heißt die Französin, die ich bei der Programmkomiteesitzung traf, bei der Gottfried Vossen Vorsitzender war?“*. Im ersten Fall qualifiziert sich keine einzelne Web-Seite; vielmehr muss man mehrere Web-Seiten im Zusammenhang sehen: die Homepage eines Saarbrücker Informatikprofessors, eine Seite über bestimmte Vorlesungen und eine dritte Seite mit Projekten. Für die zweite Suche müsste man Informationen aus unterschiedlichsten Quellen zusammenbringen und richtig miteinander verknüpfen: persönliche E-Mail-Archive, Reisekostenabrechnungen, Konferenzprogramme aus dem Web, Homepages von Wissenschaftlern. Wenn man durch solche Verknüpfungsketten auf jemanden wie Sophie Cluet von INRIA, Paris, trifft, hat man mit hoher Wahrscheinlichkeit die richtige Antwort gefunden.

Wir arbeiten an einer statistisch basierten Variante der Vision vom *„Semantic Web“*, indem alle Informationen mit Konzept-Werte-Paaren und expliziten Verknüpfungen exakter repräsentiert und semantisch tiefer gehend annotiert werden. Unser Ansatz kombiniert logikbasierte, präzise Suchverfahren mit probabilistischen Methoden. Auf diesem Weg kann der Rechner schließen, dass Sophie mit hoher Wahrscheinlichkeit ein weiblicher Vorname ist und dass mit dem Ort Paris vermutlich die Hauptstadt Frankreichs gemeint ist und nicht das durch einen Film bekannt gewordene 500-Seelen-Dorf Paris, Texas.

Die an der Abteilung entwickelte Suchmaschine TopX für Web- und semistrukturierte XML-Daten integriert automatische Annotationstechniken, Hintergrundwissen in Form von Ontologien und Thesauri und statistische Lernverfahren für die Relevanzbewertung und das Treffer-Ranking. Peer-to-Peer-Systeme, salopp kurz P2P-Systeme genannt, sind vollständig dezentralisierte Systeme mit Tausenden oder Millionen von Rechnern, die sich in einem dynamischen Verbund selbständig organisieren, um gemeinsam lang laufende Berechnungen verteilt durchzuführen oder große, verteilt vorliegende Informationsmengen zu verwalten. Bekannte Beispiele sind Musiktauschbörsen wie Gnutella oder BitTorrent; künftige Anwendungen könnten eine P2P-Suchmaschine für das Internet sowie die Langzeitarchivierung des World Wide Webs sein. Bei einer P2P-Suchmaschine stellen wir uns vor, dass jeder Benutzer eine vollständige Suchmaschine mit einem kleinen Index von z.B. 10 Gigabyte auf seinem persönlichen Rechner hat. Der Index ist für das jeweilige Interessenprofil spezialisiert, indem man z.B. unsere fokussierte Crawling-Technologie verwendet. Anfragen des Benutzers werden zunächst lokal ausgewertet; wenn das Resultat aber das Informationsbedürfnis nicht voll befriedigt, werden andere *„Peers“* automatisch kontaktiert und um Mithilfe bei der Suche gebeten. Die Art und Weise, wie Dutzende von Peers für eine einzelne Anfrage zusammenarbeiten und wie sich Millionen von Peers langfristig untereinander verbinden, sollte völlig selbst organisierend sein, also ohne menschliche Eingriffe (durch Systemadministratoren) auskommen. Eine solche Architektur ist nicht nur aus Informatiksicht interessant, sondern hat auch das Potential, die Suchresultatsgüte global für alle Benutzer zu verbessern, indem man den impliziten intellektuellen Input (Anfragen, Klicks, etc.) aller Benutzer berücksichtigt. Alleine schon die Kenntnis der Bookmarks von Millionen von Peers könnte mit geeigneten statistischen Analysen und Lernverfahren zu einem Quantensprung in der Suchresultatsgüte führen. Darüber hinaus würde dieser Ansatz die Gefahr des De-facto-Monopols einer einzigen, zentralisierten Web-Suchmaschine bannen.

Das methodische Repertoire der Abteilung erstreckt sich von der Theoriebildung bis zum praktischen Einsatz neuer Konzepte in realen Anwendungen und umfangreichen Experimenten. In den experimentellen Arbeiten kommen auch selbst entwickelte Softwaresysteme zum Einsatz: der fokussierte Web-Crawler Bingo!, die XML-Suchmaschine TopX, die Peer-to-Peer-Suchmaschine Minerva sowie die Softwarebibliothek Yago für die automatische Konstruktion und Pflege von Ontologien.

Die Abteilung ist an einer Reihe von Drittmittelprojekten beteiligt. Sie koordiniert den Themenbereich der Peer-to-Peer-Systeme im EU-Projekt DELIS über Dynamically Evolving Large-Scale Information Systems, und sie ist für die Architektur einer neuartigen Peer-to-Peer-Suchmaschine für audio-visuelle Daten im EU-Projekt SAPIR verantwortlich. Ein Mitarbeiter der Abteilung, Martin Theobald, wurde im März 2007 mit dem Dissertationspreis des Fachbereichs Datenbanken und Informationssysteme der Gesellschaft für Informatik (GI) ausgezeichnet, und die wissenschaftlichen Beiträge des Abteilungsleiters, Gerhard Weikum, wurden 2005 durch seine Ernennung zum ACM Fellow gewürdigt. :::

FG.1

Automatisierung der Logik

DR. CHRISTOPH WEIDENBACH

Die unabhängige Forschungsgruppe *Automatisierung der Logik* unter der Leitung von Dr. Christoph Weidenbach beschäftigt sich mit der kompletten Pipeline von der Erforschung formaler Beschreibungssprachen (Logiken) bis hin zum automatischen Berechnen von Sätzen der Sprachen (Formeln der Logik).

Damit die Informationstechnologie auch in Zukunft weiter Innovation vorantreiben kann, müssen ihre Produkte (Systeme, Software) noch robuster und qualitativ hochwertiger werden. Sonst sind Anwendungsszenarien, bei denen z.B. Fahrzeuge selbsttätig Bremsmanöver koordinieren, um eine Kollision zu vermeiden, oder die automatische Reaktion der Heizung unserer Wohnung auf die Urlaubsplanung im elektronischen Kalender unseres Handys nicht zufriedenstellend realisierbar.

Ein Beitrag zu robusten und qualitativ hochwertigen Systemen (Software) der Informationstechnologie ist die Unterstützung ihres Lebenszyklus (Spezifikation, Programmierung, Testen, Warten) durch formale Methoden. Das heißt: die Entwicklung von Verfahren, die den Code oder die Designmodelle von Systemen gemäß einer vorgegebenen oder frei wählbaren Spezifikation analysieren, testen und so Fehler finden. Das hört sich utopisch an, ist aber im Kleinen in der Softwareentwicklung schon lange Realität. So ist es heute Standard, dass Compiler die Typen in Programmen automatisch überprüfen und so Fehler vermeiden.

Abhängig vom Risikopotential des untersuchten Systems (der Software), fordert die Praxis das ganze Spektrum von vollständig formal mathematisch exakt verifizierten, bis hin zu getesteten Systemen. Wir beschäftigen uns mit der formalen Analyse von Systemen auf der Basis von Logik. Um hier mit den immer komplexeren Systemen Schritt zu halten, müssen die heute verwendeten Verfahren ihre Produktivität steigern. Das ist das Ziel unserer Forschungsgruppe „Automatisierung der Logik“. Wir wollen durch einen höheren Automatisierungsgrad die Produktivität formaler Methoden verbessern. :::

KONTAKT

PD Dr. Christoph Weidenbach
FG.1 Automatisierung der Logik
Sekretariat Roxane Wetzel
Telefon +49 681 9325-900
Email wetzel@mpi-inf.mpg.de

FG.2

Maschinelles Lernen

DR. TOBIAS SCHEFFER

Seit dem 1. Januar 2007 arbeitet die Forschungsgruppe *Maschinelles Lernen* unter der Leitung von Tobias Scheffer am Max-Planck-Institut für Informatik. In der Gruppe arbeiten zurzeit neun Wissenschaftler an der Entwicklung von Algorithmen, die durch die Analyse von Daten Wissen gewinnen. Verfahren des maschinellen Lernens helfen, viele Probleme unter anderem in den Bereichen Sprachverarbeitung und Informationssuche zu lösen.

Zu den Forschungsschwerpunkten der Gruppe zählt die Konstruktion von Algorithmen, die ihr in einem Bereich erworbenes Wissen auf eine veränderte Aufgabe oder Einsatzumgebung anpassen können. Beispielsweise werden Übersetzungssysteme häufig mit mehrsprachigen Parlamentsunterlagen trainiert, später aber zur Übersetzung von Gebrauchsanleitungen eingesetzt. Ein weiterer Schwerpunkt ist die Analyse von Lernproblemen im Umfeld der IT-Sicherheit. Diese Probleme sind dadurch gekennzeichnet, dass ein Gegenspieler aktiv bemüht ist, den Lernerfolg zu verhindern. Spam-Versender beispielsweise optimieren ihre Werkzeuge darauf hin, dass ihre Spam-Emails nicht von den eingesetzten Filtern erkannt werden.

Die Arbeitsgruppe führt eine Reihe von Drittmittelprojekten durch. Das Projekt „Text Mining“ wird von der Deutschen Forschungsgemeinschaft gefördert. Die Forschungsgruppe entwickelt in Zusammenarbeit mit der STRATO AG einen Filter gegen Spam, Phishing, Viren und Bilderspam. Die Gruppe verfolgt Kooperationen mit der DaimlerChrysler AG im Bereich der Analyse von Werkstattberichten und Qualitätsdatenbanken und mit der nugg.ad AG bei der Optimierung und Personalisierung von Anzeigen auf Webseiten. :::

KONTAKT

Prof. Dr. Tobias Scheffer
FG.2 Maschinelles Lernen
Sekretariat Ellen Fries
Telefon +49 681 9325-502
Email ef@mpi-inf.mpg.de

Max Planck Center for Visual Computing and Communication

PROJEKTLEITUNG: PROF. DR. HANS-PETER SEIDEL

Das Max-Planck-Institut für Informatik, Saarbrücken, und die Stanford University, USA, kooperieren seit drei Jahren in einem virtuellen „Max Planck Center for Visual Computing and Communication" (MPC-VCC). Das Bundesministerium für Bildung und Forschung (BMBF) bewilligte der Max-Planck-Gesellschaft für eine erste Phase mehrere Forschungsprojekte im Umfang von 6,9 Mio. Euro für sechs Jahre.

Die Forschungsschwerpunkte dieser Kooperation liegen auf der Grundlagenforschung im Bereich des Visual Computing and Communication, die die informationstechnischen Teilgebiete der Bildaufnahme (Bildakquisition), die Verarbeitung und Analyse von Bilddaten (Bildanalyse), die Erzeugung von Bildern und Bildsequenzen auf der Basis von Aufnahmen oder Simulationen (Bildsynthese), die Visualisierung komplexer Daten sowie den ungestörten und schnellen Austausch von Informationen in komplexen Netzwerken umfassen. Dazu bedarf es gleichzeitig der Entwicklung leistungsfähiger Personalcomputer und Betriebssysteme insbesondere als Grafik- und Multimediasysteme.

Stärkung des Wissenschafts- und Forschungsstandorts Deutschland im Bereich der Informatik

Mit diesem Programm soll zugleich dem akuten Mangel an qualifizierten Informatikern für die Hochschulen im Fach Informatik in Deutschland entgegengewirkt werden, um durch finanzielle und personelle Förderung hochbegabter Eliten aus aller Welt die Innovations- und Wettbewerbsfähigkeit des Wissenschafts- und Forschungsstandorts Deutschland zu stärken. Die Forschungsprojekte schließen jeweils einen maximal zweijährigen Forschungsaufenthalt an der Stanford University als „Assistant Professor" bzw. eine eigenverantwortliche mehrjährige Forschungstätigkeit am Max-Planck-Institut für Informatik ein, verbunden mit der Leitung einer eigenen Nachwuchsgruppe.

Derzeitiger Stand des MPC-VCC

Seit dem 01.10.2003 wurden bis Dezember 2006 vierzehn Forschungsgruppen innerhalb des Max Planck Centers aufgebaut. Zurzeit sind jeweils sechs Forschungsgruppen in Stanford und am Max-Planck-Institut für Informatik etabliert. Die Wissenschaftler arbeiten an gemeinsamen Projekten. Gruppenleiter betreuen z.T. Doktoranden in Stanford und in Saarbrücken. Die Forschungsergebnisse wurden durch zahlreiche Publikationen in namhaften Zeitschriften und auf den wichtigsten internationalen Konferenzen (z.B. SIGGRAPH 05, SIGGRAPH 07) bekannt gemacht und legen die Grundsteine für die Weiterentwicklung innovativer Technologien. In 2005/2006 wurden sieben Mitglieder des Max Planck Centers mit renommierten Preisen ausgezeichnet. Vier Gruppenleiter haben bereits Stellen als Professoren an deutschen Hochschulen inne.

Aktuelles Forschungsprogramm in Saarbrücken und Stanford

Akquisition und Modellierung Akquisition hochqualitativer 3D Objekte und deren realistische Darstellung (Dr. Hendrik Lensch) und Entwicklung von Editierungs- und Modellierungstechniken für komplexe Szenen (Dr. Michael Wand)

Bildanalyse und -synthese Lernbasierte Animation von 3D Modellen von Gesichtern und Objekten (Prof. Dr. Volker Blanz), markerfreie Bewegungsanalyse mit zukünftigen Anwendungsgebieten in der Medizin, Sportwissenschaft sowie in der Filmindustrie (Dr. Bodo Rosenhahn), Entwicklung von 3D Videos zur Erzeugung realistischer visueller Darstellungen von virtuellen dynamischen Szenen im Computer (Dr. Christian Theobalt) sowie Simulationen von physikalisch korrekten Animationen auf parallelen Rechnerarchitekturen (Dr. Robert Strzodka)

Visualisierung Entwicklung topologischer Methoden für periodisch zeitabhängige Vektorfelder zur Visualisierung von z.B. komplexen Strömungsdaten (Prof. Dr. Holger Theisel), visuelle Datenexploration bzw. die Visualisierung von Informationen (Dr. Mike Sips) sowie im Bereich des maschinellen Lernens Methodenentwicklung zur Beurteilung der wahrgenommenen Qualität von Visualisierungsalgorithmen (Dr. Joachim Giesen)

Kommunikation Optimierung optischer Sensornetzwerke (Dr. Markus Flierl), Geometrie-basierter Aufbau und Analyse von drahtlosen Sensornetzwerken zur Erkennung von so genannten „Löchern" im Kommunikationsgraphen (Dr. Stefan Funke) sowie die Konzipierung einer Softwarearchitektur für verteilte Medien Systeme (Dr. Pierpaolo Baccichet)

KONTAKT

Prof. Dr. Hans-Peter Seidel
Abt 4. Computergraphik
Telefon +49 681 9325-400
Email hpseidel@mpi-inf.mpg.d

Forschungsschwerpunkte

BIOINFORMATIK

Die Bioinformatik ist eine Schlüsseldisziplin für den schnelleren Erkenntnisfortschritt in den Biowissenschaften, wie Biotechnologie, Pharmazie und Medizin. Die Bioinformatik vertieft und beschleunigt mit Hilfe des Computers die Planung von höchst komplexen biologischen Experimenten und die Interpretation der in sehr großen Mengen anfallenden Daten.

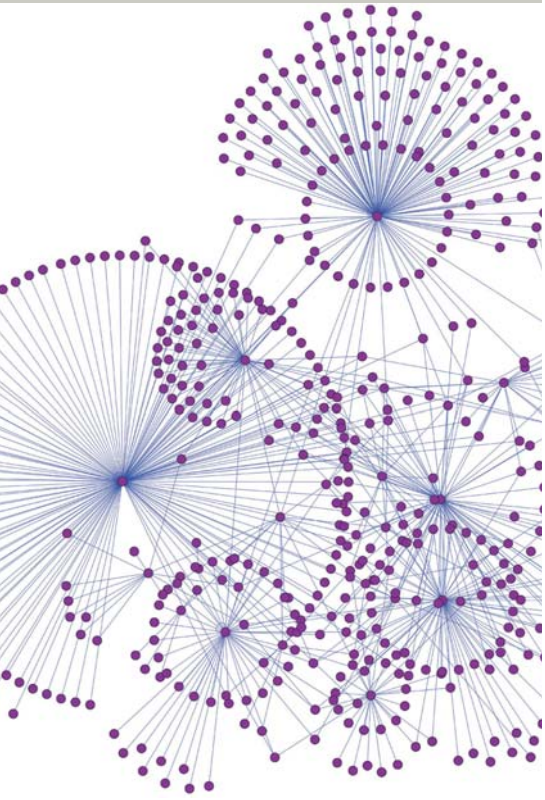
Seit etwa 10 Jahren trägt die Bioinformatik wesentlich zum Erkenntnisgewinn in den Biowissenschaften bei. Sie ist Teil einer Revolution der Biologie. Sie unterstützt Forscher bei der Planung von Experimenten, sie erhebt Daten, die aus allen Bereichen des Organismus stammen und wertet diese Daten mit computergestützten Methoden aus. Mit ihrer Hilfe dringen Wissenschaftler bis zu den molekularen Abläufen in der Körperzelle vor, der Grundeinheit von lebenden Organismen – in ein komplexes Materie, Energie und Information verarbeitendes System, in dem molekulare Prozesse auf vielen verschiedenen Ebenen zusammenwirken. Das Genom speichert den Bauplan der Zelle und den Ablaufplan ihrer Stoffwechselprozesse. Um diese Zellprozesse zu unterhalten, müssen immer wieder Teile des Genoms „abgelesen" werden, so etwa die Gene. Sie enthalten die Baupläne von Proteinen, der zellulären Molekular-Maschinen. Das Ablesen der Gene wiederum wird durch komplexe molekulare Netzwerke gesteuert. Für die Synthese von Proteinen und auch für ihren Abbau gibt es spezielle molekulare Komplexe, die selbst wieder detaillierter molekularer Steuerung unterliegen. Die Zelle wandelt Energie um, sie kommuniziert mit Zellen in ihrer Umgebung, sie nimmt unterschiedliche Strukturen an und bewegt sich. Sie reagiert auf Änderungen in ihrer Umgebung, zum Beispiel auf Veränderungen des Lichts, der Temperatur und des pH-Werts. Und sie wehrt Eindringlinge ab. Fehlsteuerungen dieser Prozesse sind die molekulare Grundlage für Krankheiten. Therapien zielen darauf ab, ein verträgliches molekulares Gleichgewicht wieder herzustellen.

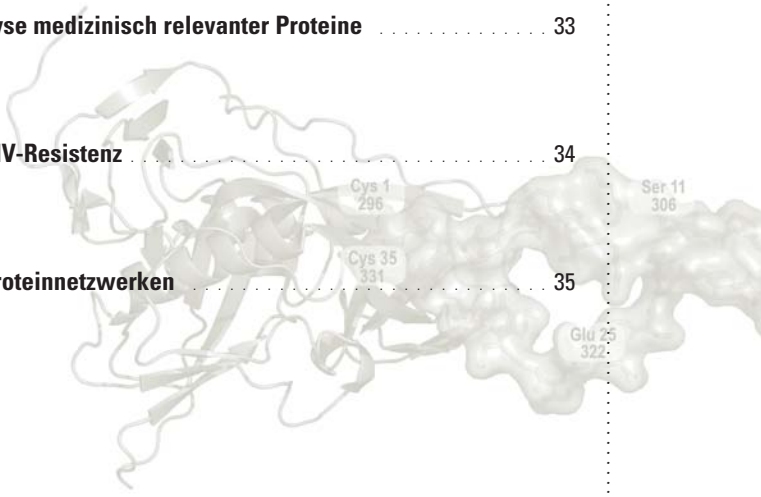
Seit gut zehn Jahren wird die klassische biologische Forschung, die bis dahin zumeist auf sehr eng eingegrenzte Teilsysteme der Zelle konzentriert war, durch Hochdurchsatz-Experimente ergänzt. Diese erfassen zellweit Daten, etwa durch eine umfassende Analyse des Genoms oder durch Messen von Häufigkeiten aller abgelesenen Gene (Transkriptom). Erfasst werden ferner die Varianten der von der Zelle verwandten Proteine (Proteom) und deren Wechselwirkungen (Interaktom). Aus diesen Daten kohärente Einsichten über die Biologie der Zelle, die Grundlagen von Krankheiten sowie Ansätze für Therapien abzuleiten, ist eine hoch komplexe informationstechnische Aufgabe. Dieser stellt sich die Bioinformatik. Am Max-Planck-Institut für Informatik wird in vielen der hier angesprochenen Bereiche geforscht. So erarbeiten die Experten am Institut beispielsweise neue Wege zur Berechnung von auf den einzelnen Patienten abgestimmten, optimalen Wirkstoffkombinationen – beispielsweise zur Behandlung von AIDS (siehe *„Analyse von HIV-Resistenz"*, Seite 34).

Damit hat die Bioinformatik den hybriden Charakter einer Grundlagenwissenschaft, die frühzeitig klare Anwendungsperspektiven definiert. Diese einzigartige Eigenschaft wird durch eine beträchtliche Zahl von Ausgründungen aus bioinformatischen Forschungsgruppen unterstrichen. So hat beispielsweise Professor Lengauer mit seinen Mitarbeitern die Firma BioSolvelT GmbH gegründet, die Software für den Entwurf von Medikamenten entwickelt und vertreibt. Zu den Nutzern dieser Software gehören weltweit über hundert Pharmafirmen.

Das von der DFG geförderte Zentrum Bioinformatik Saar, dessen Vorsitzender Professor Lengauer derzeit ist, hat unter den fünf Zentren in Deutschland bei der letzten Bewertung (2003) den ersten Platz erreicht. :::

BEITRÄGE





BIOINFORMATIK

GARANTIEN

GEOMETRIE

INTERNET

OPTIMIERUNG

SOFTWARE

STATISTISCHES LERNEN

VISUALISIERUNG

Algorithmen in der Bioinformatik

Die Bedeutung der Bioinformatik hat in den vergangenen Jahren enorm zugenommen. Durch den rasend schnellen Fortschritt in der Biotechnologie stehen Biologen immer größere Datenmengen aus Experimenten zur Verfügung. Diese können nur noch mit automatischen Methoden untersucht werden. Vor diesem Hintergrund werden leistungsstarke Algorithmen immer wichtiger.

Benötigt werden diese beispielsweise, um die biologische Funktion eines Moleküls anhand seiner Sequenz vorherzusagen. Da dies bisher noch nicht direkt möglich ist, versucht man mit Algorithmen Vergleiche anzustellen: Die Funktion eines Moleküls wird dabei durch den Vergleich seiner Sequenz mit anderen Sequenzen bestimmt, deren Funktion bereits bekannt ist.

Zum Einsatz kommen dabei verschiedene Algorithmen zur Analyse von biologischen Sequenzen. Wichtige Vertreter dieser Art Algorithmen sind zum einen Alignierungsalgorithmen, mit denen man versucht, mehrere Sequenzen bestmöglich aneinander auszurichten. Eine andere Gruppe sind Algorithmen zur Bestimmung der evolutionären Verwandtschaft.

Vereinfacht gesagt will man beispielsweise bei Alignierungsalgorithmen die Sequenzen mit einem Sonderzeichen (-) so auffüllen, dass alle Sequenzen die gleiche Länge haben und möglichst viele gleiche Buchstaben in der gleichen Position vorkommen. Ein Beispielalignment ist in Abbildung 1 zu sehen.

Abbildung 1

Ähnlich ist die Vorgehensweise beim Bestimmen der evolutionären Verwandtschaft. Man bestimmt einen so genannten Stammbaum, einen Baum dessen Blätter mit den gegebenen Sequenzen markiert sind. Haben zwei oder mehrere Sequenzen einen gemeinsamen Vorfahren, fügt man einen Knoten für diesen Vorfahren ein. Man errät seine Sequenz und verbindet ihn mit seinen Nachkommen. Dies wiederholt man so lange, bis man am gemeinsamen Vorfahren aller Sequenzen angekommen ist. Man bestimmt den Stammbaum und die Sequenzen so, dass möglichst wenig Unterschiede entlang des Baumes auftreten. In Abbildung 2 ist ein Stammbaum für die in blau geschriebenen Sequenzen zu sehen. Die Beschriftung der Kanten gibt die Anzahl der Unterschiede der zwei angrenzenden Sequenzen an. Der Stammbaum und die in rot geschriebenen Sequenzen wurden so bestimmt, dass die Gesamtzahl der Unterschiede minimal ist.

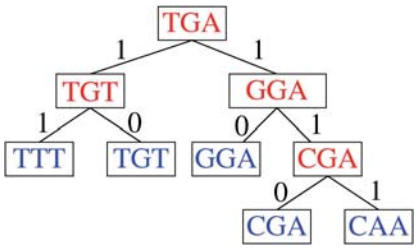


Abbildung 2

Fragestellungen wie die oben beschriebene nennt man kombinatorische Optimierungsprobleme: Die Menge aller Lösungen ist endlich, allerdings meist sehr groß. Im obigen Fall wären dies alle möglichen Stammbäume. Für jede Lösung ist ein Zielfunktionswert definiert (die Anzahl der Unterschiede entlang dieses Stammbaums). Aufgabe ist es, eine Lösung zu finden, die den kleinsten Zielfunktionswert hat – also möglichst wenig Unterschiede.

Die genannten Probleme sind allerdings, wie sehr häufig in der Bioinformatik, NP-schwer. Dabei handelt es sich um Probleme, die man eigentlich nicht effizient lösen kann.

Um die Probleme trotzdem bewältigen zu können, benutzt man entweder Heuristiken (Algorithmen, die möglichst gute, aber nicht notwendigerweise optimale Lösungen finden); oder man entwickelt ausgefeilte Optimierungsalgorithmen, um die Probleme exakt zu lösen und setzt diese in effiziente Programme um. In der Abteilung 1 des Max-Planck Instituts wird hauptsächlich an dem zuletzt genannten exakten Ansatz geforscht. Die von uns in den letzten Jahren entwickelten und implementierten Algorithmen sind die zur Zeit effizientesten Verfahren, um die zwei oben genannten Probleme (Alignierung und evolutionäre Verwandtschaft) für die jeweils betrachteten Zielfunktionen exakt zu lösen. :::

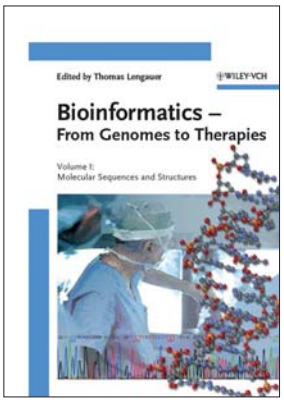
KONTAKT

Ernst Althaus
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-108
Email althaus@mpi-inf.mpg.de

Bioinformatik – Vom Genom zur Therapie

Bioinformatik ist ein sich rasch entwickelndes wissenschaftliches Gebiet. Neue experimentelle Hochdurchsatz-Technologien zur Generierung von biologischen Daten entstehen und reifen in kurzer Folge. Entsprechende Datensätze werden in großer Menge verfügbar. Die Anforderungen der biologischen und medizinischen Praxis an die Bioinformatik wachsen im gleichen Maßstab. Von der Bioinformatik verspricht man sich grundlegende Einsichten in die Entstehung von Krankheiten sowie neue Ansätze zu spezifischerer Diagnose und Prognose und letztlich neue effektivere Therapien. Als stark interdisziplinäres Gebiet berührt die Bioinformatik viele Forschungsbereiche und ist für viele Forscher und Anwender relevant. Um einen Überblick über den gegenwärtigen Stand des Gebietes zu erhalten, bedarf es ausgereifter Übersichtsliteratur, die einen breiten Kreis von Fachlesern anspricht.

In einem zweieinhalbjährigen Projekt hat Prof. Lengauer jetzt ein Referenzbuch mit dem Titel „Bioinformatics – From Genomes to Therapies" herausgegeben, das sowohl die grundlagenorientierten Methoden der Bioinformatik als auch ihre pharmazeutischen und medizinisch ausgerichteten Aspekte behandelt. In 45 Kapiteln und auf über 1700 Seiten geben 90 Autoren, die zu den internationalen Spitzenforschern des Gebietes zählen, detailliert Auskunft über den Stand der Methodik, deren Anwendungsrelevanz, sowie existierende Softwareangebote, die zum großen Teil im Internet frei verfügbar sind. Die Kapitel sind in drei Bände aufgeteilt, die dem Informationsfluss vom Genotyp zum Phänotyp entsprechen. Band (1) behandelt die molekularen Bausteine (DNS, RNS und Proteine). Band (2) befasst sich mit deren Wechselwirkungen sowie den von ihnen geformten komplexen biochemischen Netzwerken. Band (3) schließlich erläutert die biologische Funktion, die auf der Grundlage dieser Netzwerke realisiert wird, ihre Aberration in Krankheiten sowie aus dem gewonnenen molekularen Verständnis des Krankheitsprozesses resultierende Diagnose- und Therapieansätze. Der dritte Band enthält ferner Kapitel über die in der Bioinformatik verwendeten Methoden der Informationstechnologie – Datenbanken und deren Integration, webbasierte Dienste sowie Visualisierungsmethoden.
Das Buch ist eine komplette Überarbeitung und wesentliche Erweiterung des Buches „Bioinformatics – From Genomes to Drugs", das im Jahr 2002 erschienen ist. Beide Bücher sind bei Wiley-VCH erschienen.



Inhaltsangabe

Volume 1:

Molecular Sequences and Structures
Part 1 Introduction (1 Kapitel)
Part 2 Sequencing Genomes (1 Kapitel)
Part 3 Sequence Analysis (6 Kapitel)
Part 4 Molecular Structure Prediction (7 Kapitel)

Volume 2:

Molecular Interactions
Part 5 Analysis of Molecular Interactions (4 Kapitel)
Part 6 Molecular Networks (4 Kapitel)
Part 7 Analysis of Expression Data (5 Kapitel)

Volume 3:

Molecular Function
Part 8 Protein Function Prediction (8 Kapitel)
Part 9 Comparative Genomics and Evolution of Genomes (5 Kapitel)
Part 10 Basic Bioinformatics Technologies (3 Kapitel)
Part 11 Outlook (1 Kapitel) :::

KONTAKT

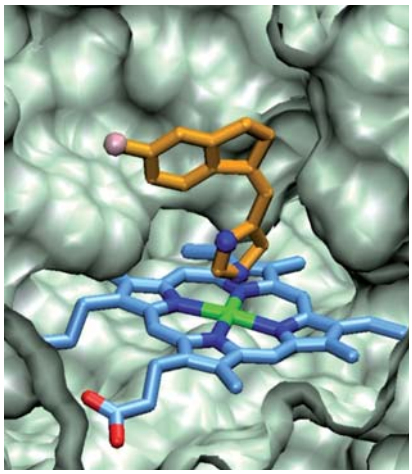
Thomas Lengauer
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email lengauer@mpi-inf.mpg.de

Docking und Wirkstoffdesign

Die Suche nach neuen Arzneistoffen hat sich in den vergangenen Jahren grundlegend gewandelt. Das traditionelle Konzept bestand darin, bereits bekannte Wirkstoffe zu verändern und heuristisch zu bewerten. Mittlerweile hat sich das strukturbasierte, rationale Design durchgesetzt. Das Hauptziel dieses Verfahrens ist, gezielt neue Substanzen zu entdekken. Die Strategie des *„Rationales Designs"* besteht darin, die dreidimensionale Struktur eines Zielmoleküls im Körper zu analysieren und Arzneistoffe zu entwickeln, die passgenau daran binden. Diese Zielmoleküle sind zumeist Proteine. Arzneistoffe binden an ihr Zielprotein nach dem *„Schlüssel-Schloss-Prinzip"*, wobei das Protein das Schloss und der Wirkstoff den Schlüssel darstellen. Wie ein Schlüssel spezifisch nur in jenes Schloss passt, für das er angefertigt wurde, sollte auch ein Arzneistoff nur an jenes Protein binden, für das er entwickelt wurde. Diese Spezifität ist von zentraler Bedeutung, da die Bindung des Arzneistoffes an das Protein dessen Funktion hemmt. Wäre der Arzneistoff nicht spezifisch, würde er auch lebensnotwendige Funktionen im Körper hemmen. Dass eine hundertprozentige Spezifität jedoch nicht immer möglich ist, lässt sich leicht an den Nebenwirkungen etablierter Arzneimittel erkennen. In sogenannten *„Screening"*-Ansätzen werden Hunderttausende von potenziellen Wirkstoffen auf ihre Bindungseigenschaften gegenüber einem Protein getestet, mit der Hoffnung einen passenden, sehr spezifischen zu finden.

Am Anfang jedes Wirkstoffdesigns steht die Auswahl eines geeigneten Zielproteins. Die Suche nach dem richtigen Kandidaten ist sehr schwierig, da an jeder Krankheit eine Vielzahl von Proteinen beteiligt ist. Zudem müssen weitere Faktoren berücksichtigt werden. So muss man beispielsweise ausschließen können, dass das Zielprotein nicht für lebensnotwendige Vorgänge zuständig ist. Diese würden sonst ebenfalls durch den Arzneistoff gehemmt. Daher sind viel Erfahrung, Wissen und Intuition nötig, um das richtige Zielprotein zu finden. Hat man das Zielprotein ausgewählt, beginnt das eigentliche Wirkstoffdesign. Zuerst wird die Stelle im Protein identifiziert, an die der Arzneistoff binden soll. Die dreidimensionale Struktur dieser Bindetasche wird analysiert. Dann versucht man, auf der Basis dieser Analyse einen Wirkstoff zu entwerfen, der optimal in diese Tasche passt. Dieser Vorgang findet inzwischen hauptsächlich über *„computergestütztes Wirkstoffdesign"* statt. Auf Grundlage der so erhaltenen Informationen werden anschließend während des so genannten *„molekularen Dockings"* mit Hilfe von speziellen Computeralgorithmen dreidimensionale virtuelle Modelle potenzieller Arzneistoffe in der Bindetasche platziert. Die Struktur des Arzneistoff-Protein-Komplexes wird anschließend danach bewertet, wie gut der Arzneistoff in das Protein passt. Jene Stoffe, die am besten passen, werden schließlich im Labor hergestellt und auf ihre biologische Wirksamkeit hin getestet.

In unserer Gruppe werden zu diesem Zweck Computerprogramme in drei Bereichen entwickelt. Ziel des ersten Bereiches ist es, dreidimensionale Modelle der Zielproteine zu erstellen und so weit zu verbessern, das sie tatsächlich ein effizientes Docking ermöglichen (Programme DynaDock und IRECS). In einem zweiten Projekt werden neue Methoden entwickelt, um das Bindungsverhalten von Peptiden (kurze Teilstücke bestimmter Proteine) an so genannte MHC-Rezeptorproteine vorherzusagen (Programm DynaPred). Peptid-MHC-Wechselwirkungen spielen eine wichtige Rolle während der Erkennung von schädlichen Viren und Bakterien durch das menschliche Immunsystem. Ihr Verständnis ist deshalb von entscheidender Bedeutung für die Entwicklung von Impfstoffen. Im dritten Bereich werden Verfahren entwickelt,



Bindetasche der Aldosteronsynthase, in blau ist der Hem-cofaktor zu sehen, an welchen der Wirkstoff bindet. Der gedockte Wirkstoff selbst ist in ocker dargestellt.

die halb-automatisch Bewertungsfunktionen und Kriterien für die Güte von Protein-Wirkstoff-Wechselwirkungen erstellen (Programm POEM).
Die so entwickelten Methoden finden in Projekten praktische Anwendung, die in enger Zusammenarbeit mit mehreren experimentellen Forschergruppen aus der pharmazeutischen Chemie, Biotechnologie und Medizin durchgeführt werden. So kooperieren wir beispielsweise mit dem Lehrstuhl für Pharmazeutische und Medizinische Chemie an der Universität des Saarlandes, Prof. Hartmann, bei der Suche nach einem Wirkstoff zur Hemmung des Proteins Aldosteronsynthase. Aldosteronsynthase ist für die Synthese von Aldosteron im menschlichen Körper zuständig. Eine Überproduktion von Aldosteron führt zu Herz-Kreislauf-Erkrankungen wie zum Beispiel Bluthochdruck. Durch Hemmung der Aldosteronsynthase kann dies verhindert werden. Erste Wirkstoffe wurden bereits identifiziert und patentiert [siehe Abbildung]. :::

KONTAKT

Iris Antes
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-319
Email antes@mpi-inf.mpg.de

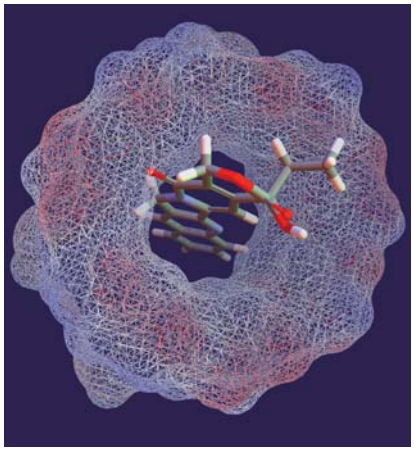
Christoph Hartmann
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-328
Email hartmann@mpi-inf.mpg.de

Chemieinformatik

Die Chemieinformatik beschäftigt sich mit der Berechnung molekularer Eigenschaften von chemischen Substanzen. Methodisch ist sie eng mit der Bioinformatik verwandt. Die Chemieinformatik beschränkt sich jedoch nicht auf biologische Moleküle, sondern umfasst darüber hinaus auch den sehr viel größeren Bereich synthetischer chemischer Strukturen. Diese große molekulare Vielfalt zu handhaben, ist eine wesentliche Herausforderung der Chemieinformatik. Die Chemieinformatik ist ein interdisziplinäres Arbeitsgebiet, das vor allem von der Pharmaforschung genutzt wird. Darüber hinaus ist sie für die Biosensorik und die Materialwissenschaften von Bedeutung. In vielen Fällen kann sie die Bioinformatik ergänzen, die der eigentliche thematische Schwerpunkt unserer Arbeitsgruppe ist. Beide Gebiete greifen insbesondere im Bereich der Wirkstoffentwicklung direkt ineinander: Bioinformatische Methoden werden hier beispielsweise zur Identifizierung von Proteinen eingesetzt, die für einen Krankheitsverlauf bedeutsam sind. Darauf baut die Chemieinformatik auf, indem sie neue Arzneistoffe identifiziert und optimiert, die an diesen Proteinen wirken.

Ein weiterer chemieinformatischer Forschungsschwerpunkt der Abteilung 3 ist die Entwicklung von Methoden zur Optimierung supramolekularer Systeme. Die diesen Systemen zugrunde liegende supramolekulare Chemie ist ein äußerst spannendes Teilgebiet der Chemie, das sich insbesondere mit der Assoziation von Molekülen zu übergeordneten supramolekularen Systemen auseinandersetzt. Oftmals bestehen solche Systeme aus einem Wirtmolekül und einem Gastmolekül. Das Wirtmolekül ist ein im Labor synthetisiertes Molekül. Wie Proteine auch ist es in der Lage Gastmoleküle zu binden. Supramolekulare Systeme werden unter anderem im Bereich der chemischen Analytik, in der Reaktionssteuerung oder auch in der Nanotechnologie in Form molekularer Maschinen eingesetzt.

In einem von der Deutschen Forschungsgemeinschaft (DFG) geförderten Projekt entwickeln wir Programme und Methoden, mit denen Strukturvorschläge für supramolekulare Komplexe berechnet und Aussagen über deren Bindungsenergien gemacht werden können. Die hier eingesetzte Algorithmik orientiert sich am Programm FlexX. Dieses wurde in den neunziger Jahren unter der Leitung von Prof. Lengauer von einer Arbeitsgruppe am GMD Forschungszentrum Informationstechnik in Sankt Augustin entwickelt. FlexX beschäftigt sich mit dem grundsätzlichen Einplatzierungsproblem eines flexiblen Gastmoleküls, in der Regel eines medizinischen Wirkstoffs in eine Proteinbindetasche (Wirtmolekül). Dabei macht das Programm die Annahme, dass sich die dreidimensionale Struktur des Proteins bei der Bindung an das Gastmolekül nicht verändert. Der Wirkstoff kann hingegen hochflexibel sein. Dieses Modell haben wir inzwischen so erweitert, dass auch das synthetische Wirtsmolekül als beweglich betrachtet werden kann. Unsere Programme können eingesetzt werden, um beispielsweise große virtuelle Bibliotheken von Wirtmolekülen zu durchsuchen und vorauszusagen, welche davon am besten zu einem gegebenen Gastmolekül passen. In Kooperation mit Forschern von der Ludwig Maximilians Universität in München und der Universität des Saarlandes in Saarbrücken wurde zum Beispiel ein synthetisches Wirtmolekül gefunden und experimentell überprüft, dass sich als Arzneimitteltransportsystem für den Krebswirkstoff Camptothecin eignet. Unsere Arbeit eröffnet damit neue Möglichkeiten, maßgeschneiderte Arzneimitteltransportsysteme zu entwickeln. :::



Camptothecin im Komplex mit einem synthetischen Wirtmolekül

KONTAKT

Andreas Steffen
ABT . 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-328
Email asteffen@mpi-inf.mpg.de

Struktur-Funktionsbeziehungen bei Proteinen

Proteine verrichten eine Vielzahl von Aufgaben in Zellen. Die meisten zellulären Prozesse beruhen auf dem Zusammenspiel mehrerer Proteine. Für viele der etwa 25.000 Protein-kodierenden Gene im Menschen ist noch nichts über die Funktion der entsprechenden Proteine bekannt – also über die Rolle der Proteine in der Zelle. Darüber hinaus ist häufig auch der molekulare Mechanismus unbekannt, nach dem das jeweilige Protein seine Funktion ausübt.

Ein Protein besteht aus einer linearen Kette von Aminosäuren die sich in der Zelle zu einer charakteristischen dreidimensionalen Struktur faltet. Diese 3D-Struktur des Proteins hat direkte Auswirkungen auf seine Funktion. So bestimmt beispielsweise die Gestalt bestimmter Bereiche an der Proteinoberfläche ob und wie gut das Protein an bestimmte andere Proteine mit komplementärer Oberfläche binden kann. Neben der Interaktion mit anderen Proteinen spielt die Struktur auch für das Zusammenspiel mit kleinen Molekülen wie etwa Wirkstoffen in Medikamenten oder mit DNA eine Rolle. Die molekularen Details haben folglich einen enormen Einfluss auf die Fragen: Wer bindet an wen? Wo wird gebunden? Wie stark ist eine Bindung? Was passiert mit dem gebundenen Partner: wird er transportiert, gespeichert, chemisch modifiziert? Was ist die Rolle dieser einzelnen Funktion im größeren Kontext zellulärer Prozesse? (Siehe auch *„Docking und Wirkstoffdesign“*, Seite 30, *„Funktionsanalyse medizinisch relevanter Proteine“*, Seite 33 und *„Analyse von Proteinnetzwerken“*, Seite 35).

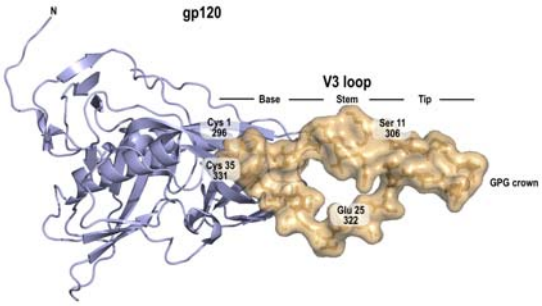


Abbildung 1: V3-Loop des HIV Oberflächenproteins gp120

Methoden und Verfahren

In der Abteilung 3: Bioinformatik des Max-Planck-Instituts für Informatik wurden mehrere Verfahren entwickelt, um folgende Fragestellungen anzugehen: (1) Charakterisierung von Protein-Protein Interaktionen zur Unterscheidung zwischen Interaktionspartnern, die sich nach verrichteter Arbeit wieder lösen und solchen, die nur gemeinsam in Form eines Komplexes auftreten. (2) Analyse von Konformationsänderungen, die beispielsweise beim Binden an einen Bindungspartner auftreten. (3) Verfahren zum Vergleich von Bindungsstellen und Interaktionsmustern in Schnittstellen zwischen zwei Proteinen, um wiederkehrende Muster und Gesetzmäßigkeiten aufzudecken (siehe auch *„Geometrische Probleme in der Bioinformatik“*, Seite 50).

Struktur-Funktionsanalyse des HIV Zelleintritts

Am Beispiel AIDS lässt sich konkret darstellen, wie eine solche Analyse von Struktur-Funktionsbeziehungen vonstatten geht. Zum Eindringen in eine menschliche Zelle benutzt der AIDS Erreger HIV eine Schleife seines Oberflächenproteins gp120. Diese so genannte V3-Loop [Abbildung 1] bindet beim Zelleintritt – und damit bei der Infektion der menschlichen Zelle – unter anderem an einen Korezeptor auf der Oberfläche der menschlichen Zelle. Grundlegend unterscheidet man die zwei Korezeptor-Typen CCR5 und CXCR4. Um den Prozess des Zelleintritts besser zu verstehen und um konkret therapeutisch eingreifen zu können, muss man wissen, welchen der beiden Korezeptoren eine Virusvariante in einem Patienten zum Zelleintritt benutzt. Eine in der Abteilung 3 entwickelte Methode ist in der Lage, basierend auf der dreidimensionalen Anordnung von physikalischen und chemischen Eigenschaften in der V3-Loop einer Virusvariante, den benutzten menschlichen

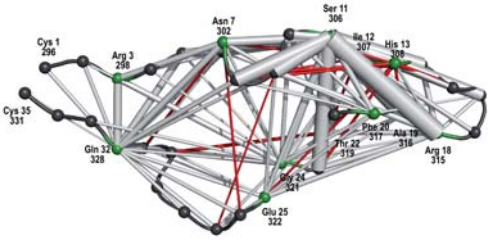


Abbildung 2: Interpretation der Wichtigkeit von Aminosäurepaaren in der V3-Loop

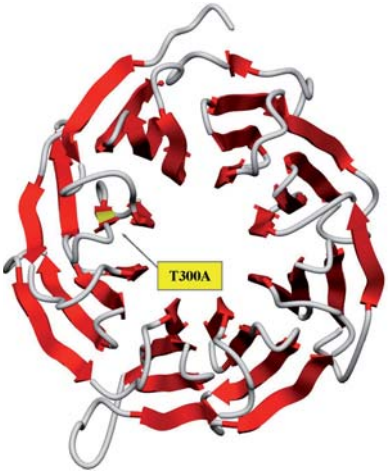
Korezeptor vorherzusagen. Darüber hinaus ist eine Aussage darüber möglich, welche Aminosäurenpaare in der Loop besonders relevant für die Auswahl des Korezeptors sind [Abbildung 2]. :::

KONTAKT

Oliver Sander
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-316
Email osander@mpi-inf.mpg.de

Francisco S. Domingues
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-304
Email doming@mpi-inf.mpg.de

Funktionsanalyse medizinisch relevanter Proteine



3D-Strukturmodell des Proteins ATG16L1. Die Position der Mutation T300A, die bei Patienten mit Morbus Crohn vorliegt, ist in Gelb markiert.

Sequenzvariation der DNA

Lebewesen unterscheiden sich grundsätzlich durch individuelle genetische Veränderungen voneinander. Am häufigsten treten so genannte *Single Nucleotide Polymorphism (SNP)* auf. Dabei handelt es sich um Punktmutationen der DNA, also Änderungen von Buchstaben im genetischen Text, die mit größerer Wahrscheinlichkeit als andere Sequenzvariationen auftreten. Populationsstudien zeigen, dass bestimmte SNPs gehäuft in einzelnen Bevölkerungsgruppen auftreten und für die Anfälligkeit gegenüber Krankheiten und die Schwere ihres Verlaufs maßgeblich sein können. Außerdem können SNPs die Verträglichkeit und Wirkung von Medikamenten beeinflussen. Die medizinische Forschung ist daher besonders an den molekularen Veränderungen interessiert, die durch SNPs verursacht werden.

Proteinstruktur

Die DNA-Sequenz eines Gens gibt vor, in welcher Reihenfolge Aminosäuren zu einem Protein zusammengesetzt werden. Die DNA-Sequenz ist gewissermaßen der Bauplan eines Proteins. Proteine haben eine charakteristische dreidimensionale Struktur, die durch atomare Wechselwirkungen innerhalb der Proteinstruktur erzeugt wird. Die meisten molekularen Prozesse laufen aufgrund von Wechselwirkungen zwischen Proteinen ab (siehe auch *„Analyse von Proteinnetzwerken“*, Seite 35). Änderungen der Proteinstruktur können diese Interaktion stören. Derartige Modifikationen können zum Beispiel durch Punktmutationen verursacht werden. So kann eine Punktmutation zum Austausch einer Aminosäure im Protein führen und bedeutende Veränderungen der intra- und intermolekularen Wechselwirkungen des Proteins zur Folge haben. Die hierdurch beeinträchtigte Funktion des Proteins und des involvierten biologischen Prozesses kann sogar Krankheiten verursachen. Beispiele sind die Sichelzellanämie und die Mukoviszidose.

Die Kenntnis der räumlichen Struktur ist ein wichtiger Beitrag zur vollständigen Analyse eines Proteins. Da die experimentelle Ermittlung der 3D-Struktur zeit- und kostenaufwendig ist, werden auch bioinformatische Methoden verwendet, um die Proteinstruktur vorherzusagen. Dazu bedient man sich der Sequenz und Struktur verwandter Proteine und erstellt daraus ein 3D-Modell.

Assoziation eines SNP mit Morbus Crohn

Während einer Kooperation zwischen der Abteilung D3: Bioinformatik und Medizinern der Universitätsklinik Kiel (Abteilung von Prof. Dr. Schreiber) konnte mit einer klinischen Assoziationsstudie ein Zusammenhang zwischen dem durch einen SNP verursachten Aminosäureaustausch im Gen *ATG16L1* und der Autoimmunerkrankung *Morbus Crohn* hergestellt werden. Morbus Crohn ist eine chronisch-entzündliche Darmerkrankung, deren molekulare Ursache noch weitgehend unbekannt ist, aber vermutlich in Verbindung mit im menschlichen Darm vorhandenen Bakterien steht. Unsere bioinformatischen Analysen gaben weiteren Aufschluss über mögliche funktionelle Veränderungen des mutierten ATG16L1-Proteins.

Proteinstrukturmodell für ATG16L1

Für unsere Kieler Kooperationspartner erstellten wir ein 3D-Modell von ATG16L1 und lokalisierten den SNP-bedingten Aminosäureaustausch T300A in der Proteinstruktur [siehe Abbildung]. Die beiden Buchstaben 'T' und 'A' bezeichnen die Punktmutation, die den Austausch der Aminosäuren Threonin gegen Alanin an der Position 300 zur Folge hat. Aus der 3D-Struktur des Proteins und ihrer Ähnlichkeit zu anderen Proteinen konnten wir folgern, dass die von der Mutation betroffene Region vermutlich an Interaktionen mit anderen Proteinen beteiligt ist. Eine Veränderung der Proteinstruktur durch den SNP kann diese Interaktionen stören und eine erhebliche Beeinträchtigung der zellulären Funktion des ATG16L1 Proteins zur Folge haben.

ATG16L1 ist ein essenzieller Bestandteil des Phagosoms, eines Organells in Immunzellen, das für den Abbau verschiedener Partikel, unter anderem von Bakterien, zuständig ist. Die Funktion des Phagosoms könnte durch das mutierte ATG16L1 beeinträchtigt werden. Eine sich daraus ergebende fehlerhafte Immunabwehr im Darm enthaltener Bakterien könnte ferner eine chronische Entzündung verursachen. Somit deuten die neuen Erkenntnisse aus unserer Studie auf eine bisher unerkannte mögliche Ursache für Morbus Crohn hin.

Auf diese Weise unterstützt die medizinische Bioinformatik klinische Forschungsarbeiten und hilft mit, genetische Befunde, wie hier von Patienten mit Morbus Crohn, in 3D zu interpretieren. Dadurch beschleunigen ihre computer-basierten Methoden die Aufklärung der molekularen Ursache von Erkrankungen und ermöglichen eine schnellere Medikamentenentwicklung. :::

KONTAKT

Gabriele Mayr
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email gabriele.mayr@mpi-inf.mpg.de

Mario Albrecht
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email mario.albrecht@mpi-inf.mpg.de
Internet http://medbioinf.mpi-inf.mpg.de

Analyse von HIV-Resistenzen

HIV Therapie

Seit der 25 Jahre zurück liegenden Entdeckung des Humanen Immundefizienz-Virus (HIV), das AIDS verursacht, sind Medikamente entwickelt worden, die in verschiedenen Phasen des viralen Lebenszyklus eingreifen. Man unterscheidet in vier Phasen:

1 Eintritt in die Wirtszelle
2 Übersetzen des viralen Erbguts (RNA) in DNA
3 Einbau der DNA in das Wirtsgenom
4 Fertigung neuer Viruspartikel

Zur Behandlung von HIV-Patienten setzt man gegenwärtig rund 20 Medikamente ein. Grund für die hohe Zahl von Medikamenten, ist die Tatsache, dass Viren immer wieder neue Resistenzen gegen die verabreichten Wirkstoffe entwickeln. Die Ursache dafür sind so genannte Resistenzmutationen – bleibende Veränderungen im viralen Erbgut, die den Krankheitserreger gegen das Medikament schützen. Die Behandlung von HIV-Patienten wird dadurch enorm erschwert. Problematisch ist zudem die Entstehung von Kreuzresistenzen. Dabei entwickelt ein Virus nicht nur gegen das verabreichte Medikament Resistenzen, sondern auch gegen andere Wirkstoffe, die der Patient noch nicht eingenommen hat. In der Abteilung 3: Bioinformatik des Max-Planck-Instituts wurden wesentliche Beiträge zur Analyse von HIV-Resistenzen geleistet.

Genotyp und Phänotyp

Bei der Wahl eines geeigneten Medikaments müssen Mediziner nicht nur mögliche Nebenwirkungen, sondern vor allem auch Resistenzen des Virus berücksichtigen. Es wurden deshalb klinische Verfahren entwickelt, um mögliche Resistenzen des Virus gegenüber einem Wirkstoff bereits vor Verabreichung festzustellen. Dabei kommen zwei unterschiedliche Vorgehensweisen zum Einsatz:

genotypische Analyse: die RNA-Sequenz des Virus wird bestimmt, und mögliche Resistenzen werden mit Hilfe von Regeln, die von medizinischen Experten erstellt und fortgeschrieben werden, ermittelt.

phänotypische Analyse: der Phänotyp des Virus (die Resistenz gegenüber einem Medikament) wird im Laborversuch gemessen und ergibt einen leicht interpretierbaren Wert pro Medikament: den Resistenzfaktor.

Die genotypische Analyse ist deutlich kostengünstiger als die phänotypische. Ihr Nachteil aber ist die mangelnde Interpretierbarkeit, denn bei der Betrachtung der Sequenz werden nur bekannte Mutationen untersucht. Dabei wird die Tatsache, dass sich Mutationen gegenseitig beeinflussen, häufig außer Acht gelassen. Am Max-Planck-Institut für Informatik wurde mit der Software geno2pheno ein Verfahren entwickelt, das die Vorteile beider Analyse-Methoden vereint. Auf einer Datenbank, die einander zugeordnete genotypische und phänotypische Messdaten enthält, wurden mit Hilfe von Methoden des Maschinellen Lernens Modelle erstellt, die anhand des viralen Erbguts die Resistenz des Virus berechnen. Die geno2pheno-Werkzeuge liefern somit die Information des teuren Verfahrens *(Phänotypisierung)* mit den Kosten des billigen Verfahrens *(Genotypisierung)*. Die berechnete Resistenz kann genau wie der durch teure Phänotyp-Analyse im Labor bestimmte Wert interpretiert werden [Abbildung 1].

Therapieauswahl

HIV-Medikamente werden heute nur noch in Kombination verabreicht, damit die Entwicklung von resistenten Virus-Varianten erschwert wird. Um behandelnde Ärzte bei der Auswahl einer wirksamen und lange anhaltenden Kombinationstherapie zu unterstützen, wird am Max-Planck-Institut für Informatik ein weiteres bioinformatisches Hilfsmittel entwickelt. Auf Grundlage einer Datenbank, die alle relevanten Informationen einer HIV-Therapie speichert, ist ein statistisches Modell erstellt worden,

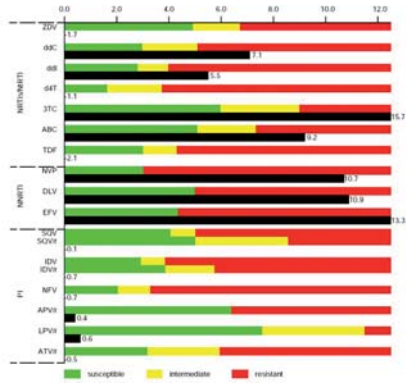


Abbildung 1: geno2pheno[resistance] schätzt auf Basis des Virusgenoms für alle gegenwärtig verwendeten Medikamente den Resistenzfaktor. Liegt ein Medikament im grünen Bereich, so ist seine Wirksamkeit unbeeinträchtigt, bei Rot liegt vermutlich eine Resistenz vor. Eine Anwendung ist nicht ratsam.

das nach Eingabe des viralen Genoms und der gewählten Medikation die Erfolgswahrscheinlichkeit der Therapie berechnet. Die Software THEO *(THErapy Optimizer)* verwendet dieses Modell und ordnet alle in Betracht kommenden Medikamentenkombinationen nach ihrer Wirksamkeit in eine Rangliste ein [Abbildung 2]. Zusätzlich zum detaillierten Wissen über den Patienten erleichtert diese Information dem behandelnden Arzt die Wahl einer optimalen Therapie. :::

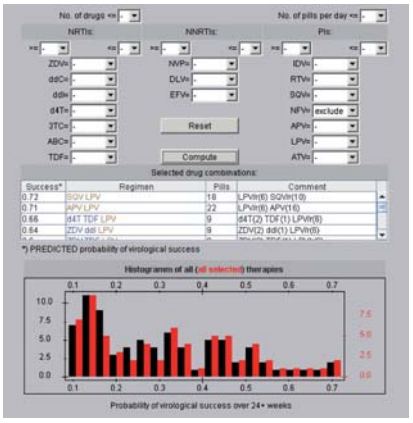


Abbildung 2: Bewertung verschiedener Kombinationstherapien durch die Software THEO

KONTAKT

André Altmann
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-308
Email altmann@mpi-inf.mpg.de
Internet http://www.geno2pheno.org

Analyse von Proteinnetzwerken

Proteininteraktionen in Zellen

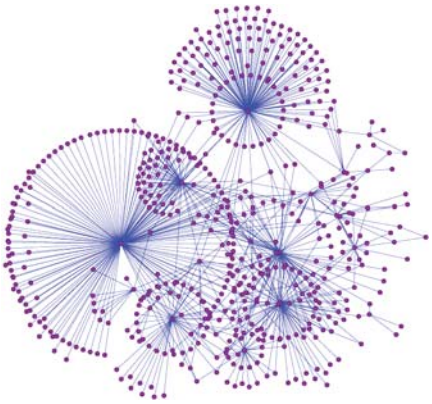
Proteine sind an nahezu allen Lebensvorgängen in der Zelle beteiligt. Oft interagieren sie miteinander, um ihre biologischen Funktionen zu erfüllen. Sie bilden dabei molekulare Maschinen, die beispielsweise Stoffe transportieren, biochemische Vorgänge beschleunigen oder Krankheitserreger abwehren. Um die entsprechenden biologischen Prozesse in den Zellen zu verstehen, ist es notwendig, die verschiedenen molekularen Interaktionen zwischen den Proteinen zu kennen. Die Gesamtheit aller Proteininteraktionen eines Organismus wird als Interaktom bezeichnet. Es wird geschätzt, dass es etwa 25.000 menschliche Gene gibt. Diese liefern die Information für den Bau von mehreren Zehntausend Proteinen, die in der Summe zwischen 200.000 und 300.000 Interaktionen eingehen [siehe Abbildung].

Störungen von Interaktionen zwischen Proteinen und die damit verbundene Beeinträchtigung von zellulären Vorgängen können zum Auftreten verschiedener Krankheiten führen (siehe auch *„Funktionsanalyse medizinisch relevanter Proteine“*, Seite 33). Fachleute hoffen, dass die Aufklärung aller Proteininteraktionen und die nähere Analyse des menschlichen Interaktoms neue Ansätze zur Behandlung von Erkrankungen liefern. Ein Beispiel dafür sind Infektionen mit Viren wie zum Beispiel dem AIDS verursachenden HI-Virus oder dem Hepatitis C Virus. Hier ist es von besonderer Bedeutung, Interaktionen zwischen viralen und menschlichen Proteinen zu identifizieren, um die molekularen Mechanismen der Infektion besser zu verstehen. Dieses Verständnis ist letztendlich nötig, um bessere Therapien und neue Medikamente gegen Viren entwikkeln zu können.

Analyse von Interaktionen menschlicher Proteine

Mittlerweile gibt es verschiedene experimentelle Techniken, um Interaktionen zwischen Proteinen im Labor zu bestimmen. Da diese Experimente aber sehr zeit- und kostenintensiv sind, werden auch rechnerische Methoden entwickelt, um Proteininteraktionen vorherzusagen. Noch sind aber weder die experimentellen noch die computergestützten Methoden ganz ausgereift. Zum einen werden viele Proteininteraktionen noch nicht entdeckt. Zum anderen sind einige Interaktionen Artefakte, die in der Realität nicht vorkommen. Aus diesem Grund ist eine genaue Analyse der vorhandenen Interaktionsdaten und ihrer Verlässlichkeit nötig, bevor diese in der biomedizinischen Forschung verwendet werden können.

In einer umfangreichen Vergleichsstudie haben wir mehrere weltweit verfügbare Datensätze von Interaktionen zwischen menschlichen Proteinen untersucht, die mit unterschiedlichen Methoden vorhergesagt oder experimentell bestimmt worden waren. Unsere Analyse zeigt, dass viele Interaktionen nur jeweils in einem Datensatz vorhanden sind. Das führen wir darauf zurück, dass die unterschiedlichen Methoden verschiedene Teilbereiche des gesamten humanen Interaktoms abdecken. In anderen Fällen wurden bestimmte Interaktionen von mehreren Methoden zugleich vorhergesagt. Das macht die einzelne Interaktion glaubwürdiger. Der Anteil dieser zuverlässigen Interaktionen an der gesamten bis jetzt vorliegenden Datenmenge von Proteininteraktionen ist allerdings noch gering.



Beispiel eines molekularen Netzwerkes aus Interaktionen menschlicher Proteine. Violette Kreise repräsentieren Proteine, deren Interaktionen durch blaue Linien gekennzeichnet sind.

Ein weiterer Schritt bei der Untersuchung von Proteininteraktionen ist die Analyse der molekularen Funktion der beteiligten Proteine. Meist dienen Interaktionen zwischen Proteinen der Ausführung einer gemeinsamen Aufgabe. Anhand eines von uns entwickelten funktionellen Ähnlichkeitsmaßes erkennt man bei Interaktionen mit bekannter hoher Verlässlichkeit, dass die beteiligten Proteine auch eine hohe funktionelle Ähnlichkeit aufweisen. Anhand bereits bekannter Funktionen von Proteinen kann man so im Umkehrschluss überprüfen, ob eine gefundene Interaktion zwischen zwei Proteinen funktionell sinnvoll ist und sie vermutlich tatsächlich in der lebenden Zelle stattfindet – oder aber, ob die Interaktion auf einen technischen Fehler der angewandten Methode zurückzuführen ist. Unser Ähnlichkeitsmaß macht es also möglich, die verfügbaren Datensätze von Proteininteraktionen zu filtern, um nur wirklich zuverlässige Interaktionen in nachfolgenden biologischen und medizinischen Untersuchungen weiter zu verwenden. :::

KONTAKT

Andreas Schlicker
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email andreas.schlicker@mpi-inf.mpg.de

Mario Albrecht
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email mario.albrecht@mpi-inf.mpg.de
Internet http://medbioinf.mpi-inf.mpg.de

GARANTIEN

Software soll verlässlich sein. Das wichtigste Kriterium für Verlässlichkeit ist die Korrektheit. Fast genauso wichtig aber ist oft die Performanz: Eine korrekte Antwort, die man nicht rechtzeitig bekommt, ist unnütz. Die Suche nach Korrektheits- und Performanzgarantien gehört für viele Abteilungen des Instituts zu den zentralen Fragestellungen.

Computer, Netzwerke und mikroprozessorgesteuerte Systeme sind heute ein allgegenwärtiger Teil unseres Lebens. Wir benutzen sie ständig – teils bewusst, wie den Rechner auf dem Schreibtisch, die Internet-Suchmaschine oder das Handy, teils unbewusst, wie die elektronische Steuerung im Auto, im Flugzeug oder in der Waschmaschine. Und je mehr wir unser Leben abhängig von Hard- und Software machen, umso mehr stellt sich die Frage, ob das Vertrauen, das wir in diese Produkte setzen, gerechtfertig ist. Können wir garantieren, dass eine Hardware, eine Software oder ein eingebettetes System, das mit seiner Umgebung interagiert, wie gewünscht funktioniert? Diese Frage durchzieht die Arbeiten mehrerer Abteilungen des Instituts.

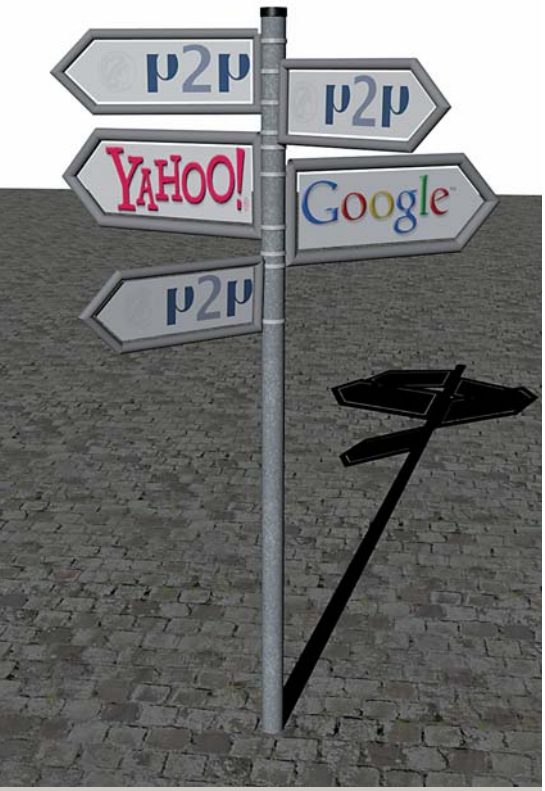
Die einleuchtende Forderung, die man gewöhnlich an eine Hard- oder Software stellt, ist Korrektheit. Wir erwarten, dass eine E-Mail nur an den angegebenen Adressaten zugestellt wird, dass ein Routenplaner uns tatsächlich zum gewünschten Ziel bringt, dass eine Eisenbahnsteuerung ein Gleis erst freigibt, wenn eine Schranke geschlossen ist. Um solche Eigenschaften nachzuweisen, benötigen wir deduktive Systeme, die überprüfen, ob eine Eigenschaft aus anderen, bereits bekannten Eigenschaften folgt. Ein erster Schritt ist dabei die Modularisierung: die (möglichst automatische) Zerlegung eines großen Problems in kleine, handhabbare Teilprobleme. Ein zweites wichtiges Hilfsmittel ist die Abstraktion. Vereinfacht ausgedrückt versucht man dabei, die unendlich vielen möglichen Daten, mit denen ein Programm arbeiten soll, in endlich viele Klassen zu gruppieren, deren Elemente sich im Wesentlichen gleich verhalten.

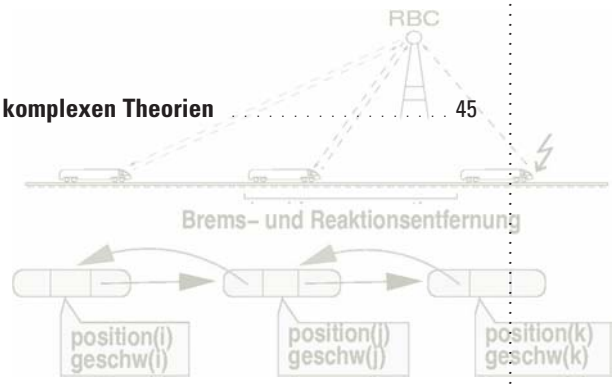
Performanzgarantien sind unter Umständen ebenso wichtig wie Korrektheitsgarantien: Eine Suchmaschine, die Resultate erst nach Tagen liefert, ist für den Benutzer inakzeptabel; eine Flugzeugklappensteuerung, die für die Berechnung der erforderlichen Landeklappeneinstellung länger braucht als vorgesehen, ist lebensgefährlich. Neben dem Verbrauch von Rechenzeit interessiert man sich hier auch für den Verbrauch von Speicherplatz, den Kommunikationsbedarf oder den Energieverbrauch.

Bei Such- und Optimierungsproblemen stellt man fest, dass die Forderungen nach optimalen Ergebnissen und effizienter Berechnung oft nicht gleichzeitig zu erfüllen sind. Manchmal liegt dies an der schieren Größe der zu verarbeitenden Datenmenge, beispielsweise bei der Beantwortung einer komplexen Suchmaschinenanfrage. In anderen Fällen, etwa bei graphentheoretischen Fragestellungen, zeigt sich, dass selbst Probleme von durchaus überschaubarer Größe nicht in akzeptabler Zeit exakt algorithmisch behandelt werden können. Ein Ausweg besteht hier in der Entwicklung approximativer Verfahren. Solche Verfahren liefern Lösungen, die zwar in der Regel nicht optimal sind, aber deren Qualität von der optimalen Lösung höchstens um einen festgelegten Wert abweicht. Dieser Weg ist oftmals viel schneller als eine exakte Methode. So lässt sich beispielsweise die Beantwortung einer Suchmaschinenanfrage erheblich beschleunigen, wenn man in Kauf nimmt, unter Umständen 10 bis 20 Prozent der besten Treffer zu verpassen.

Korrektheits- und Performanzgarantien sind von zentraler Bedeutung für einen großen Teil der Arbeiten im Institut. Die folgenden Beispiele kommen aus vier Abteilungen. :::

BEITRÄGE

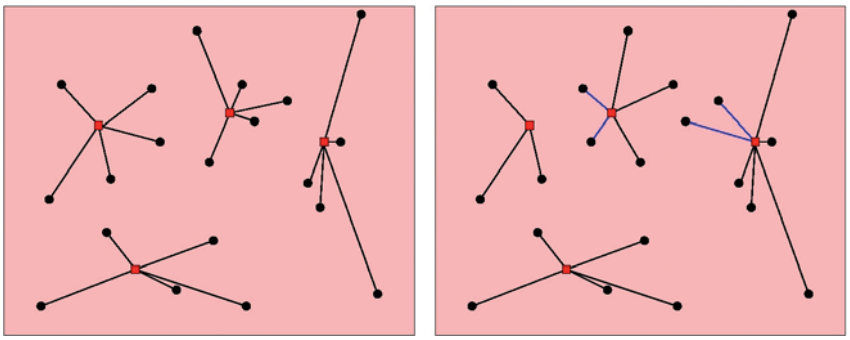




Approximationsalgorithmen

Wo auch immer ein Dienstleistungsunternehmen eine neue Filiale eröffnen will, muss es sichergehen, dass sich der neue Standort lohnt. Dass sich eine solche Suche durch informatische Methoden – insbesondere durch die Lösung so genannter Facility-Location-Probleme – erleichtern lässt, zeigt folgendes Fallbeispiel: Das Management eines großen Lebensmitteldiscounters plant, sein Filialnetz auszubauen. Eine Marktstudie hat ergeben, dass die meisten Kunden nicht bereit sind, mehr als zwei Kilometer zu einer Filiale zurückzulegen. Die Kosten der Neueröffnung einer Filiale sind bekannt, hängen jedoch vom jeweiligen Standort ab. Das Management hat ein Budget von 100 Millionen Euro für den Ausbaus des Filialnetzes in Aussicht gestellt, das möglichst optimal genutzt werden muss. Das bedeutet also, dass neue Filialstandorte so geplant werden müssen, dass die Kundenabdeckung maximiert wird. Diese und viele ähnliche Aspekte gehören in die Kategorie der **Facility Location** Probleme. Diese stehen schon seit Jahren im Brennpunkt vieler Forschungsarbeiten aus der Informatik und der Unternehmensforschung.

Die meisten Facility-Location-Probleme sind **NP-schwer**. Das heißt, dass die Generierung einer exakten Lösung selbst für moderate Problemgrößen (zum Beispiel 50 potenzielle Filialstandorte) mehrere Jahre dauern würde. Eine Alternative sind so genannte **Approximationsalgorithmen**. Dabei handelt es sich um schnelle Verfahren, die keine optimale Lösung anstreben, sondern Ergebnisse, die beweisbar nahe am Optimum liegen. Damit lassen sich Problemlösungen in einer vertretbaren Zeit finden. In den letzten 20 Jahren wurden große Fortschritte auf dem Gebiet der Approximationsalgorithmen für verschiedenste Probleme erzielt. Ebenso wurden fundamentale Schranken bezüglich der (Nicht-)Approximierbarkeit bewiesen.



Die einem Klienten nächstgelegene Facility kann sich ändern, wenn eine Facility ersetzt wird, ebenso die Summe der Distanzen der Klienten zu den jeweils nächstgelegenen Facilities.

Ein klassisches Facility-Location-Problem ist das **k-Median-Problem**. Ziel ist es hier, k Facilities *(Filialen)* so zu platzieren, dass die Summe der Distanzen der Klienten *(Kunden)* zu den jeweils nächsten Facilities minimiert wird. Eine naheliegende Heuristik für dieses Problem basiert auf der so genannten **lokalen Suche**. Ein solcher Algorithmus beginnt mit k beliebig gewählten Facilities und versucht dann durch Ersetzen einer Facility zu einer besseren Lösung zu gelangen. Dieser Ersetz-Schritt wird wiederholt, bis keine lokale Verbesserung mehr möglich ist.

Eine Analyse dieser Heuristik zeigt, dass die berechnete Lösung maximal um Faktor 5 schlechter als die optimale Lösung ist. Falls man erlaubt, zwei Facilities gleichzeitig zu ersetzen, verbessert sich die Qualitätsgarantie auf Faktor 4. Beim Ersatz von p Facilities ergibt sich sogar eine Garantie von (3+2/p). Allerdings erhöht sich die Laufzeit des Algorithmus stark mit wachsendem p.

Ein alternativer Lösungsansatz für das Facility-Location-Problem ist die lineare Programmierung. Das Problem wird hierbei als ganzzahliges lineares Programm formuliert. Das korrespondierende lineare Programm (LP) wird dabei nach Relaxierung der Ganzzahligkeitsbedingungen gelöst. Die Lösung für dieses LP, das durchaus nicht-ganzzahlige Einträge enthalten kann, wird dann auf spezielle Art und Weise zu einer ganzzahligen zulässigen Lösung gerundet.

Für das k-Median Problem existiert bereits ein LP-basierter Algorithmus, dessen Lösung maximal um Faktor 4 vom Optimum abweicht. In unserer Arbeitsgruppe untersuchen wir eine Variation dieses Algorithmus, der diese Schranke auf Faktor 3 verbessert. Ein solcher Algorithmus wäre viel schneller als ein auf lokaler Suche basierender Algorithmus. :::

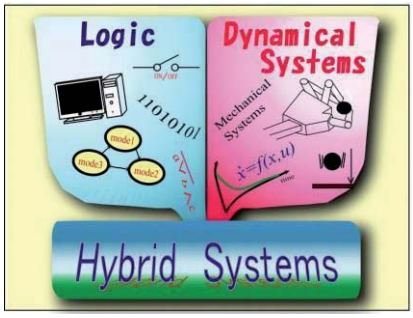
KONTAKT

Naveen Garg
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-115
Email naveen@mpi-inf.mpg.de

Automatische Verifikation von Stabilitätseigenschaften für hybride Systeme

Der Verifikation von Software- und Hardwaresystemen kommt eine immer größere Bedeutung zu – insbesondere aus industrieller Sicht. Mit *„Verifikation"* bezeichnet man den Nachweis, dass ein Software- oder Hardware-System die ihm zugedachten Eigenschaften, seine Spezifikationen, tatsächlich erfüllt. Für den Einsatz in sicherheitskritischen Anwendungen sind Verlässlichkeitsgarantien und damit formale Modelle unabdingbar. Diese formale Modellierung lässt sich mit so genannten *hybriden Systemen* durchführen. Hybride Systeme beschreiben das Zusammenwirken von kontinuierlichen (analogen) und diskreten (digitalen) Komponenten. Die kontinuierlichen Größen stammen aus der physikalischen Umwelt – beispielsweise die Temperatur in einem Raum. Es gilt, diese Größen mit diskreten Steuerungen (Hardware oder Software) so zu beeinflussen, dass das Gesamtsystem gewisse gewünschte Eigenschaften erfüllt. Ein Beispiel ist das Halten der Raumtemperatur zwischen 18 und 22 Grad Celsius. Weitere Beispiele für hybride Systeme sind Steuerungen von Verkehrssystemen wie Eisenbahnen und Flugzeuge oder die Überwachung chemischer Prozesse.

Ein wichtiger Punkt bei der Analyse sicherheitskritischer Systeme ist der Nachweis der *Stabilität* bezüglich einer Sicherheitsanforderung. Befindet sich ein System erst einmal in einem stabilen Zustand, lässt es sich nämlich so steuern, dass es ihn nicht mehr verlässt. Innerhalb dieses Zustandes ist garantiert, dass das System die Sicherheitsanforderung erfüllt, dass beispielsweise die Raumtemperatur in einem bestimmten Bereich liegt.
Um eine gewünschte Sicherheitsanforderung zu gewährleisten, muss demnach überprüft werden, ob das System in jedem Fall einen stabilen Zustand erreicht. Dazu wurde ein Verfahren entwickelt, das automatisch verifiziert, dass ein hybrides System unabhängig von seinem Startzustand (im Beispiel: unabhängig von der Anfangstemperatur des Raums) in endlicher Zeit den stabilen Bereich erreicht. Das Verfahren wurde implementiert und erfolgreich an einer Reihe von Beispielen getestet. :::



Hybride Systeme ermöglichen die formale Modellierung von Software- oder Hardwaresystemen.

Anwendung finden die neuen Kontrollmethoden z.B. bei Zugsicherungssystemen im europaweiten ETCS-Standard.

Avacs: Transregionaler Sonderforschungsbereich der Universitäten Oldenburg, Freiburg und Saarbrücken sowie des Max-Planck-Instituts für Informatik

KONTAKT

Silke Wagner
ABT . 2 Logik der Programmierung
Telefon +49 681 9325-221
Email swagner@mpi-inf.mpg.de
Internet http://www.avacs.org

Programmanalyse und -verifikation

Softwaresysteme wie zum Beispiel Web- und Mail-Server oder Datenbanksysteme werden in der Regel aus vielen verschiedenen Komponenten zusammengestellt. Der Nutzer erwartet von diesen Komponenten zurecht, dass sie zügig und exakt zum richtigen Zeitpunkt angeforderte Resultate liefern. Falls aber auch nur eine Komponente unerwartet nicht mehr funktioniert, kann es passieren, dass das gesamte System ausfällt.

Wir entwickeln Algorithmen, die verschiedene Systemfunktionen automatisch analysieren, sodass ein fehlerfreier Betrieb gewährleistet werden kann. Eine solche Analyse ist ausgesprochen schwierig, da die dafür notwendige moderne Software zu den kompliziertesten Konstruktionen zählt, die Ingenieure und Programmierer heutzutage entwickeln. Die Analyse einer Systemfunktion oder eines entsprechenden Programms ist ausgesprochen komplex. Deshalb entwickeln wir Algorithmen, die eine klare logische Analyse der Programme, der in ihnen ablaufenden Berechnungen und ihrer Funktionalitäten ermöglichen.

Die logische Analyse der Computerprogramme mithilfe unserer Algorithmen erlaubt uns, während der Analyse das richtige Abstraktionsniveau zu treffen, sodass nur wirklich relevante Programmmerkmale betrachtet werden. Der Abstraktionsprozess wird gebraucht, um leistungsfähige Analysewerkzeuge zu erschaffen, die tatsächlich anwendungstauglich sind. Es ist dabei unabdingbar, eine ausreichende Menge an Informationen über das zu analysierende Programm zu sammeln, damit die gewünschten Programmeigenschaften nachgewiesen werden können. Darüber hinaus ist es unverzichtbar, die nicht wesentlichen, im Wege stehenden Details zu eliminieren. Die Hauptherausforderung liegt dabei in der automatischen Bestimmung des passenden Abstraktionsniveaus.

Bereits seit 1970 hatte man versucht Algorithmen zu finden, mit denen sich nachweisen lässt, dass Systemkomponenten ihre Aufgaben erfüllen und dabei die angeforderten Resultate liefern. Besonders schwierig war es, so genannte Lebendigkeitseigenschaften zu prüfen – jene Eigenschaften, die dafür sorgen, dass ein Programm immer korrekt auf Anfragen reagiert – also „lebendig bleibt". Eins der größten Hindernisse bei der Überprüfung der Lebendigkeitseigenschaften war das Fehlen adäquater Abstraktionstechniken.

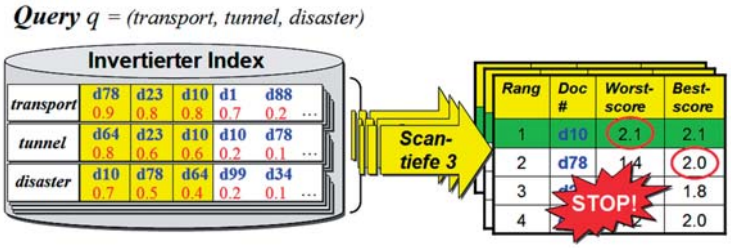
Wir stellen eine Lösung dieses Problems dar, indem wir eine neue Art von logischen Hilfsaussagen definieren, die als Transitionsinvarianten bezeichnet werden. Diese Hilfsaussagen überwinden die Einschränkungen der traditionellen Methoden zum Beweisen der Lebendigkeitseigenschaften eines Programms. Wir bieten einen Algorithmus an, der ein für die Anwendung von Transitionsinvarianten adäquates Abstraktionsniveau automatisch berechnet. Mit diesem von uns entwickelten Algorithmus ist es möglich, eine automatische Analyse der Lebendigkeitseigenschaften von großen Programmfragmenten durchzuführen.

Unsere theoretischen Beiträge sind eine Basis für die Entwicklung von praktisch einsetzbaren Verifikationswerkzeugen für die Analyse von Lebendigkeitseigenschaften. Wir haben ein Werkzeug implementiert, das erfolgreich für die Analyse von entscheidenden Fragmenten von Systemsoftware eingesetzt wurde, beispielsweise in Zusammenarbeit mit dem Forschungslabor von Microsoft und dem durch das Bundesforschungsministerium geförderten Projekt Verisoft. Zu den Hauptanwendungen gehören Routinen in Gerätetreibern und in einem Betriebsystem. Die erstellten Lebendigkeitsnachweise erhöhen unsere Zuversicht, dass die Systemkomponenten tatsächlich stets die nötige Funktionalität liefern. :::

KONTAKT

Andrey Rybalchenko
ABT . 2 Logik der Programmierung
Telefon +49 681 9325-206
Email rybal@mpi-inf.mpg.de

Wie suche ich schnell im Web?



Mit der stetig exponenziell wachsenden Größe von Informationsquellen wie dem Internet, Informationsportalen von Firmen oder auch wissenschaftlichen Datenbanken wird die automatisierte und effiziente Informationssuche von enormen Datenmengen unterschiedlichster Art auch in Zukunft weiter an Bedeutung gewinnen. Laut Moores Gesetz verdoppelt sich zwar dem allgemeinen technologischen Fortschritt zufolge die Leistungsfähigkeit von Rechnern und die Kapazität von Massenspeichern beinahe jährlich; bei genauer Betrachtung zeigt sich allerdings, dass sich andere wichtige Technologieparameter leider nicht in dem für das schnelle Datenwachstum erforderlichen, ebenfalls exponenziellen Tempo weiter entwickeln: Dies gilt insbesondere für die für die schnelle Suche kritischen Zugriffsgeschwindigkeiten von Festplatten, die über das letzte Jahrzehnt hinweg sogar beinahe stagniert haben. Neben der mit hohen Kosten verbundenen Verteilung von Suchindizes auf viele, dicht vernetzte und hochgradig redundante Rechnerstrukturen ist daher die Verbesserung von Effizienz und Skalierbarkeit der lokalen Anfrageauswertung auf jedem einzelnen dieser Rechner von tragender Bedeutung, um trotz stagnierender Zugriffsgeschwindigkeiten von Festplatten mit dem ständigen Datenwachstum auch weiterhin Schritt halten zu können.

Bei allen aktuellen Suchmaschinen basiert der Suchindex auf vorausberechneten Listen von Dokumenten zu jedem möglichen Suchbegriff, der in dem gegebenen Korpus, wie beispielsweise einem Firmennetzwerk oder sogar dem gesamten Web, von der Suchmaschine erkannt wurde. Die Indexliste eines Schlüsselworts enthält zu jedem Dokument, das das Wort enthält, einen mit raffinierten Wortstatistiken berechneten numerischen Güte- oder Relevanz-Score, der widerspiegelt, wie gut eine Seite mit dem jeweiligen Suchbegriff übereinstimmt. Wegen der enormen Anzahl unterschiedlichster Suchbegriffe und ihrer möglichen Treffer besteht eine besondere Herausforderung in der effizienten Auswertung mehrdimensionaler Anfragen, also von Anfragen mit mehreren Suchbegriffen, kurzen Phrasen oder sogar ganzen Sätzen. Eine Anfrage mit mehreren Suchbegriffen wird dabei durch paralleles Traversieren der entsprechenden Indexlisten eines jeden in der Anfrage vorkommenden Schlüsselwortes ausgewertet, wobei die Treffer für jeden einzelnen Begriff und deren Scores möglichst effizient zu einem Gesamt-Score des jeweiligen Dokumentes kombiniert werden müssen. Das Anfrageergebnis ist dann eine nach diesen Gesamt-Scores absteigend sortierte, potenziell sehr lange Rangliste von möglichen Trefferdokumenten. Wie wahrscheinlich jeder aus eigener Erfahrung bestätigen kann, ist dabei der Benutzer in der Regel jedoch nur an den 10 oder 20 Top-Treffern interessiert, so dass ein Großteil der Ergebnislisten niemals betrachtet wird.

Unsere Arbeiten haben das Ziel, bestehende Algorithmen zur effizienten Berechnung dieser besten, so genannten Top-k Treffer weiter zu beschleunigen. Dazu vermeiden wir zum einen nichtsequenzielle Plattenzugriffe, da diese teilweise um mehrere Größenordnungen langsamer als sequenzielle Zugriffe sind. Zudem nutzen wir eine ausgeklügelte Priority-Queue als zentrale Datenstruktur der Anfrageauswertung. Anstelle der unnötig teuren, vollständigen Auswertung aller möglichen Kandidaten, beschränken wir uns während des Traversierens der Indexlisten auf die Menge der Top-k-Kandidaten zusammen mit Garantien für die oberen und unteren Schranken ihres jeweils bestmöglichen und schlechtestmöglichen aggregierten Scores. Ein Kandidat kann dabei aus der Priority-Queue entfernt werden, wenn sein bestmöglicher Score nicht ausreicht, um ihn zu einem der Top-k Treffer zu machen. Auf diese Weise ist der Algorithmus in der Lage, die Anfragebearbeitung zu terminieren, ohne notwendiger Weise alle Indexlisten vollständig traversieren zu müssen, und spart so einen Großteil an teuren Plattenzugriffen.

Über diese Optimierung hinaus lässt sich das oben beschriebene konservative Abbruchkriterium mit Hilfe von Statistiken über die Verteilungen der Scores in Form eines probabilistischen aber approximativen Top-k-Algorithmus weiter verbessern. Probabilistische Abbruchkriterien führen dabei zu einer früheren und drastischen Verkleinerung der Kandidatenmenge, wodurch letztlich eine wesentlich schnellere Terminierung der Indextraversierungen erreicht wird. Unsere Experimente mit großen Datenmengen und in internationalen Benchmark-Wettbewerben, wie etwa dem TREC Terabyte Benchmark, zeigen, dass wir gegenüber den besten Verfahren mit konservativem Abbruchkriterium unsere Laufzeiten um einen weiteren Faktor von 10 bis 20 verkürzen können. Dabei büßen wir aber nur etwa 10 bis 20 Prozent an Präzision in der Top-k Ergebnisliste ein. Durch die Kombination von aggressiven Abbruchkriterien und durch geschicktes Scheduling der Indexlistenzugriffe können so häufig sogar Laufzeitverbesserungen um zwei Größenordnungen gegenüber bestehenden Top-k Verfahren erreicht werden. :::

KONTAKT

Martin Theobald
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-500
Email martin.theobald@mpi-inf.mpg.de

Informationssuche in Peer-to-Peer Systemen

Die Vorstellung, nicht nur mit dem eigenen PC nach Informationen im World-Wide-Web zu suchen, sondern dafür gleich Tausende über den Erdball verteilte Computer zu nutzen, ist verlokkend. Die Möglichkeiten für eine derartige effiziente Suche sind derzeit allerdings noch sehr beschränkt und ohne zentralen Server praktisch nicht möglich. Traditionelle Ansätze einer derartigen verteilten Informationssuche scheitern zumeist an der erwarteten Dynamik eines solchen Verbundes oder an seiner enormen Größe. So genannte Peer-to-Peer-Architekturen für verteilte Computersysteme aber könnten in naher Zukunft den Durchbruch bringen. Denn mit ihnen lässt sich das Potenzial dezentralisierter Computer- und Informationsressourcen tatsächlich effizient nutzen.

Im Laufe der vergangenen Jahre sind Peer-to-Peer-Systeme, kurz P2P-Systeme, immer populärer geworden. Dazu zählen Vertreter wie Napster, Gnutella oder BitTorrent. Diese Musiktauschbörsen haben sich durch ihre dezentralisierte Struktur einer effizienten Überwachung durch Behörden entzogen. Damit erlangte letztlich der ganze Peer-to-Peer-Bereich ungerechterweise einen zweifelhaften Ruf. Im wissenschaftlichen Sinne ist ein Peer-to-Peer-System ein Netzwerk von verteilten, autonomen Computern, die gegenüber anderen Computern jeweils sowohl als Client als auch als Server agieren können. Die Daten- und Lastverteilung und Organisation bewältigt das gesamte Netzwerk von selbst. Zu den erwarteten Stärken des Peer-to-Peer-Ansatzes gehören neben der potenziell besseren Fehlertoleranz und Datenverfügbarkeit sowie der Ausnutzung enormer, heute zumeist brachliegender Rechnerkapazitäten insbesondere die Resistenz gegen Überlast, Systemdynamik, Attacken, Manipulationen oder Zensur.

Gerade der letzte Punkt rückt zunehmend ins Interesse der Öffentlichkeit. Bereits im Juni 2004 kritisierte der Unterausschuss Neue Medien des Bundestags neben der Tatsache, dass keine Suchmaschine das Internet auch nur annähernd vollständig erfasse, insbesondere die Unterwanderung von Suchergebnissen durch getarnte Werbeeinträge sowie den fortschreitenden Monopolisierungsprozess auf diesem Sektor und forderte die Förderung freier Suchmaschinen.

Unsere Forschung greift diese Sorgen auf und verfolgt eine vollständig dezentralisierte, selbstorganisierende Architektur zur Informationssuche im World-Wide-Web. Diese beruht auf der Kommunikation zwischen einer a priori unbegrenzten Zahl von kooperierenden Computern. Dabei erweitern wir das Fundament der Informationssuche über den heutigen Stand der Technik hinaus, indem wir neben der herkömmlichen, schlüsselwortbasierten Suche das intellektuelle Potential großer Benutzergemeinden ausnutzen. So liefern beispielsweise die Bookmarks einzelner Benutzer und das in den Click-Streams der Benutzer reflektierte Feedback zu früheren Suchergebnissen die Basis für Empfehlungen für künftige Suchergebnisse. Dieses bislang unerschlossene Wissen lässt sich mit Hilfe von Data-Mining-Techniken und statistischen Lernverfahren ausschöpfen. Auch die effektive Nutzung zunehmend verbreiteter manueller Annotationen von Objekten *(Social Tagging)* wie etwa bei der populären Fotoplattform *Flickr* kann helfen, die Qualität der Suchergebnisse deutlich zu verbessern. Letztlich soll sich eine Suchmaschine mit der Zeit auf die Vorlieben des Benutzers einstellen, indem sie dessen Interaktionen beobachtet. So kann sie ihm künftig individuell zugeschnittene Ergebnisse präsentieren. Für den Suchbegriff „Java“ zum Beispiel erhielte ein Informatiker andere Ergebnisse (Hinweise auf die Programmiersprache) als ein Weltreisender (Hinweise auf die indonesische Insel).

Einer unserer aktuellen Forschungsschwerpunkte ist das so genannte Query-Routing. Bei diesem Verfahren sollen wenige, aber vielversprechende Peers, an die eine bestimmte Benutzeranfrage geleitet werden soll, anhand von kompakten Daten-Statistiken dieser verschiedenen Peers effizient ausgewählt werden. Unsere Verfahren werden in das Prototypsystem MINERVA (siehe auch *„Minerva“*, Seite 75) integriert und experimentell erprobt.

In dem gerade gestarteten EU-Projekt *SAPIR (Search in Audio Visual Content Using Peer-to-Peer Information Retrieval)* wird gemeinsam mit europäischen Partnern eine umfassende Architektur entwickelt, die die verteilte Suche nach Webseiten, Fotos, Videos und Musik in integrierter Form realisiert. :::

KONTAKT

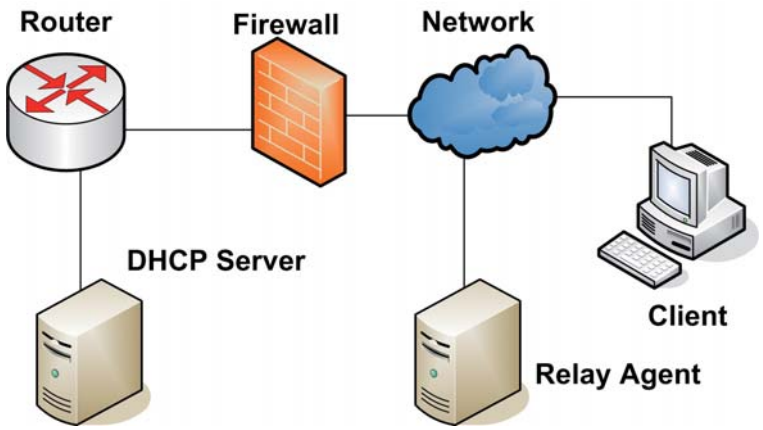
Matthias Bender
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-500
Email mbender@mpi-inf.mpg.de

Leistungsfähige Beweissysteme: SPASS, SPASS+T, Waldmeister

Um mathematisch exakt Eigenschaften von Systemen (Programmen) zu beweisen, werden die Eigenschaften erst in einer formalen Sprache (Logik) beschrieben und dann mittels computergestützter Verfahren bewiesen. Die Weiterentwicklung dieser Verfahren, d.h. die Automatisierung von Schlussfolgerungstechniken in Logiken und damit des Beweisens, ist ein zentrales Thema der aktuellen Forschung. Es ist dann aber noch einmal ein großer Schritt von den Schlussfolgerungstechniken bis hin zu praktisch funktionierenden Systemen, bei dem insbesondere effiziente Algorithmen zur Realisierung der Techniken gefunden und implementiert werden müssen. Die Leistungsfähigkeit des Beweisers SPASS für die Prädikatenlogik mit Gleichheit, der Erweiterung SPASS+T von SPASS um arithmetische Theorien und des Beweisers Waldmeister für Gleichungslogik, wollen wir im Folgenden an jeweils einem Anwendungsszenario vorstellen.

In großen IT-Infrastrukturen ist es nicht trivial, die Konfigurationen von Netzwerkkomponenten (Routern, Firewalls), Servern und Computern von Endanwendern immer konsistent zu halten. Oft ist es schon sehr langwierig, die Firewall des DSL-Anschlusses zu Hause richtig zu konfigurieren. Mit Hilfe von SPASS lösen wir eine Reihe von Fragestellungen in diesem Zusammenhang vollautomatisch. Dazu gehört die Frage, ob ein Computer auf einem (beliebigen) Netzwerk einwandfrei funktionieren wird (in Abhängigkeit von den Server und Netzwerkkonfigurationen) oder welche Firewall-Einstellungen fehlen, um einen bestimmten Dienst zu erlauben.

Arbeitet man mit großen Wissensbasen, kann man gewöhnlich davon ausgehen, dass nicht alle enthaltenen Daten absolut zuverlässig sind. Stattdessen sind die Daten mit einem Zahlenwert versehen, der die Unsicherheit der Information beschreibt. Wenn man aus solchen Formeln Schlüsse ziehen will, reicht es nicht aus, die Formeln logisch zu verknüpfen, sondern es müssen auch die Unsicherheiten der Einzelinformationen kombiniert werden. SPASS+T erweitert



Analyse des DHCP-Dienstes

SPASS um die Möglichkeit, effizient mit solchen numerischen Werten zu arbeiten. Diese Erweiterung erfolgt auf verschiedene Weise: Einerseits kann ein leistungsfähiges Entscheidungsverfahren, das für den Umgang mit mathematischen Formeln spezialisiert ist, an SPASS angebunden werden. Andererseits sind manche Rechenregeln unmittelbar in SPASS+T eingebaut.

Waldmeister zeichnet sich aus durch seine effiziente Inferenzmaschine, durch mächtige Simplifikationstechniken und durch die geschickte Ansteuerung der Beweissuche. Seit nunmehr zehn Jahren gewinnt Waldmeister auf dem internationalen Beweiserwettbewerb CASC in seiner Kategorie und unterstreicht damit seine Effizienz.

Ein prominenter Waldmeister-Anwender ist Stephen Wolfram. Er hat mit Mathematica eines der meistgenutzten mathematisch-naturwissenschaftlichen Programmpakete geschaffen. In seinem Buch *„A New Kind of Science"* geht er unter anderem der Frage nach, wie aus einfachen Mustern große Komplexität entstehen kann. Mit Hilfe von Waldmeister konnte er zeigen, dass sich Boolesche Algebren durch ein einziges Axiom axiomatisieren lassen:
$((x \mid y) \mid z) \mid (x \mid ((x \mid z) \mid x)) = z$

Dabei verhält sich der spezifizierte Operator $\mid$ wie die negierte Disjunktion. :::

KONTAKT

Thomas Hillenbrand
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-217
Email hillen@mpi-inf.mpg.de

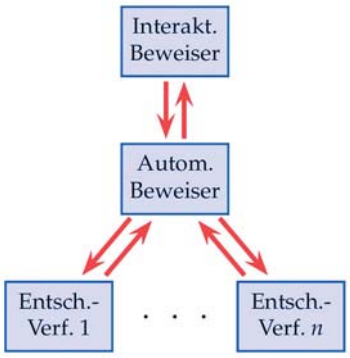
Uwe Waldmann
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-205
Email uwe@mpi-inf.mpg.de

Christoph Weidenbach
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-900
Email weidenbach@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/departments/rg1/software.html

Automatisches Beweisen

Um garantieren zu können, dass eine Hardware oder Software korrekt arbeitet, muss man sie verifizieren – das heißt, die Korrektheit formal nachweisen. Kernelement einer jeden Verifikation ist die Untersuchung, ob bestimmte Eigenschaften aus anderen, bereits bekannten Eigenschaften eines Systems folgen. Mit der Frage, wie man Computerprogramme zum Lösen solcher Beweisaufgaben einsetzen kann, beschäftigen sich Wissenschaftler bereits lange Zeit. Schon seit den fundamentalen theoretischen Ergebnissen von Gödel und Turing zu Beginn des zwanzigsten Jahrhunderts weiß man, dass nicht alles, was im mathematischen Sinne *wahr* ist, auch *beweisbar* ist, und dass nicht alles, was *beweisbar* ist, *automatisch beweisbar* ist. Deduktionssysteme unterscheiden sich dementsprechend deutlich in ihrer Ausdrucksstärke und ihren Eigenschaften: Entscheidungsverfahren sind auf eine bestimmte Art von Daten (etwa reelle Zahlen) spezialisiert und können innerhalb dieses Bereichs garantiert die Korrektheit oder Inkorrektheit einer Aussage nachweisen. Automatische Beweiser für die so genannte erststufige Logik können mit beliebigen, in einem Programm definierten Datentypen umgehen. Hier steht aber nur fest, dass sie einen Beweis finden, falls er existiert; falls keiner existiert, dann suchen sie möglicherweise erfolglos weiter, ohne jemals anzuhalten. Interaktive Beweiser arbeiten mit noch ausdrucksstärkeren als der erststufigen Logik. Sie funktionieren allerdings nur mit Benutzerunterstützung und ohne jede Vollständigkeitsgarantie.

Es ist offensichtlich, dass für praktische Anwendungen im allgemeinen die Kombination aller Methoden nötig ist: Man braucht interaktive Beweiser, um Probleme komfortabel in einer ausdrucksstarken Logik formulieren zu können und um größere Beweise durch eine Auswahl geeigneter Strategien steuern zu können. Man benötigt automatische Beweiser, um den Automatisierungsgrad des interaktiven Beweisers so weit wie möglich zu erhöhen. Und schließlich benötigt man Entscheidungsverfahren, um beispielsweise rein arithmetische Teilaufgaben effizient zu lösen. In unserer Arbeitsgruppe versuchen wir die logischen sowie technischen Schwierigkeiten zu überwinden, die eine solche Kombination mit sich bringt.



Kombination deduktiver Systeme

Neben der Kombination von Deduktionssystemen beschäftigen wir uns zudem seit langem intensiv mit dem automatischen Beweisen. Wie arbeitet ein automatischer Theorembeweiser? Ein Programm zu schreiben, das aus gegebenen Formeln neue Formeln logisch korrekt ableitet, ist nicht schwer. Eine logisch korrekte Ableitung ist allerdings nicht unbedingt eine sinnvolle Ableitung. Wer zum Beispiel $2\cdot a+3\cdot a$ erst in $2\cdot a+3\cdot a+0$ und dann in $2\cdot a+3\cdot a+0+0$ umwandelt, macht zwar keinen Rechenfehler, kommt seinem Ziel aber keinen Schritt näher. Die eigentliche Herausforderung besteht also darin, aus unendlich vielen *korrekten* Ableitungen die wenigen *sinnvollen* Ableitungen herauszusuchen. Dabei stellt man zunächst fest, dass es nützlich ist, Gleichungen so anzuwenden, dass sich das Ergebnis vereinfacht, also etwa $x+0=x$ nur von links nach rechts und nicht umgekehrt.

$$x+0 = x$$
$$x+(-x) = 0$$
$$\frac{x\cdot z}{y\cdot z} = \frac{x}{y}$$

kompliziert → einfach

Gleichungsanwendung

Dieser Ansatz reicht allerdings nicht immer aus. Deutlich wird das beispielsweise bei der Bruchrechnung: Bekanntlich muss man einen Bruch hin und wieder erweitern, bevor man damit weiterrechnen kann. Beim Erweitern passiert aber genau das, was man eigentlich vermeiden möchte: Die Gleichung $(x\cdot z)/(y\cdot z)=x/y$ wird von rechts nach links angewendet – aus einem einfachen Ausdruck wird ein komplizierterer. Der 1990 von Bachmair und Ganzinger entwickelte Superpositionskalkül bietet einen Ausweg aus diesem Dilemma. Einerseits rechnet er vorwärts, andererseits aber identifiziert und repariert er systematisch die möglichen Problemfälle in einer Formelmenge, für die ein Rückwärtsrechnen unvermeidbar sein könnte. Superposition ist damit die Grundlage fast aller heutigen Theorembeweiser für erststufige Logik mit Gleichheit. Das gilt auch für unsere am Institut entwickelten Beweiser SPASS, Waldmeister und Bliksem. Derzeit beschäftigen wir uns nicht nur mit der oben angesprochenen Kombinationsproblematik, sondern insbesondere auch mit speziellen Optimierungstechniken für verschiedene Anwendungen, wie beispielsweise die Analyse von Netzwerkprotokollen oder Ontologien. :::

KONTAKT

Uwe Waldmann
RG. 1 Automatisierung der Logik
Telefon +49 681 9325-205
Email uwe@mpi-inf.mpg.de

Modulares Beweisen in komplexen Theorien

Die großen Fortschritte in der Entwicklung der Informationstechnik haben dazu geführt, dass heutzutage komplexe, rechnergesteuerte Systeme fast überall eingesetzt werden: im Haushalt, in Autos, Zügen, Flugzeugen oder Kraftwerken. Insbesondere in den letztgenannten sicherheitskritischen Bereichen können Fehler katastrophale Folgen haben. Es ist deshalb sehr wichtig, das korrekte Funktionieren solcher Systeme zu garantieren, das heißt mathematisch zu beweisen. Eigentlich wäre es wünschenswert, solche Korrektheitsbeweise völlig automatisch vom Rechner durchführen zu lassen. Fundamentale theoretische Ergebnisse von Gödel, Church und Turing zeigen aber, dass das nicht möglich ist. Für konkrete Anwendungsbereiche existieren jedoch effektive automatische Verifikationsverfahren.

Unser Ziel ist es, Rahmenbedingungen zu identifizieren, unter denen effiziente Verifikationsverfahren für komplexe Systeme existieren. Die formale Beschreibung eines komplexen Systems ist aus Teilen zusammengesetzt, die verschiedenen Bereichen entstammen, so finden sich beispielsweise numerische Formeln neben Aussagen über Datenstrukturen. Es ist daher sehr wichtig, effizient in komplexen Theorien, die als Kombinationen verschiedener Bestandteile entstehen, schlussfolgern zu können. Wir sind daran interessiert, Beweisverfahren zu entwikkeln, die die modulare Struktur der komplexen Theorien ausnutzen, und es erlauben, spezialisierte Beweiser für das Schlussfolgern in den Teiltheorien zu benutzen. Solche modularen Verfahren sind besonders flexibel und effizient und in vielen Bereichen anwendbar (wie etwa in der Mathematik, Verifikation oder Wissensrepräsentation).

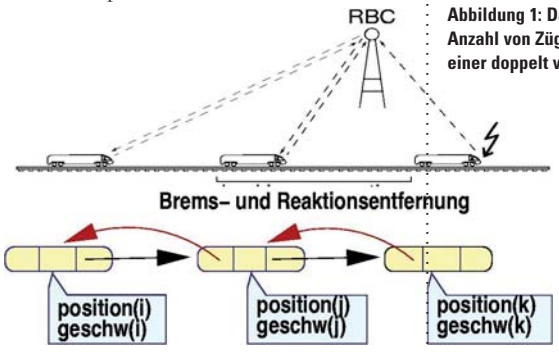


Abbildung 1: Darstellung einer unspezifizierten Anzahl von Zügen auf einer Bahnstrecke mit Hilfe einer doppelt verketteten Liste

Die einfachste Form von komplexen Theorien sind *Erweiterungen* einer gegebenen Theorie (hier als Basistheorie bezeichnet) mit zusätzlichen Funktionen. Theorieerweiterungen kommen z.B. in parametrischen Ansätzen zur Verifikation reaktiver oder hybrider Systeme vor, in denen bestimmte Größen (Zeit, Geschwindigkeit) sowie ihre Veränderungen als Parameter betrachtet werden. Alternativ kann auch die Anzahl von Komponenten ein Parameter sein (Abbildung 1 illustriert die Darstellung einer unspezifizierten Anzahl von Zügen auf einer Bahnstrecke mit Hilfe einer doppelt verketteten Liste, die auch numerische Informationen über Position und Geschwindigkeit der Züge enthält). Im Allgemeinen ist es schwierig, solche Erweiterungen zu behandeln, auch wenn effiziente Verfahren für die Basistheorie vorhanden sind.

Wir gehen die Lösung dieses Problems an, indem wir eine Klasse von Theorieerweiterungen identifizieren, in denen es möglich ist, das ursprüngliche Problem auf ein Problem im Basisbereich zu reduzieren. Dieses kann dann mit einem für die Basistheorie spezialisierten Verfahren gelöst werden. Das allgemeine Prinzip eines solchen hierarchischen Verfahrens ist in Abbildung 2 dargestellt.

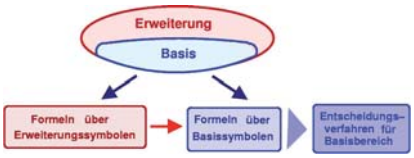


Abbildung 2: Hierarchisches Schließen: Das allgemeine Prinzip

Darüber hinaus entwickeln wir Verfahren für modulares Schließen in *Kombinationen* von Theorien. Wenn wir zum Beispiel bei der Programmverifikation eine Aussage beweisen wollen, in der sowohl von Listen von Zahlen als auch von Arrays von Zahlen die Rede ist, dann können wir dieses Problem in zwei Teilprobleme zerlegen – eines mit Listen und Zahlen, eines mit Arrays und Zahlen – und für jedes davon ein existierendes Beweisverfahren benutzen. Wenn beide Verfahren nun ausreichend viele Informationen über die gemeinsamen Daten (also die Zahlen) austauschen, dann ist gewährleistet, dass zum Schluss die Einzellösungen beider Verfahren zu einer Gesamtlösung kombiniert werden können. Diese Idee ist in Abbildung 3 dargestellt.

Unsere theoretischen Beiträge bilden die Basis für die Entwicklung von praktisch einsetzbaren Verifikationswerkzeugen für die Verifikation sicherheitskritischer Systeme, insbesondere im Rahmen des SFB Transregio Projektes AVACS (Automatic Verification and Analysis of Complex Systems). :::

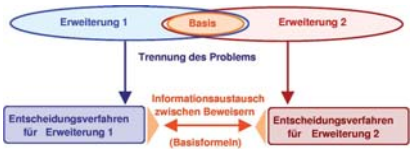


Abbildung 3: Modulares Schließen: Das allgemeine Prinzip

KONTAKT

Viorica Sofronie-Stokkermans
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-207
Email sofronie@mpi-inf.mpg.de

GEOMETRIE

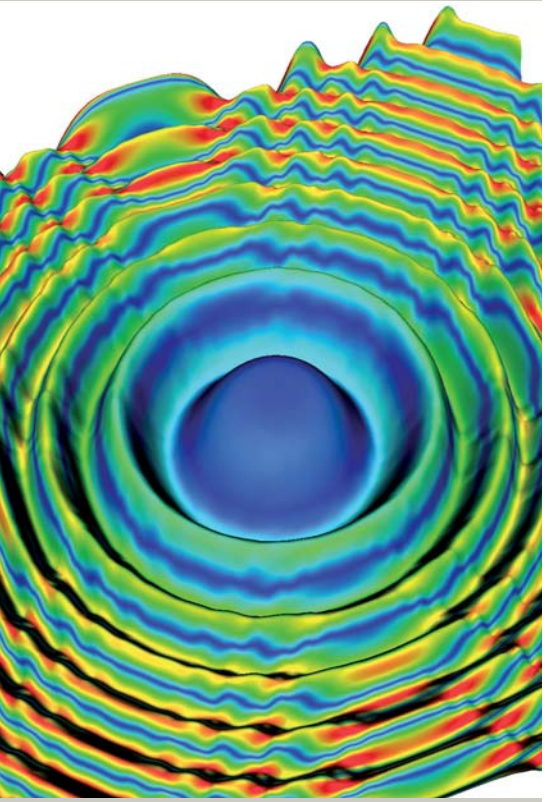
Geometrische Daten sind allgegenwärtig und ihre Verarbeitung ist und bleibt eine der größten Herausforderungen der Informatik.

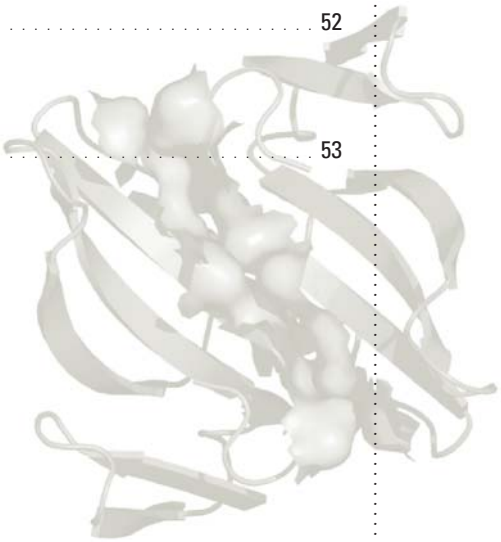
Das Max-Planck Institut für Informatik beschäftigt sich traditionell insbesondere mit geometrischen Fragestellungen. Von besonderem Interesse ist dabei die Verarbeitung digitaler Modelle (Shapes) von Objekten unserer dreidimensionalen Welt. Dazu zählt unter anderem die Erzeugung von Shapes, ihre Modifizierung, ihre Speicherung und die Suche nach ähnlichen Shapes.

Shapes lassen sich oft durch Operationen wie Schneiden und Vereinigen aus einfacheren Shapes erzeugen. Im Rahmen des Exacus-Projektes werden solche Operationen auf Shapes untersucht, die sich durch einfache algebraische Gleichungen beschreiben lassen. In der Produktion von Animationsfilmen erzeugt man Shapes meist als Unterteilungsflächen, da sie dem Designer ein großes Maß an Flexibilität geben und einfach zu modifizieren sind. Flächenmodelle spielen auch in der Bioinformatik in Form von Proteinflächen eine große Rolle. Proteinflächen können dabei helfen, zu entscheiden, ob zwei Proteine aneinander docken können – eine Frage, die fundamental für die Entwicklung neuer Medikamente ist. Das Docking-Problem ist ein Spezialfall der Schwierigkeit, zumindest partiell ähnliche Shapes zu finden. Von großer ökonomischer Bedeutung ist dieses Problem auch beim Verwalten von CAD-Modellen in einer Datenbank. Um in einer solchen Shape-Datenbank effizient Suchen zu können, braucht man geeignete Ähnlichkeitsmaße von Shapes, die sich schnell berechnen lassen und auch partielle Ähnlichkeiten berücksichtigen. Das Max-Planck Institut für Informatik arbeitet aktiv an all diesen Fragestellungen – in dem von der EU geförderten Projekt AIM@SHAPE auch gemeinsam mit internationalen Partnern.

Geometrie taucht aber nicht nur im Zusammenhang mit der Verarbeitung von Shapes auf. Ein Anwendungsgebiet, das in den letzten Jahren weltweit Forscher und Praktiker gleichermaßen beschäftigt, sind drahtlose Sensornetzwerke. Sensoren werden oftmals idealisiert als miteinander kommunizierende Punkte in der Ebene betrachtet. Einsichten in die Geometrie solcher Sensorpunktmengen helfen, gute Kommunikationsprotokolle für die unterschiedlichsten (zum Teil auch geometrischen) Probleme zu finden, die man mit dem Sensornetzwerk lösen will. :::

BEITRÄGE

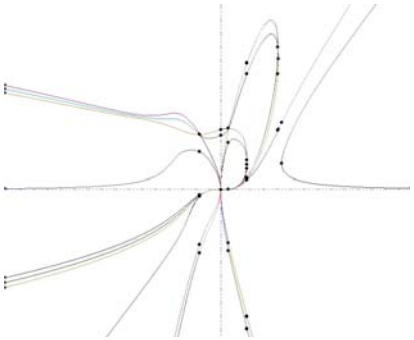




EXACUS: Effiziente und exakte Algorithmen für Kurven und Flächen

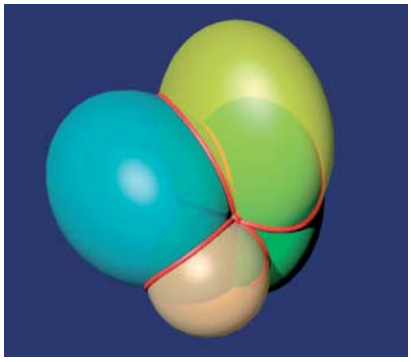
EXACUS *(Efficient and Exact Algorithms for Curves and Surfaces)* ist eine umfangreiche Sammlung von C++-Software-Bibliotheken zur Untersuchung von Kurven und Flächen. Im Gegensatz zur klassischen algorithmischen Geometrie beschränkt sich EXACUS im Wesentlichen auf das Studium linearer Objekte. Das Projekt wurde am Max-Planck-Institut für Informatik im Rahmen des EU-Projektes ECG *(Effective Computational Geometry)* gegründet und stellt auch im Folgeprojekt ACS *(Algorithms for Complex Shapes)* eine wichtige Grundlage für eine Vielzahl von Implementierungen dar.

In der klassischen algorithmischen Geometrie liegt der Fokus auf dem Studium von linearen Objekten wie Geraden, Strecken und Ebenen. Dieser Ansatz wird nun auf die Betrachtung gekrümmter Objekte ausgedehnt. Bei der Implementierung geometrischer Algorithmen treten häufig Schwierigkeiten auf, die sich aus Rechenungenauigkeiten als Folge von Rundungsfehlern ergeben: Falsche Resultate, Crashes oder nichtterminierende Programme können die Folge sein. Während sich bei der Untersuchung linearer Objekte in vielen Fällen die Algorithmen unter Verwendung exakter Numerik implementieren lassen, stößt man im Falle von gekrümmten Kurven und Flächen schnell an die mathematischen Grenzen. Ein Grund dafür ist, dass Lösungen von allgemeinen algebraischen Gleichungssystemen nicht mehr Elemente in Wurzelerweiterungen der rationalen Zahlen sind. Ferner stellen Berechnungen innerhalb solcher Körpererweiterungen eine relativ schlechte Ausgangslage für effiziente Implementierungen dar. In bisherigen Ansätzen, so auch in der praktischen Anwendung bei allen CAD-Systemen, setzt man ausschließlich auf numerische Algorithmen, die nur dann korrekte Ergebnisse liefern, wenn man einen „hinreichend generischen" Input voraussetzt. Beispiele zeigen jedoch, dass das nicht der Weisheit letzter Schluss ist. In unserer Arbeit kombinieren wir deshalb symbolische Verfahren aus der Computeralgebra mit garantierten numerischen Verfahren und effizienten Algorithmen. Auf diesem Wege ist es uns möglich, alle Fälle korrekt und schnell zu untersuchen.



Kurve vom Grad 23 mit hoher Multiplizität im Ursprung

Im Moment können wir die Topologie einer ebenen algebraischen Kurve beliebigen Grades exakt bestimmen (AlciX). Die Kurve wird dabei in Teilsegmente zerlegt, die charakteristische Punkte der Kurve wie Extremstellen und Singularitäten verbinden. Darüber hinaus verfügen wir über eine exakte Visualisierung auf AlciX-Basis. Boolesche Operationen auf Polygonen mit gekrümmten Rändern spielen eine zentrale Rolle in CAD-Systemen. Wir befassen uns mit dem algorithmisch anspruchsvollen Teil der Arrangementberechnung von Kurven in der Ebene: Das Arrangement für eine gegebene Menge von Kurven beschreibt die Zerlegung der Ebene in Kurvenstücke, Schnittpunkte und berandete Flächenstücke. Für Grad-3-Kegelschnitte und -Kurven bestehen bereits Implementierungen. Für Kurven allgemeinen Grades soll eine solche in Kürze verfügbar sein. Im dreidimensionalen Raum arbeiten wir an einer Implementierung von Booleschen Operationen von Flächen, insbesondere Quadriken, also Grad-2-Flächen. Dabei werden zwei Ansätze verfolgt: Im Projektionsansatz werden die Schnitt- und die Silhouettekurve von Quadriken in die Ebene projiziert. Das Arrangement der projizierten Kurven gibt dann Aufschluss über das Arrangement der Quadriken im Raum. Im zweiten Ansatz werden exakte Parametrisierungen der Quadriken betrachtet. Die Berechnungen im Erweiterungskörper stellen hierbei die höchsten Ansprüche an eine effiziente Implementierung dar. Beide Ansätze befinden sich in einem weit fortgeschrittenen Stadium und werden voraussichtlich im laufenden Jahr fertiggestellt.



Arrangement von Quadriken

Exacus hat im August 2006 mit der aktuellen Version 1.0 Marktreife erlangt und ist unter einer Open-Source-Lizenz verfügbar. Derzeit transferieren wir das vorhandene Know-how nach CGAL (siehe auch *„CGAL"*, Seite 72). Das Ziel des Transfers ist es, CGAL künftig als Basis für alle Implementierungen auf diesem Gebiet zu nutzen. Damit schaffen wir eine wesentliche Basis für die Kooperation mit unseren EU-Partnern (ECG und ACS Projekt), mit denen wir erfolgreiche und effektive Softwareprojekte angehen wollen. :::

KONTAKT

Michael Sagraloff
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-106
Email msagralo@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/EXACUS/

Geometrie drahtloser Sensornetzwerke

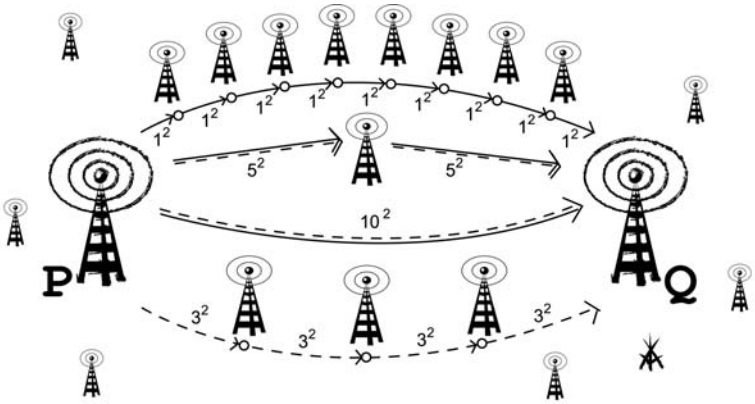
Dank der Fortschritte bei der Miniaturisierung der IT-Technik ist es mittlerweile möglich, einen vollständigen Computer mitsamt drahtloser Kommunikationseinheit, autonomer Stromversorgung und verschiedensten Sensoren zur Erfassung von Umgebungsparametern auf der Fläche eines Fünfmarkstücks unterzubringen. Solche Kleinstrechner – so genannte *Sensorknoten* – können aufgrund ihrer Größe und autonomen Arbeitsweise fast überall verteilt und eingesetzt werden, insbesondere auch in schwer zugänglichen Umgebungen. In Naturreservaten zum Beispiel könnten sie großflächig Bodenklimadaten erfassen. Die erste Generation solcher *Sensornetzwerke* wurde hauptsächlich zur Datenakquisition eingesetzt. Dabei wurden die erfassten Daten zur Auswertung an einen zentralen Rechner übermittelt. Bei neueren Sensornetzwerkarchitekturen kommt hingegen der netzwerkinternen Verarbeitung und Auswertung eine immer größere Bedeutung zu. Kennzeichnend für die Funktionsweise von drahtlosen Sensornetzwerken ist der starke Einfluss der „Geographie" des Netzwerks. Ob zwei Netzwerkknoten direkt miteinander kommunizieren können, hängt stark davon ab, wie weit die beiden voneinander entfernt sind. Was ein Sensor messen kann, ist wiederum stark mit der geographischen Distanz korreliert. Ein Feuchtigkeitssensor zum Beispiel kann die Feuchtigkeit nur in seiner unmittelbaren Umgebung messen. Eine Herausforderung ist bereits die netzwerkinterne Kommunikation: Typischerweise ist es keinem Knoten möglich, direkt mit allen anderen Netzwerkknoten drahtlos zu kommunizieren; aufgrund der geographischen Distanzen müssen Nachrichten also über mehrere Stationen von Knoten zu Knoten übertragen werden.

Energieeffizienter Betrieb von drahtlosen Sensornetzwerken

Sensorknoten werden typischerweise durch eine Batterie betrieben – gegebenenfalls unterstützt durch Solarzellen. Die Energieeffizienz aller verwendeten Protokolle und Algorithmen spielt für die Lebenszeit des Netzwerks also eine zentrale Rolle. In unseren Arbeiten entwickeln wir Algorithmen, die die Aktivitäten der Sensorknoten koordinieren, um die Lebenszeit zu maximieren. Das kann zum Beispiel dadurch geschehen, dass Sensorknoten nach einem berechneten Zeitplan in einen Schlafmodus versetzt werden. Dieser Zeitplan sollte jedoch so strukturiert sein, dass zu keinem Zeitpunkt Lücken in den akquirierten Daten entstehen. Auch bei der Übermittlung einer Nachricht von einem Knoten im Netzwerk zu einem anderen, in der netzwerkinternen Kommunikation also, lässt sich der Gesamtenergieverbrauch durch geschickte Nutzung von Zwischenstationen stark reduzieren. Einfach deshalb, weil die benötigte Energie zur Übertragung einer Nachricht superlinear mit der zu überbrückenden Distanz wächst.

Wieviel Geometrie versteckt sich in Konnektivität?

Beim Einsatz einer sehr großen Zahl von Knoten etwa wäre es teuer, jeden einzelnen dieser Knoten mit einer eigenen GPS-Einheit zur Positionsbestimmung auszustatten. Nach einer typischerweise eher unkontrollierten Ausbringung der Knoten aber – zum Beispiel von einem Flugzeug aus – kennen die meisten Sensorknoten ihre geographische Position nicht. Dennoch ist eine Lokalisierung möglich, denn gewisse Informationen über die Knotenpositionen sind implizit im *Kommunikationsgraphen* enthalten. Der Kommunikationsgraph charakterisiert, welche Knoten mit welchen anderen Knoten direkt in Verbindung stehen können. Ziel unserer Arbeiten ist es, durch Analyse dieses Kommunikationsgraphen Rückschlüsse auf die geographischen Positionen im Netzwerk zu ziehen, insbesondere wenn keiner oder nur wenige der Netzwerkknoten mit GPS-Einheiten ausgestattet sind. :::



Geschickte Nutzung von Zwischenstationen bei der Übermittlung von Nachrichten kann den Energieverbrauch reduzieren.

Bei der Berechnung eines optimalen Zeitplans oder bei der Identifizierung guter Zwischenstationen spielt der geometrische Aspekt folglich eine besondere Rolle.

KONTAKT

Stefan Funke
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-108
Email funke@mpi-inf.mpg.de

Geometrische Probleme in der Bioinformatik

Proteine spielen eine Schlüsselrolle in lebenden Organismen (siehe auch *„Struktur-Funktionsbeziehungen bei Proteinen“*, Seite 32). Das Wissen um die dreidimensionale *(3D)* Struktur von Proteinen ist essenziell, weil diese die Interaktion mit Pharma-Wirkstoffen und mit anderen Proteinen entscheidend beeinflusst. Nur mit profundem Wissen über diese Proteinstruktur kann man letztlich biochemische Zusammenhänge verstehen und neue Pharma-Wirkstoffe zielgerichtet entwickeln. So lässt sich beispielsweise durch rechnerisches Ausprobieren feststellen, ob kleine Moleküle an ein Protein binden können und sich damit als Wirkstoff eignen. Die bioinformatische Analyse von 3D-Strukturen von Proteinen und potenziellen Wirkstoffen schließt zahlreiche geometrische Fragestellungen ein.

Doch die Analyse von Protein-Strukturen hat es in sich. Energieminimale 3D-Strukturen von kleinen Molekülen lassen sich noch berechnen. Bei Proteinen aber ist der Konformationsraum – also die Vielfalt der möglichen geometrischen Varianten – so komplex, dass die Struktur nur durch Experimente bestimmt werden kann. Für derartige Experimente setzt man beispielsweise die Röntgenkristallographie ein. Dieses Verfahren liefert als Ergebnis typischerweise räumliche Koordinaten mit den Hauptaufenthaltswahrscheinlichkeiten für jedes Atom und die erwarteten Abweichungen. Wir beziehen diese Koordinaten überwiegend aus öffentlichen Datenbanken. Mittlerweile stehen darin mehr als 37.000 Proteinstrukturen zur Verfügung.

Beispiele für geometrische Probleme bei Strukturanalyse von Proteinen sind die Definition einer molekularen Oberfläche, die Suche nach ähnlichen Oberflächen oder nach strukturell ähnlichen Teilstücken von Proteinen.

Um die strukturelle Suche nach ähnlichen oder komplementären Proteinteilen zu ermöglichen, entwickeln wir zurzeit Deskriptoren, mit denen sich Teile – insbesondere Oberflächenstücke – von Proteinen effizient beschreiben, inventarisieren und suchen lassen.

Solche Deskriptoren fassen dabei die relevanten geometrischen und chemischen Eigenschaften des Proteins in einer Vektorrepräsentation zusammen. In dieser Form lassen sich in wenigen Sekunden Tausende von Teilstücken miteinander vergleichen. Abbildung 1 zeigt eine Interaktion zwischen zwei Proteinen. In Abbildung 2 ist die herausgelöste Bindungsstelle eines Proteins zu sehen. Häufig gibt es zu beiden Bindungspartnern evolutionär verwandte Proteine. Interessant ist dann die Frage, ob die verwandten anderen Proteine ganz wie die ursprünglichen Bindungspartner auch miteinander reagieren würden. Dies lässt sich rechnerisch untersuchen, indem man die Bindungsmodi analysiert und testet, ob sie sich auf die ähnlichen Proteine übertragen lassen. Somit kann man überprüfen, ob sich eine Funktionshypothese von einem Paar von Proteinen auf ein anderes übertragen lässt. Diese Überprüfung ist wichtig, da sich wegen des riesigen Aufwands nicht alle möglichen Funktionshypothesen experimentell bestätigen lassen.

Bei Interaktionen zwischen Proteinen spielt die komplementäre Geometrie der beteiligten Oberflächen eine große Rolle. Methoden der algorithmischen Geometrie lassen sich jedoch kaum ohne spezielle Anpassungen in der Bioinformatik einsetzen. Das liegt daran, dass die Analyse molekularer Strukturen nicht nur geometrische, sondern auch energetische Aspekte hat: Die Moleküle und deren Teile müssen nicht nur ineinander passen wie ein dreidimensionales Puzzle. Berücksichtigt werden muss auch, dass ihre Teile darüber hinaus noch beweglich sind, weil zwischen den verschiedenen molekularen Komponenten komplexe Kräfte herrschen. Die Richtung dieser Kräfte wird wiederum durch die zugrunde liegende Chemie bestimmt.

Es bedarf also komplexer Methoden, um sowohl den geometrischen als auch den energetischen Aspekt angemessen zu berücksichtigen. Die entsprechenden Probleme können teilweise exakt gelöst werden. In der Regel werden die Lösungen aber mit Heuristiken und statistischen Methoden lediglich angenähert. Vor diesem Hintergrund kommt der Validierung der Methoden, also der Überprüfung mit experimentell gemessenen Daten, eine besondere Bedeutung zu.

Am Max-Planck-Institut für Informatik werden geometrische Methoden vor allem beim Docking, bei der Suche nach Wirkstoffen und bei der Struktur- und Funktionsanalyse von Proteinen eingesetzt. :::

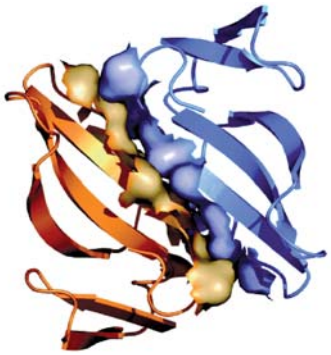


Abbildung 1: Zwei Proteine, blau und orange, in typischer Cartoondarstellung; die Interaktion ist als Oberflächendarstellung hervorgehoben.

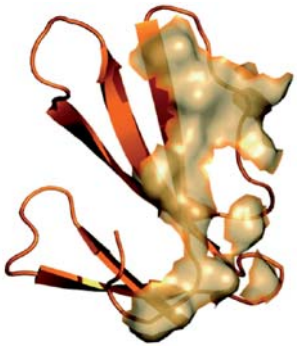


Abbildung 2: Leicht gedreht die Bindungsstelle des einen Proteins

KONTAKT

Ingolf Sommer
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-306
Email sommer@mpi-inf.mpg.de

Digitale Geometrieverarbeitung

Sphärische Parametrisierung

Das Ziel der Parametrisierung von Flächen ist es, Abbildungen zu finden, die komplexe Flächen auf einfachere, äquivalente Bereiche, zum Beispiel Ebenen oder Sphären projizieren. Eine solche Parametrisierung ist wesentlich für viele Anwendungen der digitalen Geometrieverarbeitung wie Texturabbildungen, Formanalyse, Kompression oder Verformung. Der sphärische Fall ist aus Dimensionsgründen deutlich komplexer als der planare. Einer unserer Forschungsschwerpunkte ist es, effiziente und stabile Lösungen für dieses Problem zu finden. Dazu formulieren wir das Problem so in kurvilineare Koordinaten um, dass wir es auf den zweidimensionalen Fall zurückführen können. Dadurch erhalten wir zusätzlich eine bessere Verzerrungskontrolle. Mithilfe dieser Technik können Netze mit nicht-trivialer Geometrie und Zehntausenden von Dreiecken innerhalb weniger Sekunden verarbeitet werden. Diese Rechenleistung überbrückt die bislang große Kluft zwischen sphärischer und planarer Parametrisierung.

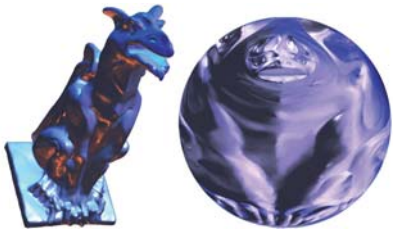


Abbildung 1: Sphärische Parametrisierung eines hochaufgelösten Gargoyls

Abbildung 2: Sphärische Parametrisierung einer hochaufgelösten Schildkröte (Arbeit von Rhaleb Zayer, SMI '06)

Entrauschen von Geometriedaten

Ein wichtiger Aufgabenbereich in der geometrischen Datenverarbeitung ist das Entrauschen von Punktdaten, die beispielsweise durch mehrere Laserentfernungsmessungen gewonnen werden. Eine solche Vorverarbeitung ist nötig um entstandene Messungenauigkeiten auszugleichen, die bei jedem physikalischen Aufnahmeprozess unweigerlich entstehen. Des Weiteren wird die anschließende Verarbeitung der Daten, wie beispielsweise die Rekonstruktion einer Oberfläche, erleichtert [Abbildung 3]. Ein besonders großes Problem ist die Entfernung von Punkten mit groben Messfehlern, so genannten *„outliern"*. Dieses lässt sich mit robusten statistischen Methoden behandeln. Darüber hinaus ist im Zuge der Entwicklung neuer Aufnahmeverfahren, die Vorverarbeitung von zeitlich variierenden geometrischen Daten von großem Interesse.

Abbildung 3: Verrauschte Punktdaten vor der Vorverarbeitung *(links)* und eine rekonstruierte Oberfläche nach dem Entrauschvorgang *(rechts)*, (Arbeit von Oliver Schall, SPBG '05)

Verallgemeinerte baryzentrische Koordinaten

Häufig gibt es verschiedene Möglichkeiten, die Lage eines Ortes zu beschreiben. So kann man für einen Ort globale Koordinaten angeben *(„Saarbrücken liegt auf 49°14' Nord, 7°0' Ost")*. Eine Alternative ist eine Beschreibung mithilfe lokaler Bezugspunkte *(„Saarbrücken liegt in der Mitte zwischen Trier, Metz und Straßburg")*. Dabei kommt es auf den jeweiligen Kontext an, welche

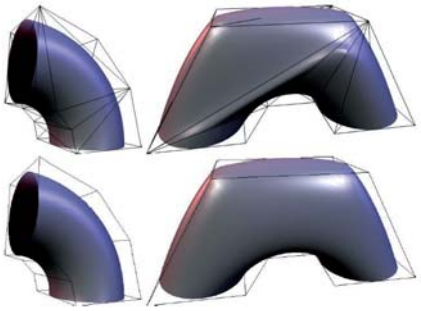


Abbildung 4: Verformung eines Rohrstückes (Arbeit von Torsten Langer, SGP '06)

der beiden Beschreibungen geeigneter ist. Oftmals aber ist die zweite, lokale Variante anschaulicher.

Wenn diese lokale Beschreibung präzise genug ist, um die Position von Saarbrücken exakt zu bestimmen (zum Beispiel durch Angabe von Entfernungen und Winkeln zu Trier, Metz und Straßburg), spricht man von *baryzentrischen Koordinaten*. Diese wurden schon 1827 von Möbius definiert, allerdings nur für den Fall, dass die Bezugspunkte die Eckpunkte eines Dreiecks bilden. Wenn wir aber einen vierten Bezugspunkt hinzunehmen (etwa Kaiserslautern), funktioniert Möbius' Methode nicht mehr. In unserer Arbeitsgruppe haben wir diese baryzentrischen Koordinaten auf beliebige Körper verallgemeinert. Diese Methode kann beispielsweise in der 3D-Verarbeitung genutzt werden, um Objekte zu verformen. In Abbildung 4 ist eine Röhre zu sehen, deren Oberfläche bezüglich des schwarzen Kontrollnetzes in verallgemeinerten baryzentrischen Koordinaten beschrieben wird (links). Wird nun das Kontrollnetz verformt, folgt die Röhre dieser Bewegung (rechts). Es ist deutlich zu erkennen, dass dabei ungewollte Effekte auftreten können, wenn das Kontrollnetz nur aus Dreiecken besteht (oben). Mit unserer Methode sieht das Ergebnis deutlich besser aus (unten). :::

KONTAKT

Alexander Belyaev
ABT . 4 Computergraphik
Telefon +49 681 9325-400
Email belyaev@mpi-inf.mpg.de

Dezentrale 3D Verarbeitung

Mit dem schnellen Wachstum des Internets werden digitale Bibliotheken und Datensammlungen eine immer wichtigere Informations- und Datenquelle für Wissenschaftler, Forscher und Studenten. Insbesondere die online verfügbare Menge an geometrischen 3D-Informationen wächst dramatisch. Diese explosionsartige Wissensvermehrung in der Entwicklung und Benutzung von 3D-Inhalten stellt die Wissenschaftler vor neue Forschungsherausforderungen. Um diese anzugehen, wurde AIM@SHAPE (*http://www.aimatshape.net*) gegründet, ein Zusammenschluss von 14 Forschungseinrichtungen, die im Rahmen des sechsten Forschungsrahmenprogramms der EU (Exzellenznetz #506766) vier Jahre lang finanziell unterstützt werden.

Auch unsere Forschungsgruppe *„Geometrische Modellierung"* nimmt am AIM@SHAPE-Projekt teil. Eine unserer Hauptaktivitäten besteht in der Bereitstellung einer hochentwickelten Digital-Shape-Workbench (DSW). Deren Hauptbestandteile sind das Shape-Repository (*http://shapes.aim-at-shape.net*, wird von uns gewartet), das eine Vielzahl von digitalen 3D-Modellen zur Verfügung stellt, das Tools-Repository (*http://www-sop.inria.fr/aim-at-shape*, wird von unseren Partnern bei INRIA gewartet), das Softwarewerkzeuge zum Bearbeiten von 3D-Modellen anbietet, und das zentrale Internet-Portal (wird von uns gewartet).

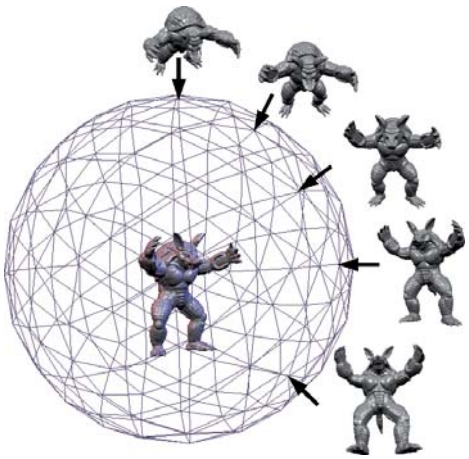
Das Hauptziel des Shape-Repositories ist es, eine gemeinsame Datenbasis von Standardmodellen für Vergleichstests anzubieten. Damit wird eine effiziente Entwicklung von Prototypen und eine praxisnahe Bewertung realer, hochaufgelöster 3D-Modelle ermöglicht. Sein Hauptmerkmal ist eine vollständige Dokumentation der interessantesten geometrischen Merkmale, die in Form detaillierter Metainformationen durch eine gemeinsame 3D-Ontologie angegeben wird. Hochauflösende Modelle werden speziell für das Repository erfasst. Programme aus dem Tool-Repository wiederum werden benutzt, um automatisch Metadaten für bestimmte 3D-Kategorien zu extrahieren.

Alle Projektpartner stellen in das Shape-Repository Modelle aus ihrer aktuellen Forschung ein. Zurzeit (Februar 2007) enthält das Repository über 700 digitale 3D-Objekte. Fast alle wurden speziell zu diesem Zweck erzeugt. Das Repository wird regelmäßig in Fachartikeln dieses Forschungsbereichs zitiert. In den letzten 30 Monaten hatte es über 250.000 Besucher, die mehr als 25.000 Modelle heruntergeladen haben.

Das Repository enthält eine Visualisierungssoftware und einige Werkzeuge aus dem Tools-Repository, mit dem sich die unterstützten Dateiformate bearbeiten lassen. Metadaten werden automatisch erzeugt. Eine Darstellung der Modelle ist in mehreren Auflösungen verfügbar.

Um die 3D-Objekte automatisch mit den Softwarewerkzeugen in Verbindung zu bringen, sollen die verschiedenen Komponenten der DSW vollständig miteinander verknüpft werden. Damit lässt sich zudem aufzeigen, in welchen Publikationen die Objekte erscheinen. Zu den Komponenten der DSW zählen das Shape-Repository, das Tools-Repository, die digitale Bibliothek und das Ontologie- und Metadatenverzeichnis. Diese Integration wird eine intelligente, semantik-basierte Suche ermöglichen (zurzeit in der Entwicklung). So lassen sich durch die Metadaten semantische Bezüge aufspüren, die durch herkömmliche Textsuche nicht auffindbar sind. Unsere Partner entwickeln Module zum Verarbeiten natürlicher Sprache. Zusammen mit einem Index befreien diese den Benutzer von der Notwendigkeit, unklare Fachbegriffe benennen und kennen zu müssen. Außerdem entwickeln unsere Partner eine geometrische Suche. Diese fahndet im Shape-Repository nach Objekten, die einem bestimmten Referenzmodell ähneln.

Die Arbeit an der Entwicklung des Shape-Repositories und der zugehörigen Softwarewerkzeuge stimuliert unsere eigene Forschung erheblich. Gibt sie doch neue Anreize für das automatische Finden, Bezeichnen und Darstellen von 3D-Modellen. Zu unseren wichtigen aktuellen Entwicklungsarbeiten zählt insbesondere ein Aspekt, der in letzter Zeit auf dem Gebiet der Mustererkennung für Aufsehen gesorgt hat: Gegeben sei ein 3D-Modell. Die Aufgabe besteht darin, eine oder mehrere Perspektiven zu finden, von denen aus das Modell gut betrachtet werden kann. Unsere neuesten Ansätze bezüglich dieses Problems sind Lernmodelle, die automatisch herausfinden, welche Teile eines Objektes sich oben und welche sich unten befinden. :::



Ein 3D Objekt kann aus vielen verschiedenen Blickrichtungen betrachtet werden. Wir berechnen eine Anzahl repräsentativer Blickrichtungen, um eine effiziente Darstellung des Objektes zu erhalten.

KONTAKT

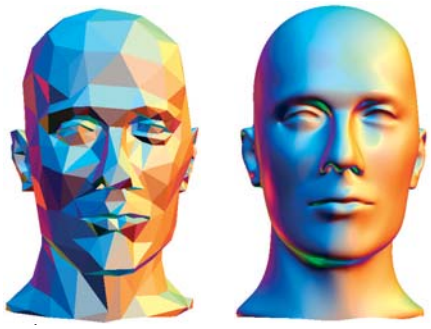
Alexander Belyaev
ABT . 4 Computergraphik
Telefon +49 681 9325-400
Email belyaev@mpi-inf.mpg.de

Freiformflächen und Visualisierung

Unterteilungsflächen

Beim Entwurf von geometrischen Modellen, beispielsweise für Animationsfilme, sind Unterteilungsflächen *(subdivision surfaces)* mittlerweile unverzichtbar. Gründe dafür sind die hohe Flexibilität eines solchen Flächenmodells und die einfache Handhabung: Der Designer gibt nur den groben Verlauf der gewünschten Fläche vor. Dann wird eine glatte, ästhetische Fläche automatisch erzeugt, indem die bestehenden Flächenstücke *(Polygone)* immer feiner unterteilt werden.

Bereits nach wenigen solcher Unterteilungsschritte liegt ein Ergebnis vor, das visuell glatt wirkt. Für die Untersuchung und den Entwurf derartiger Unterteilungsverfahren sind folgende Fragestellungen besonders interessant: Was passiert nach unendlich vielen Unterteilungsschritten? Welche Klassen von Flächen werden beschrieben, und wie glatt sind sie in mathematischer Hinsicht? Unsere Abteilungen befassen sich mit theoretischen Problemen dieser Art aber auch mit praktischen Aspekten. Hier gilt es beispielsweise, Unterteilungsflächen möglichst gut an Datenpunkte anzupassen, die mit einem 3D-Scanner gemessen wurden. Solche Flächenmodelle dienen etwa der effizienten Speicherung der Geometriedaten oder als Ausgangsdaten für einen Künstler, der sie weiter verformt und gestaltet.



Graphik einer Büste mit grober und feiner Flächenunterteilung

Gestaltung von Flächen

Hat man mit Polygonnetzen und Unterteilungsflächen ein geeignetes Geometriemodell gefunden, muss man eine weitere Herausforderung meistern: Das Geometriemodell muss künstlerisch und technisch ausgestaltet und designed werden – ein Prozess, der mit digitaler Bildbearbeitung vergleichbar aber in vielerlei Hinsicht anspruchsvoller ist. Nicht zuletzt wegen der hohen Zahl an Gestaltungsmöglichkeiten.

So können Flächen etwa beliebig verformt werden. Dabei gibt der Benutzer mit minimalen Aufwand wenige Parameter vor. Das Ergebnis muss schnell (in Echtzeit) visualisiert werden und natürlich aussehen. Neben reiner Deformation sind auch viele andere Arten der Manipulation denkbar, etwa eine Kombination von Flächenstücken durch Einpassen und Verschmelzen.

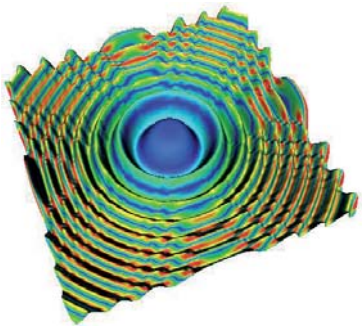
Dem Benutzer bleibt in der Regel verborgen, dass das, was so einfach aussieht, eine Reihe von wissenschaftlich interessanten Problemstellungen birgt. Besonders knifflig sind Fragen wie: Was bedeutet Ästhetik für einen Rechner, oder wie lässt sie sich mathematisch ausdrükken? Wie kann man die Physik von analogen technischen Prozessen geeignet nachbilden? Wie kann das alles möglichst effizient algorithmisch formuliert werden?

Verschmelzung von zwei Flächen

Visualisierung von Volumendaten

Viele Anwendungen stützen sich auf so genannte Volumendaten: In einem räumlichen Bereich werden bestimmte Dichtewerte gemessen. Ein typisches Beispiel hierfür sind medizinische Daten, etwa aus einem Computertomographen. Zur Visualisierung solcher Daten werden bestimmte Oberflächen *(Iso-Flächen)* extrahiert. Dabei handelt es sich um Übergangsflächen zwischen verschiedenen gemessenen Materialien, wie zum Beispiel Gewebe und Knochen.

Die gemessenen Eingabedaten bestehen aus einer Vielzahl von einzelnen *(diskreten)* Datenpunkten. Um diese Daten überhaupt sinnvoll aufbereiten zu können, muss ein geeignetes *(kontinuierliches)* mathematisches Modell geschaffen werden. Dieses schätzt plausible Datenwerte gewissermaßen auch zwischen den gegebenen Punkten ab. Dabei gilt es, die goldene Mitte zwischen verschiedenen Anforderungen zu finden, beispielsweise zwischen Effizienz und Genauigkeit. Ein allgemeiner Ansatz ist die Zusammensetzung von sehr einfachen mathematischen Objekten *(Polynomen)* zu einem umfassenden Modell. Die Herausforderung besteht in der Wahl möglichst einfacher Bausteine und deren geschickter Verzahnung. :::



Visualisierung eines Testdatensatzes

KONTAKT

Christian Rössl
ABT. 4 Computergraphik
Telefon +49 681 9325-400
Email roessl@mpi-inf.mpg.de

INTERNET

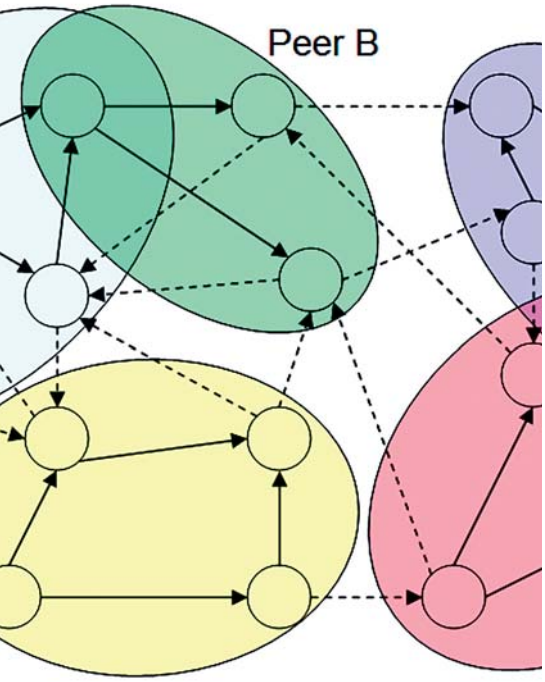
Das Internet hat sich innerhalb weniger Jahre zu einer der wohl weltweit wichtigsten Wissensquellen entwickelt. Informationsdienste, auf die über das Internet zugegriffen werden kann, tragen wesentlich dazu bei. Dazu gehören digitale Bibliotheken, virtuelle Museen, Proteindatenbanken und andere wissenschaftliche Datenarchive sowie Suchmaschinen und E-Commerce-Anbieter. Die Dienstqualität dieser vielfältigen „E-Services" ist jedoch alles andere als befriedigend: Der Nutzer sieht sich mit inakzeptabel langen Antwortzeiten konfrontiert. In anderen Fällen sind zu wichtigen Zeiten bestimmte Dienste nicht verfügbar. Und häufig erhält der Nutzer unbrauchbare Suchresultate.

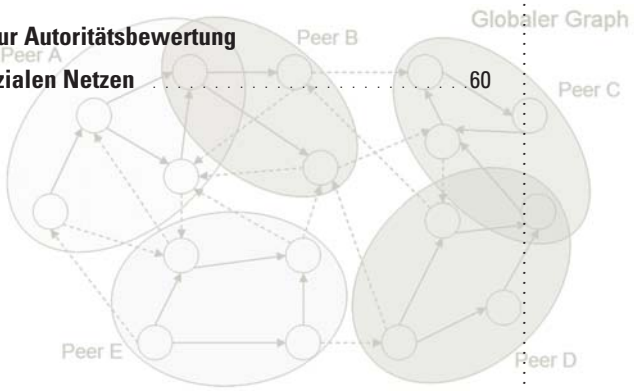
Ein zentrales Ziel der Informatikforschung muss daher sein, das Internet der Zukunft nicht einfach wie bisher dem Kommerz und den Zufälligkeiten industrieller Trends und Interessen zu überlassen, sondern aktiver als bisher systematisch mitzugestalten und zu einer hochgradig verlässlichen Infrastruktur unserer Informationsgesellschaft zu entwickeln, in der Informationsdienste mit Leistungsgarantien betrieben werden. In erster Linie muss die Suchtechnologie für das Web, auf der im Wesentlichen auch das Suchen in Intranets, digitalen Bibliotheken und wissenschaftlichen Datenarchiven beruht, signifikant verbessert werden, um die anspruchsvolleren Informationsbedürfnisse von Wissenschaftlern, Studenten und anderen Individualisten mit hoher Präzision und akzeptabler Effizienz befriedigen zu können.

Am Max-Planck-Institut für Informatik werden diese Themen in verschiedenen Abteilungen unter verschiedenen Blickwinkeln untersucht. Die aktuellen Arbeiten reichen von der intelligenten und effizienten Informationssuche über die Gewährleistung von „Quality-of-Service"-Eigenschaften und die Optimierung von Internet-Protokollen bis zu Anwendungen in der Jurisprudenz. Dort werden die Relevanz der Globalisierung und die damit einhergehenden Fragestellungen mehr als deutlich. Das Institut ist an mehreren großen Forschungsprojekten der EU beteiligt, bei denen Internet-Aspekte im Vordergrund stehen.

Ein grundlagenorientiertes EU-Projekt, in dem die Abteilung *„Algorithmen und Komplexität"* und die Abteilung *„Datenbanken und Informationssysteme"* intensiv zusammenarbeiten, ist *DELIS (Dynamically Evolving Large-Scale Information Systems)*. Hier werden Prinzipien selbstorganisierender komplexer Systeme untersucht. Von Interesse sind dabei Fragestellungen wie diese: Wie etwa kann ein Wissenschaftler, der nach Spezialinformation sucht, schnell mit Hilfe des Internets Daten- und Wissensressourcen sowie andere Wissenschaftler finden, die sich aktuell mit einschlägigen Themen beschäftigen? Und wie kann er diese in einen dynamischen E-Science-Verbund mit einbeziehen? Wie stellt man sicher, dass in einem großen Rechnerverbund mit Tausenden oder Millionen von hochgradig fluktuierenden Knoten, einem so genannten *„Peer-to-Peer-System"*, jeder jeden erreichen kann, und zwar schnell und mit möglichst geringer Belastung des Netzwerks? Mit welchen Mechanismen aus der ökonomischen Spieltheorie kann man Anreize zur elektronischen Kooperation schaffen und Trittbrettfahrer identifizieren? Lösungen dieser aktuellen Forschungsfragen würden die Vision der Welt als globalem Dorf in greifbare Nähe rücken und entscheidend zu einer intelligenten und verlässlichen, Internet-basierten Infrastruktur unserer Gesellschaft beitragen. :::

BEITRÄGE





Finden, was man sucht

Konzepte statt Worte

Gut die Hälfte aller Suchanfragen an herkömmliche Suchmaschinen scheitert daran, dass es nicht die Worte der Suchanfrage sind, die gesucht wurden, sondern eigentlich das dahinterstehende Thema oder ein bestimmter Kontext. Sucht man in Tausenden seiner Mails nach *„Geburtstag“* findet man möglicherweise nicht die eine gewünschte Nachricht. Vielleicht ist das einfach so, weil in der gesuchten Mail lediglich erzählt wird, dass jemand *„50 Jahre alt wird“*, ohne dass das Wort *„Geburtstag“* explizit erwähnt ist. In anderen Fällen befördert eine Suchanfrage nach *„Matrix Zerlegung“* möglicherweise nur Fachartikel zutage, in denen von *„Faktorisierung“* die Rede ist. Oder sie liefert Ergebnisse, die sich nicht mit *„Matrizen“*, sondern ausschließlich mit *„linearen Gleichungssystemen“* befassen. Es scheint, als müsse man zur Lösung dieses Problems Sprachwissen in die Suchmaschine einbauen. Zum einen Wissen darüber, welche Worte ähnliche Bedeutungen haben, wie zum Beispiel *„Zerlegung“* und *„Faktorisierung“* (Synonymie); zum anderen Wissen über die verschiedenen Bedeutungen ein- und desselben Wortes, zum Beispiel *„Surfen auf dem Wasser”* und *„Surfen im Internet“* (Polysemie). Solches Wissen elektronisch zu erfassen, ist allerdings ein sehr aufwändiges und wartungsintensives Unterfangen. Besonders interessant ist daher die Frage, ob sich diese Arbeit nicht wenigstens teilweise automatisieren lässt.

Konzeptsuche = Matrixzerlegung

Eine kompakte und bewährte Art, eine Menge von Dokumenten im Rechner zu repräsentieren, ist die so genannte Wort-Dokument-Matrix. Dabei entspricht jede Zeile einem der irgendwo vorkommenden Wörter und jede Spalte einem Dokument. Ein bestimmter Eintrag in der Matrix gibt dann an, mit welcher Signifikanz – zum Beispiel einfach wie oft – das entsprechende Wort in dem entsprechenden Dokument vorkommt. Für die Menge der von Google indizierten Dokumente wäre das beispielsweise eine zirka 2 Millionen x 4 Milliarden große Matrix. Diese ist allerdings sehr dünn besetzt, denn die meisten Einträge sind Null, weil ja jedes Dokument nur einen Bruchteil aller möglichen Wörter enthält.

Verblüffenderweise lässt sich das eingangs beschriebene Problem der Konzeptsuche als rein mathematisches Matrixzerlegungsproblem formulieren. Dies ist in den nebenstehenden Abbildungen anhand eines Beispiels illustriert. Dass das ganze Konzept überhaupt funktionieren kann, liegt daran, dass die erstellten Dokumente selbst schon sehr viel Information darüber enthalten welche Wörter thematisch zusammengehören. Schließlich ist in die Dokumente viel Sprachintelligenz eingeflossen.

Unsere Forschung

Der Matrixzerlegungsansatz funktioniert in der Praxis erstaunlich gut. Allerdings konnte man bislang nicht schlüssig erklären, warum dem so ist oder unter welchen Bedingungen er funktioniert. Letztlich hatte man bisher in der Regel durch schlichtes Ausprobieren – beispielsweise durch Variieren der geeigneten Zahl von Konzepten – Lösungen gefunden. Vor zwei Jahren aber haben wir ein neues Modell vorgestellt, das erstmals überzeugend erklärt hat, wie Verfahren der beschriebenen Art Paare von zusammenhängenden Wörtern finden können. Das Prinzip: Das Verfahren erkennt, dass bestimmte Wörter gemeinsam mit anderen auftreten – beispielsweise sowohl *„Internet“* als auch *„Online“* zusammen mit dem Wort *„Surfing“*.

Inzwischen haben wir dieses Modell dahingehend weiterentwickelt, dass es auch eine *„Richtung“* in den Wortbeziehungen zu erkennen vermag, beispielsweise dass *„Hawaii“* eine *„Insel“* ist, aber nicht umgekehrt. Wir haben daraus das erste auf Matrixzerlegung basierende Verfahren entwickelt, das vollautomatisch eine ganze Hierarchie von Schlagwörtern aus einer gegebenen Textmenge extrahieren kann. :::

Internet	**0**	**2**	**0**	**1**	**0**	**0**
Online	**2**	**1**	**0**	**0**	**0**	**0**
Surfen	**1**	**1**	**0**	**1**	**1**	**1**
Insel	**0**	**0**	**1**	**1**	**1**	**1**
Hawaii	**0**	**0**	**2**	**2**	**2**	**1**

Abbildung 1 Eine einfache Wort-Dokument-Matrix. Jede Spalte entspricht einem Dokument. Das zur dritten Spalte gehörende Dokument beispielsweise enthält einmal das Wort *„Insel“* und zweimal das Wort *„Hawaii“*, aber keins der anderen Wörter. Eine konventionelle Suchmaschine würde für die Suchanfrage *„Internet“* nicht zwischen Dokument 1 und 3 unterscheiden: keins von beiden enthält das Wort Internet, und die Suchmaschine weiß nicht, dass Dokument 1 zum Thema *„Internet“* passt, Dokument 3 dagegen nicht.

Internet	**2**	**0**
Online	**2**	**0**
Surfen	**1**	**1**
Insel	**0**	**1**
Hawaii	**0**	**2**

1.0	**1.0**	**0.0**	**0.5**	**0.0**	**0.0**	**„Thema” Computer**
0.0	**0.0**	**1.0**	**0.5**	**1.0**	**1.0**	**„Thema” Urlaub**

Abbildung 2 Das Produkt der beiden Matrizen ergibt eine Matrix, die der aus Abb. 1 sehr ähnlich ist. Jede der beiden Spalten der linken Matrix kann als Konzept aufgefasst werden, wobei die Einträge angeben, wie sehr ein Wort auf das Konzept hinweist. Die rechte Matrix stellt jedes Dokument als Kombination dieser Konzepte dar. Eine Suchanfrage nach *„Internet“* könnte jetzt eindeutig dem ersten Konzept zugeordnet werden, womit jetzt deutlich erkannt wird, dass Dokument 1 relevanter ist als Dokument 3.

KONTAKT

Holger Bast
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-100
Email bast@mpi-inf.mpg.de

R4eGov – Sicherheit in der elektronischen Verwaltung

Hintergrund

Das vereinte Europa wächst wirtschaftlich immer weiter zusammen. Die internationale Zusammenarbeit von Verwaltungsbehörden der verschiedenen Mitgliedsstaaten gewinnt damit stark an Bedeutung. Ziel dieser Zusammenarbeit ist es, Verwaltungsabläufe sowohl effizienter als auch für den Bürger transparenter und nutzungsfreundlicher zu gestalten. Während bereits in vielen Mitgliedsstaaten Verwaltungsprozesse elektronisch ausgeführt werden, sind bei der geplanten Kollaboration computerunterstützter Verwaltungen mehrere Herausforderungen zu meistern. Dazu gehören sowohl generelle Anforderungen, wie die Beachtung landesspezifischer Gesetze und die Bewahrung der Privatsphäre der Bürger, aber auch technologische Herausforderungen, die durch die bereits vorhandenen technischen Systeme und Prozesse gegeben sind. In dem von der Europäischen Union geförderten Projekt *R4eGov (Architecture for eGovernment)* haben sich neben dem Max-Planck-Institut für Informatik weitere 20 europäische Partner aus Forschung und Wirtschaft zusammengeschlossen, um eine Plattform zur europaweiten Zusammenarbeit öffentlicher Verwaltungen zu entwickeln.

Das Max-Planck-Institut für Informatik wird in diesem Projekt Methoden zur formalen Analyse der elektronischen Verwaltungsprozesse entwickeln und anwenden. Die formale Analyse der Prozesse ermöglicht Korrektheitsaussagen sowohl über Sicherheits- als auch Interoperabilitätsaspekte.

Sicherheit

Sicherheit hat im Rahmen des R4eGov-Projektes zwei Bedeutungen: Zum einen müssen sichere Kommunikationswege gefunden werden, um das gesamte System vor Unbefugten zu schützen. Zum anderen muss eine geeignete Zugriffskontrolle innerhalb des Systems verwaltet werden. Dazu gehören die Zugriffe auf Daten durch befugte Bearbeiter aber auch die Verwaltung dieser Zugriffsrechte unter Berücksichtigung möglicher Delegationen von Rechten oder Rechteentzug. Hierbei treten insbesondere Probleme bei transitiver Delegation und bei der Verwaltung von aus Einzelsystemen zusammengesetzten Zugriffskontrollsystemen auf. Über diese Art der statischen Zugriffskontrolle hinaus werden Systeme analysiert, die Rechte dynamisch verwalten. Ein Befugter kann beispielsweise das Recht haben, Dokumente sowohl zu verfassen als auch zu begutachten, jedoch kann er dasselbe Dokument nicht zugleich verfassen und begutachten.

Interoperabilität

Interoperabilitätsaspekte bilden den Schwerpunkt im R4eGov-Projekt. Selbst wenn Verwaltungsprozesse unabhängig voneinander korrekt ausgeführt werden können, so gilt deren Korrektheit nicht notwendigerweise im Zusammenspiel mit anderen Behörden. Die Untersuchung der Korrektheit wird dadurch erschwert, dass von den jeweiligen Organisationen und Institutionen nicht erwartet werden kann, dass sie ihre internen Prozesse vollständig offen legen. Hierzu werden am Max-Planck-Institut für Informatik Methoden entwickelt, die eine genaue Schnittstellenspezifikation ohne die Veröffentlichung interner Details ermöglicht. Neben der Korrektheit interoperierender Kontroll- und Informationsflüsse werden Synergieeffekte untersucht, die durch die Kollaboration von Institutionen mit gleichen Bearbeitern (Rollen) auftreten können. Diese Synergieeffekte führen einerseits zu Effizienzsteigerungen, andererseits ist die Zugriffskontrolle, insbesondere die dynamische, schwerer zu verwalten.

Unsere Lösung

Die Forscher des Max-Planck-Institutes für Informatik verwenden für die Modellierung und Analyse kollaborierender Verwaltungsprozesse formale Methoden wie zum Beispiel Petrinetze. Mit Hilfe dieser Methoden werden semiformale Modellierungssprachen, die in kommerziellen Workflow-Management-Systemen zum Einsatz kommen, formalisiert. Zur Implementierung serviceorientierter Architekturen werden Modelle kollaborierender Verwaltungsprozesse automatisch in vom Computer ausführbare Sprachen übersetzt. Prototypen dieser Werkzeuge werden bei Projektpartnern wie dem Bundeskanzleramt Österreich, Europol, Eurojust und dem Bundesgerichtshof Deutschland eingesetzt. :::

R4eGov
Making eGovernment work

KONTAKT

Jörn Freiheit
ABT . 2 Logik der Programmierung
Telefon +49 681 9325-220
Email freiheit@mpi-inf.mpg.de
Internet http://www.R4eGov.info

Individualisiertes Ranking von Webseiten

Die Fülle an Informationen im World Wide Web wächst rasend schnell. Leider aber nimmt nicht nur die Anzahl qualitativ hochwertiger Seiten zu. Auch die Zahl unseriöser, für den Benutzer meist unbrauchbarer Seiten steigt. Dies stellt Suchmaschinen, die das Web nach Informationen durchforsten, vor neue Herausforderungen. Dienste wie Google und Yahoo! haben jedoch nicht nur mit einer größeren Suchbasis, sondern auch mit gezielten Attacken auf ihre zugrundeliegenden Suchmechanismen zu kämpfen. Ein einfaches Beispiel hierfür sind Webseiten, die zahlreiche Wörter für den Benutzer unsichtbar in weißer Schrift auf weißem Hintergrund enthalten. Damit erhöhen die Betreiber die Trefferquote bei Suchabfragen.

Die Anbieter von Suchmaschinen sehen sich folglich nicht nur mit einer größeren Datenbasis konfrontiert, sondern müssen außerdem ständig die Mechanismen zur Bestimmung der Rangordnung von Suchresultaten verbessern. Wie Benutzerstudien belegen, betrachtet die Mehrheit der Benutzer lediglich die erste Resultatseite einer Suchanfrage. Wird das Suchinteresse durch das erste Resultat nicht hinreichend befriedigt, entscheiden sich die meisten Benutzer eher dafür, die Anfrage umzuformulieren. Dies verdeutlicht wie entscheidend es ist, sowohl Anzahl als auch Position von relevanten Resultatseiten, also die Präzision der Suche, zu optimieren. Nur so kann man dem Benutzer künftig eine zufriedenstellende Sucherfahrung bescheren.

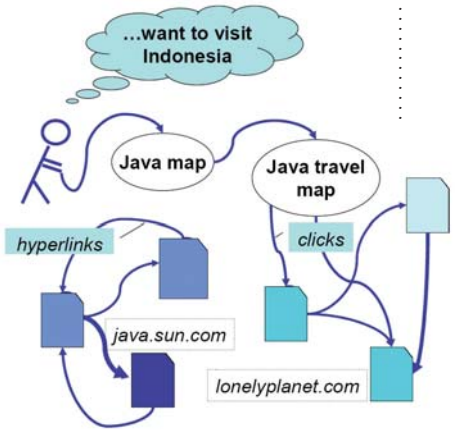
Bei der Erstellung des Rankings von Suchergebnissen verwenden Suchmaschinen heute meist eine Kombination verschiedener Ordnungskriterien. Einen besonders hohen Stellenwert haben dabei hyperlinkbasierte Qualitätsmaße oder solche Kriterien, die die inhaltsbasierte Ähnlichkeit zwischen Dokument und Anfrage berücksichtigen. Zu den hyperlinkbasierten Qualitätsmaßen gehört beispielsweise der so genannte **PageRank**-Wert, der von den Google-Gründern entwickelt wurde. Er spricht einer Webseite höhere Güte zu, wenn viele Seiten hoher Qualität auf sie zeigen. Verweist beispielsweise die Webseite der Max-Planck-Gesellschaft auf die Homepage eines Forschers, führt dies zu einem höheren PageRank-Wert als ein Link von der Homepage eines Schachfreundes.

Implizites Benutzerfeedback

Eine weitere Informationsquelle, die Rückschlüsse auf die Güte von Suchresultaten erlaubt, sind die Benutzer von Suchmaschinen selbst. Wann immer ein Benutzer eine Anfrage stellt und zurückgelieferte Dokumente aufgrund der präsentierten Kurzzusammenfassung gezielt besucht, gibt er für dieses Dokument implizit positives Feedback ab. Hier setzt unsere Forschung an. So beschäftigen wir uns damit, die Güte von Suchresultaten unter Zuhilfenahme von beobachtbarem Benutzerverhalten zu verbessern. Dabei adressieren wir mehrere Ebenen, indem wir Benutzerfeedback in Modelle zur Ermittlung von linkbasierten Qualitätsmaßen sowie in die Berechnung inhaltsbasierter Ähnlichkeit einfließen lassen. Außerdem betrachten wir Benutzerverhalten in verschiedenen Granularitäten, sowohl aggregiert über eine kohärente Benutzergruppe als auch individuell für einen einzelnen Benutzer.

Individualisiertes Ranking

Neben einer allgemeinen Verbesserung von Suchresultaten erlaubt implizites Benutzerfeedback auch, Suchresultate auf einen einzelnen Benutzer zurechtzuschneiden. Schließlich variiert die Interessenslage von Benutzer zu Benutzer meist deutlich. Auch die Einschätzung der Qualität einer Webseite unterliegt subjektiven Einflüssen. Da ein häufiger Einwand gegen eine Individualisierung von Suchergebnissen im Schutz der Privatsphäre begründet liegt, konzentriert sich unsere Arbeit auf ein Client-seitiges Anwendungsszenario. Das Protokollieren aller Interaktionen des Benutzers mit der Suchmaschine, sowie seines Verhaltens beim Surfen durch das Web, ermöglicht es uns, ein Benutzerprofil zu erstellen. Auf dessen Grundlage untersuchen wir Methoden, neue Anfragen gemäß der Benutzerinteressen zu erweitern und Suchergebnisse individuell neu zu ordnen. Können wir beispielsweise dem Profil eines Benutzers entnehmen, dass sein Interesse weniger der Java-Programmierung als vielmehr Reisen auf die Insel Java gilt, lassen sich Doppeldeutigkeiten in der Suchanfrage verringern. :::

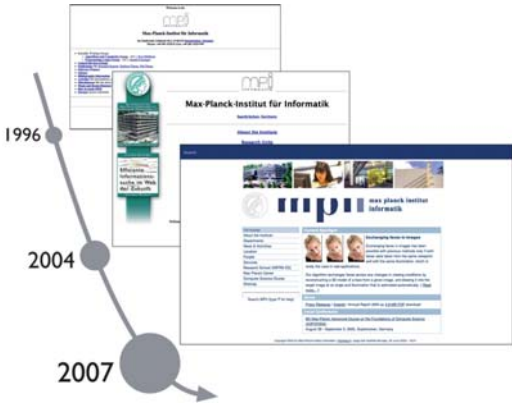


KONTAKT

Julia Luxenburger
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-521
Email julialux@mpi-inf.mpg.de

Zeitreise in Web-Archiven

Das World-Wide-Web (kurz: Web) verändert sich ständig: Inhalte kommen hinzu, werden modifiziert oder verschwinden. Diese Evolution spiegelt aktuelles Geschehen wider und ist ein wichtiges Zeitdokument, das es zu erhalten gilt. Dieser Aufgabe haben sich verschiedene Organisationen angenommen, darunter Nationalbibliotheken und, als wohl bekanntestes Beispiel, das in San Francisco ansässige Internet Archive (*http://www.archive.org*). In so genannten Web-Archiven speichern diese Institutionen Momentaufnahmen von Web-Seiten zu verschiedenen Zeitpunkten.



Die Web-Site des Max-Planck-Instituts für Informatik im Zeitverlauf, wie sie in einem Web-Archiv bewahrt ist

Der Zugriff auf archivierte Web-Seiten über solche Web-Archive ist heute weitgehend auf das Nachschlagen einzelner Seiten beschränkt: Anhand ihrer Adresse wird eine Liste verfügbarer Versionen angezeigt. Die einzelnen Versionen können dann eingesehen werden. Eine Suche anhand von Schlüsselwörtern wird nicht unterstützt oder aber behandelt Versionen als eigenständige Dokumente und ignoriert damit die Zeitachse.

An diesem Punkt setzt unsere gegenwärtige Forschung an. Wir beschäftigen uns mit Effektivitäts- und Effizienzaspekten von Zeitreise-Anfragen auf Web-Archiven. Unter einer Zeitreise-Anfrage verstehen wir hierbei eine aus Schlüsselwörtern bestehende Anfrage wie *„Prognosen zur Bundestagswahl"*, die um einen zeitlichen Kontext, beispielsweise *„August 2005"*, erweitert ist. Als mögliche Resultate für die Anfrage sollen nur solche Versionen betrachtet werden, die im genannten Zeitraum tatsächlich existiert haben.

Effektivität spiegelt sich in der Resultatsgüte wider, also darin, wie gut das vom Benutzer in der Anfrage formulierte Informationsbedürfnis befriedigt wird. Bestehende Information-Retrieval-Modelle (wie beispielsweise OKAPI BM25) sind nicht auf den Umgang mit Versionen und Zeitaspekten ausgelegt. Existieren mehrere Versionen einer Webseite, so werden diese als eigenständige Dokumente behandelt und können daher unabhängig voneinander in einem Anfrageresultat enthalten sein. Dies ist im Allgemeinen nicht erwünscht. Stattdessen soll dem Benutzer eine Liste der Webseiten im zeitlichen Kontext gezeigt werden, wobei verschiedene Versionen unter einem einzigen Eintrag zusammengefasst werden. In einer kürzlich erschienenen Arbeit wurden daher verschiedene Möglichkeiten definiert, wie sich die Relevanz einer Webseite aus den Relevanzwerten ihrer Versionen aggregieren lässt.

Eine Herausforderung bei der effizienten Bearbeitung von Zeitreise-Anfragen sind die zu bewältigenden Datenmengen. Selbst kleine Web-Archive erreichen leicht Größen im Bereich mehrerer Terabytes. Geeignete Indexstrukturen zur Unterstützung von Zeitreise-Anfragen müssen daher sowohl im Hinblick auf Platzverbrauch als auch auf Performanz hervorragend skalieren. Unser Ansatz baut auf einer existierenden, diesen Anforderungen entsprechenden Indexstruktur auf: dem invertierten Index. Der invertierte Index enthält pro Wort eine Liste. In dieser sind Informationen über das Vorkommen des Wortes in einzelnen Dokumenten enthalten. Wir erweitern die pro Vorkommen des Wortes gespeicherte Information um ein Gültigkeit-Zeitintervall und nutzen zusätzlich aus, dass sich zeitlich benachbarte Versionen einer Webseite häufig nur geringfügig unterscheiden – etwa weil lediglich Tippfehler korrigiert wurden. Unser Ansatz entfernt solche Unterschiede, sofern sie einen kalibrierbaren Schwellwert nicht überschreiten, und reduziert so den Platzbedarf drastisch, ohne Anfrageresultate merklich zu verfälschen.

Bei einer Anfragebearbeitung werden in der Regel viele Einträge gelesen, die irrelevant sind, da sie sich auf Versionen beziehen, die nicht im zeitlichen Kontext existiert haben. Dies ist ineffizient und vermeidbar. Um dieses Problem zu umgehen, materialisiert unser Ansatz pro Wort mehrere kürzere Indexlisten, die ausschließlich Informationen zu den im zugehörigen Zeitraum existierenden Versionen enthalten. Dadurch steigt wiederum der Platzbedarf, da Einträge eventuell über mehrere dieser kürzeren Indexlisten repliziert werden. Es besteht somit ein Zielkonflikt zwischen Effizienz und Platzverbrauch. Verfahren, die in diesem Zielkonflikt vermitteln und unter verschiedenen Vorgaben (beispielsweise einer vorgegebenen Beschränkung des Platzbedarfs) die optimalen kürzeren Indexlisten materialisieren, sind Gegenstand unserer gegenwärtigen Forschung. :::

KONTAKT

Klaus Berberich
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-521
Email kberberi@mpi-inf.mpg.de

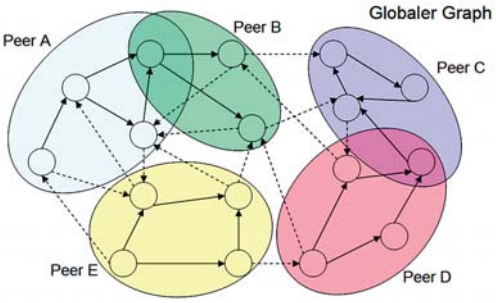
Verteilte Linkanalyse zur Autoritätsbewertung in Webgraphen und sozialen Netzen

Informationsmanagement in Internet-*Communities* ist ein brandaktuelles Thema. Viele Benutzer, die gewisse Interessen gemeinsam haben, schließen sich mit ihren persönlichen Daten sowie ihren Annotationen und Meinungen zu Daten anderer in einem elektronischen Portal zusammen und schaffen einen Mehrwert der insgesamt gesammelten Informationen. Die Interaktionen der Benutzer und die dynamisch entstehenden Untergruppen bilden eine Form von *sozialen Netzen*, deren Analyse interessante Erkenntnisse über die Community-Struktur und die Qualitätsverteilung der Informationen und Meinungen liefern kann. Eines der erfolgreichsten Community-Portale dieser Art ist der Photo-Sharing-Service *www.flickr.com*. Dort platzieren Millionen von Hobbyphotographen ihre besten Schnappschüsse. Benutzer annotieren und bewerten ihre Photos gegenseitig und erzeugen damit sowohl eine soziale Struktur als auch indirekt ein Ranking der besten Photographen und Photographien.

Eine mathematische Methode zur Analyse solcher vernetzten Strukturen ist die *Spektralanalyse* der zugrundeliegenden Graphen. Der bekannteste Sonderfall dieser Analysen ist die Berechnung von Googles *PageRank-Maß* für Webseiten. Dies ist ein Maß für die Autorität oder anfrageunabhängige Wichtigkeit von Webseiten, das auf der Interpretation von Hyperlinks als Empfehlung von Seiten beruht. Eine Seite hat umso höhere Autorität, je öfter sie via Link empfohlen wird und je höher die Autorität der Empfehlenden ist. Die Idee lässt sich auf die Analyse sozialer Netze verallgemeinern. Dazu muss man statt des Webgraphen – mit Webseiten als Knoten und Hyperlinks als Kanten – geeignete Graphstrukturen definieren, zum Beispiel mit Benutzern, Annotationen und Photos als Knoten und den Benutzerempfehlungen und -interaktionen sowie Zugehörigkeiten von Photos zur gemeinsamen Themengruppen als Kanten. Anschließend kann man mit Spektralanalysemethoden die wichtigsten Benutzer und Photos bestimmen.

Auch wenn die größten der derzeitigen Communities auf großen Server-Farmen in zentralisierter Form betrieben werden, drängt sich eigentlich eine verteilte, dezentralisierte Form des Community-Managements auf. Dabei würde jeder Benutzer seine Daten auf seinem eigenen Rechner halten, und die Rechner aller Benutzer würden Community-weit untereinander in einem *Peer-to-Peer-Netzwerk* kooperieren. Dies hätte den Vorteil, dass Benutzer in expliziter und flexibler steuerbarer Weise Herr über ihre Daten und deren Sharing und Weiterverbreitung bleiben und das soziale Netz weniger anfällig für etwaige Manipulationen wäre. Man würde für die oben erwähnten Autoritäts- und Qualitätsanalysen quasi auf eine Art der Basisdemokratie setzen. Zugleich hätte man damit potentiell riesige Speicher- und Rechenressourcen zur Verfügung, verteilt über die persönlichen Computer von Millionen von Benutzern. Gerade für die Spektralanalyse sehr großer Graphen – der Webgraph hat zum Beispiel mehr als 10 Milliarden Knoten und vermutlich Billionen von Kanten –, wäre eine verteilt-parallele Berechnung sehr nützlich, wenn sie denn mit der Anzahl der eingesetzten Rechner gut skalieren würde.

Der JXP-Algorithmus ist ein von uns entwickeltes, neues Verfahren, das eine solche massiv dezentralisierte Berechnung von Autoritäts- und Qualitätsmaßen in skalierbarer und effizienter Weise ermöglicht. Dabei trägt es der für selbstorganisierende Peer-to-Peer-Strukturen typischen Eigenschaft Rechnung, dass man die Datenverteilung nicht einfach top-down planen kann. Da Peers autonom über ihre Datensammlungen und ihre Beziehungen zu anderen Daten und Benutzern bestimmen, muss man mit beliebigen Untergraphen auf den einzelnen Peers rechnen. Insbesondere können sich die Untergraphen verschiedener Peers überlappen, und die Peers wissen a priori nicht, ob solche Überlappungen vorliegen und wie groß sie gegebenenfalls sind.



Beispiel eines Graphs, der auf mehrere Peers verteilt ist

JXP steht für *Juxtaposed Approximate PageRank*, weil es primär für die PageRank-Berechnung in einem Peer-to-Peer-Netz entwickelt wurde, weil es approximativ arbeitet und weil sein Grundprinzip auf bilateralen Rendezvous zwischen jeweils zwei Peers beruht. Die durch lokale Berechnungen und die Peer-Rendezvous approximierten Autoritätswerte konvergieren mathematisch beweisbar gegen die korrekten globalen Werte, die man erhalten würde, wenn man die gesamte Berechnung zentralisiert auf dem globalen Graphen durchführen würde. :::

KONTAKT

Josiane Xavier Parreira
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-508
Email jparreir@mpi-inf.mpg.de

OPTIMIERUNG

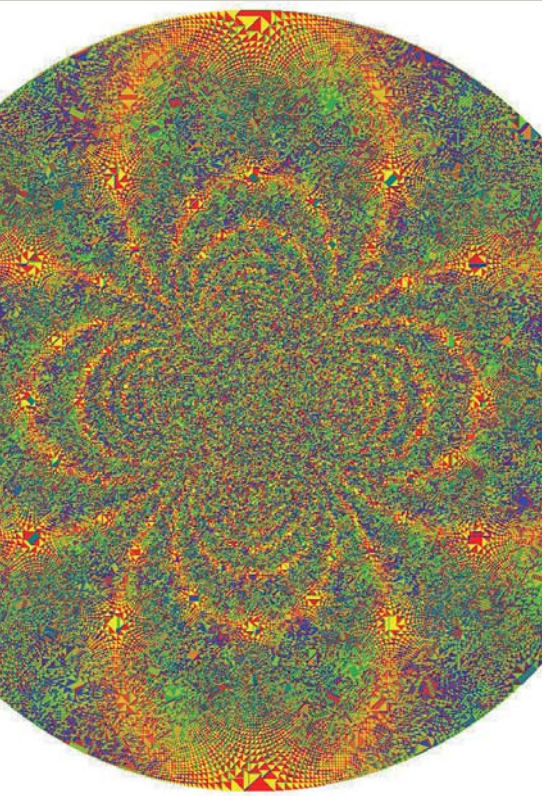
Die Optimierung ist heutzutage von sehr großer Bedeutung. Sie wird zum Beispiel eingesetzt, um den Bedarf an teuren Ressourcen wie etwa Arbeit einzusparen. Die Herausforderung an die Wissenschaft ist es, Verfahren zu entwickeln, mit denen sich schnell optimale Lösungen finden lassen oder zumindest solche, die nahe am Optimum liegen.

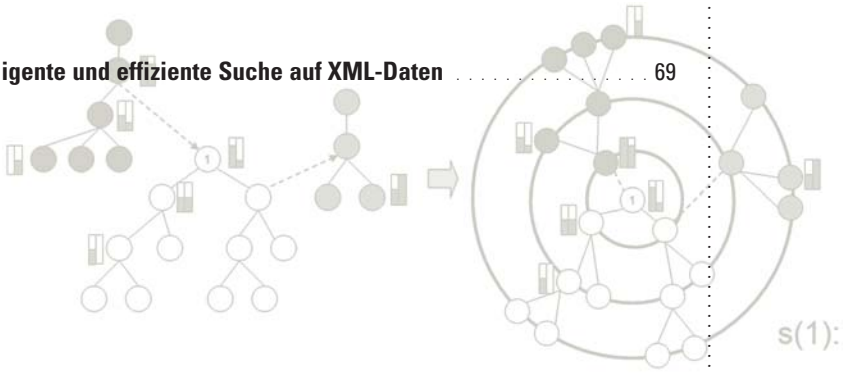
Gute Optimierungsverfahren sind für viele große Unternehmen entscheidend für die Wettbewerbsfähigkeit. Durch sorgfältige Planung können in Industrieprojekten oft große Beträge eingespart werden. Allerdings müssen in einem solchen Plan viele Bedingungen berücksichtigt werden. Das macht es für den Computer schwer, optimale oder zumindest sehr gute Pläne zu finden. Weiterhin kann die Datenmenge, die den Problemen zugrunde liegt, so groß sein, dass sie nicht von einem Menschen überschaut werden kann und auch für den Computer eine Herausforderung darstellt. Trotzdem werden Lösungen immer schneller benötigt.

Am Max-Planck-Institut für Informatik beschäftigen wir uns mit derartigen schwierigen Optimierungsproblemen. Zum einen entwickeln wir ausgefeilte Verfahren, um höchst effizient optimale Lösungen zu finden, auch wenn die zu behandelnde Datenmenge immer größer wird. Ist das zugrunde liegende Problem zu schwierig, um die optimale Lösung schnell zu berechnen, entwickeln wir Verfahren um zumindest eine Lösung zu finden, die nahe am Optimum liegt. Außerdem forschen wir an einem allgemeinen Optimierungsverfahren, der „ganzzahligen Programmierung“, mit dem sich viele Probleme einfach formulieren lassen und mit dem wir letztlich doch effizient optimale Lösungen berechnen können.

Da die Optimierung in sehr vielen verschiedenen Bereichen eine wesentliche Rolle spielt, untersuchen Wissenschaftler aus allen Forschungsgebieten, die am Max-Planck-Institut betrachtet werden, Optimierungsprobleme. Optimierung ist heutzutage ein wesentlicher Schlüssel zur Wettbewerbsfähigkeit von Unternehmen. Diese Bedeutung wird weiter zunehmen. :::

BEITRÄGE

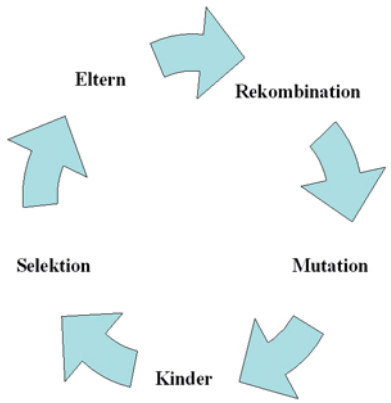




Theorie evolutionärer Algorithmen

Evolutionäre Algorithmen *(EAs)* sind allgemeine Suchverfahren, die in den Ingenieurdisziplinen und im Bereich der kombinatorischen Optimierung vielfältig angewendet werden. Diese Klasse von Lösungsverfahren folgt dem Vorbild der Evolution und dem Darwinschen Prinzip des *„survival of the fittest"*. Angelehnt an das natürliche Evolutionsprinzip, wird ein spezieller Lösungskandidat als *Individuum* und eine Menge solcher Kandidaten als *Population* bezeichnet. Eine so genannte Fitnessfunktion, welche vom gegebenen Problem abhängt, bewertet die Lösungskandidaten. Nach dem biologischen Prinzip wird aus einer Eltern-Population eine Kinder-Population erzeugt. Dies geschieht durch so genannte Veränderungsoperatoren, die das genetische Material der Eltern an die Kinder vererben. Die wichtigsten Operatoren sind in diesem Fall *Rekombination* und *Mutation*. Die Rekombination erzeugt gewöhnlich aus zwei Eltern ein Kind, während die Mutation zusätzlich dafür sorgt, dass das Kind weitere neue Eigenschaften aufweist. Von einer Startpopulation ausgehend, ist es das Ziel, für ein gegebenes Problem eine Menge möglichst guter Lösungskandidaten zu erhalten. Nachdem zunächst durch Veränderungsoperationen Kinder erzeugt worden sind, werden anhand der Fitnessfunktion aus der Eltern-Kind-Menge Individuen ausgesucht und eine so neue Elternpopulation geschaffen.

Evolutionäre Algorithmen werden insbesondere dann eingesetzt, wenn für ein gegebenes (neues) Problem kein guter problemspezifischer Algorithmus vorhanden ist. Es kann nicht erwartet werden, dass EAs speziell für ein Problem entworfene Lösungsverfahren übertreffen. Es ist also nicht Ziel der Forschung, zu zeigen, dass EAs problemspezifischen Algorithmen überlegen sind. Vielmehr steht im Vordergrund, die Arbeitsweise evolutionärer Verfahren zu verstehen.



Ablaufschema eines evolutionären Algorithmus

Forschungsschwerpunkt

Während evolutionäre Verfahren bereits vielfach erfolgreich angewendet werden, steckt das theoretische Verständnis dieser Algorithmen im Vergleich zu klassischen Algorithmen noch in den Kinderschuhen.

Hierbei wird untersucht, wie evolutionäre Suchverfahren in der Lage sind, bestimmte Probleme zu lösen, und mit welchen Strukturen EAs gut oder schlecht umgehen können. Das Hauptaugenmerk ist darauf gerichtet, wie viel Zeit EAs benötigen, um für ein gegebenes Problem eine optimale Lösung zu generieren. Da evolutionäre Algorithmen eine spezielle Klasse randomisierter Algorithmen sind, kann man auf eine große Zahl klassischer Analysemethoden zurückgreifen. Des Weiteren werden neue Analysemethoden entwickelt, die insbesondere evolutionäre Verfahren analysieren.

Es zeigt sich, dass evolutionäre Algorithmen oft gute Lösungen für bekannte Probleme finden. Sie sind bei vielen Problemen in der Lage, sich ähnlich wie problemspezifische Algorithmen zu verhalten. So wurde exemplarisch in mehreren Arbeiten untersucht, wie evolutionäre Algorithmen Eulerkreise in einem gegebenen Graphen finden können. Hierbei zeigt sich, dass die Kodierungsart möglicher Lösungen und die Wahl des verwendeten Mutationsoperators einen großen Einfluss auf die Laufzeit eines EAs haben. Mit rigorosen Laufzeitanalysen konnte gezeigt werden, dass sich gut gewählte EAs ähnlich effizient verhalten wie die besten problemspezifischen Algorithmen. Im Gegensatz hierzu kann eine eher allgemeinere Kodierung im Zusammenhang mit einem (für das Problem) schlechten Mutationsoperator zu einem vollkommen ineffizienten Verfahren führen. :::

KONTAKT

Benjamin Doerr
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-104
Email doerr@mpi-inf.mpg.de

Frank Neumann
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-117
Email fne@mpi-inf.mpg.de

Auf schnellstem Weg durchs Straßennetz

Ein computergestütztes Navigationsgerät gehört heute schon zur Auto-Standardausstattung. Trotzdem wissen viele Nutzer nicht, wie diese Geräte in wenigen Sekunden die schnellsten Routen in einem Straßennetzwerk berechnen.

Kürzeste Pfade in Graphen

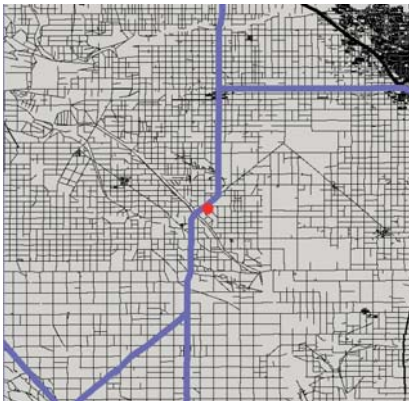
In der Informatiksprache modelliert man die Bestimmung einer schnellsten Reiseroute in einem Straßennetzwerk als eine *„Kürzeste-Pfadanfrage"* in einem Graphen. Die Kantengewichte des Graphen entsprechen dabei den Reisezeiten. Aufgrund seiner Effizienz wird sehr oft der so genannte Dijkstras Algorithmus zur Beantwortung von Kürzeste-Pfad-Anfragen in Graphen eingesetzt. Ein Beispiel aus der Praxis: Das US-amerikanische Straßennetzwerk besteht aus zirka 24 Millionen Kreuzungspunkten und 58 Millionen Streckenabschnitten. Ein handelsüblicher 2-GHz-Rechner beantwortet eine Routenanfrage innerhalb von etwa zehn Sekunden. Was sich zunächst sehr schnell anhört, würde auf eher rechenschwachen Navigationsgeräten zu Anfragezeiten im Minutenbereich führen. Undenkbar wäre es auch, auf dieser Basis einen Routenplaner im Internet aufzusetzen, da er mehrere hundert Anfragen pro Sekunde verarbeiten muss. Um effizienter arbeiten zu können, müssen die speziellen Eigenschaften eines Straßennetzwerks ausgenutzt werden.

Ausnutzung von Straßenhierarchien

Die hierarchische Struktur von Straßennetzwerken spiegelt sich schon in der Unterteilung in verschiedene Straßentyp-Kategorien wie Autobahnen, Landstraßen, Feldwege wider. Je weiter jemand verreist, umso häufiger benutzt er wichtige Straßen wie Autobahnen oder Schnellstraßen. Um nun die Routenplanung schneller zu berechnen, gibt es folgende Möglichkeit: Im Umfeld von Start und Ziel der Reise werden alle Straßenkategorien berücksichtigt, aber im mittleren Routenbereich nur wichtige Straßen wie beispielsweise Autobahnen. Tatsächlich lässt sich mit dieser Einschränkung die schnellste Route etwa 1000 mal effizienter berechnen – damit spielen sich Anfragen nun im Millisekundenbereich ab. Mit diesem Verfahren können selbst rechenschwache Plattformen schnell arbeiten und Server Hunderte von Anfragen pro Sekunde beantworten.

Unser neuer Ansatz: schnellste Routen via Transitknoten

Der folgende natürliche Umstand erlaubt es sogar, die Anfragezeiten nochmals um den Faktor 100 zu verbessern: Eine Person möchte mit dem Auto verreisen. Wie viele Möglichkeiten gibt es nun für sie, die nähere Umgebung zu verlassen? Das Ziel der Reise ist in diesem Zusammenhang unerheblich. Typischerweise existiert nur eine Handvoll solcher Routen.



Routen auf denen man ein kleines Dorf südlich von Fresno (USA) verlassen würde

Die Idee unseres Ansatzes ist es nun, wenige Kreuzungspunkte – so genannte Transitknoten – zu identifizieren, sodass jede längere Reise auf schnellstem Wege durch mindestens einen dieser Transitknoten verläuft. Außerdem kennt jeder Ort im Straßennetzwerk die wenigen für ihn relevanten Transitknoten. Mithilfe einer Tabelle, die die schnellsten Pfade zwischen allen Transitknoten aufführt, können damit Anfragen im Mikrosekundenbereich beantwortet werden, also um Größenordnungen schneller als zuvor.
Dieses Verfahren gehört zu den jüngsten Entwicklungen unserer Arbeitsgruppe. :::

KONTAKT

Holger Bast
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-120
Email bast@mpi-inf.mpg.de

Stefan Funke
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-108
Email funke@mpi-inf.mpg.de

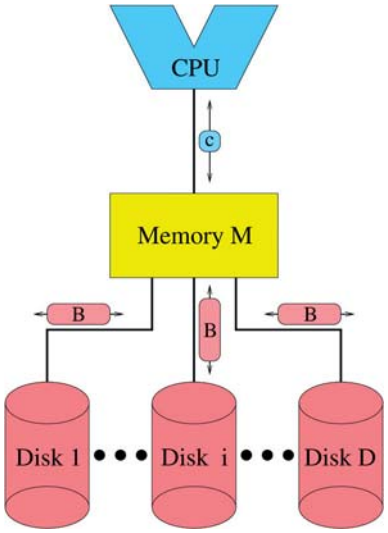
Algorithmen für Speicherhierarchien

Zwischen Theorie und Praxis

Im Forschungsbereich Algorithmen und Datenstrukturen für Speicherhierarchien betreiben wir Grundlagenforschung zu einem praxisrelevanten Thema, das für viele Anwendungen der Ingenieurs- und Naturwissenschaften stetig an Bedeutung gewinnt: In Algorithmen, die große Datenmengen verarbeiten, hängen die Kosten für Speicherzugriffe davon ab, wo die Daten abgelegt sind. Während der Wartezeit für einen Hauptspeicherzugriff können moderne Prozessoren leicht 1.000 arithmetische Operationen durchführen; für Festplattenzugriffe kann dieser Faktor sogar auf mehrere Millionen ansteigen. Im klassischen Algorithmenentwurf hingegen werden für alle Operationen und Speicherzugriffe identische Kosten angesetzt *(Von-Neumann-Modell)*. Sobald die Daten nicht mehr in den Hauptspeicher passen, kommt es oftmals zu verheerenden Folgen für die Programmlaufzeiten.

Das Externspeicher-Modell

Die folgende einfache Erweiterung zum so genannten Externspeicher-Modell liefert oft schon brauchbare Vorhersagen: Ein Von-Neumann-Rechner mit begrenztem internen Speicher für M Maschinenworte ist mit einem externen Speicher verbunden. In einem Ein-/Ausgabeschritt kann ein Block mit B nebeneinander stehenden Maschinenworten zwischen internem und externem Speicher bewegt werden. Wegen des großen Geschwindigkeitsunterschieds zwischen Haupt- und Festplattenspeicher spielen meist diese beiden Ebenen die Rolle von internem und externem Speicher. Neben der Anzahl der Operationen und Zugriffe auf den internen Speicher analysieren wir deshalb zusätzlich die Anzahl der Ein-/Ausgabeschritte. Ziel ist es, diese zu minimieren.



Das Externspeicher-Modell

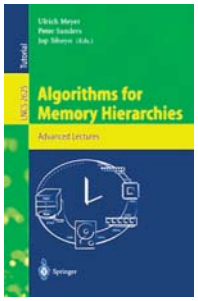
Goldene Regeln

In diesem Zusammenhang ist es oft notwendig, die Speicherzugriffsmuster im Vergleich zum jeweils besten Von-Neumann-Algorithmus radikal umzugestalten. Wie genau, hängt vom konkreten Problem ab. Es gibt aber einige wenige goldene Regeln. Ist ein Datenelement zum Beispiel einmal im internen Speicher abgelegt, sollte es bestmöglich und umfassend genutzt werden, bevor es wieder ausgelagert wird. Weiterhin sollten die Zugriffe auf den Externspeicher ein hohes Maß an Lokalität aufweisen: wenn beispielsweise eine externe Speicherstelle gelesen wird, sollten die benachbarten B Maschinenworte (die durch den Blockzugriff automatisch mitgeliefert werden) ebenfalls von Interesse sein. Ebenso sollten unstrukturierte Zugriffe vermieden werden. Dies ist besonders schwierig bei vielen der so genannten Graphenalgorithmen, die hinter zahlreichen Anwendungen stehen. Dies steht im Zusammenhang mit Querverweisen zwischen Daten.

Anwendungen im WWW

Ein Beispiel für einen sehr großen Graph ist das World Wide Web (WWW). Typische Graphenprobleme im Web beziehen sich auf seine Zusammenhangsstruktur (welche Seiten sind beispielsweise verzahnt, welche isoliert) und seine Distanzen (kürzeste Pfade: Wie viele Verweise müssen verfolgt werden, um von einer bestimmten Seite zu einer anderen zu gelangen?). Da Verweise im allgemeinen nicht-lokal sind, müssen für effiziente Ein-/Ausgabe-Graphenalgorithmen oft unorthodoxe Strategien entworfen werden. Zwei fundamentale Formen des Kürzesten-Pfad-Problems (Breitensuche, kürzeste Pfade bei einem Startpunkt und beschränkten positiven Gewichten) konnten wir entscheidend verbessern, indem wir durch einen – eigentlich unerwünschten – größeren Berechnungsaufwand viele unstrukturierte Einzelzugriffe einsparen konnten: die Implementierung unseres Ansatzes ist bis zu 1.000 mal schneller als die besten bisherigen Verfahren.

Einen Überblick über Algorithmen und Datenstrukturen für zwei- und mehrstufige Speicherhierarchien liefert unser Buch *Algorithms for Memory Hierarchies*, Springer, 2003. :::



KONTAKT

Ulrich Meyer
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-100
Email umeyer@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/departments/d1/areas/models.html

Deterministische Irrfahrten

In der Natur spielt der Zufall eine zentrale Rolle. Auch in der Informatik ist er der Grundbaustein für viele Anwendungen, beispielsweise für evolutionäre Algorithmen.

Zufällige Irrfahrten

Eine zufällige Irrfahrt *(englisch: Random walk)* ist ein Weg, der dadurch entsteht, dass man an jeder Kreuzung in eine zufällige Richtung geht. Diese Richtungsentscheidungen sind unabhängig voneinander und jede Richtung ist gleich wahrscheinlich. Die Eigenschaften solcher Irrfahrten wurden seit Anfang des letzten Jahrhunderts eingehend untersucht. Daraus haben sich einfache, aber hocheffiziente Algorithmen entwickelt. Ein Beispiel dafür ist die Erkundung von unbekanntem Terrain mit einer zufälligen Irrfahrt.

Quasi-Zufällige Irrfahrten

Wir untersuchen die Frage, wie viel Zufall bei einer solchen Irrfahrt tatsächlich nötig ist. Dazu betrachten wir ein *„quasi-zufälliges"* Analogon der zufälligen Irrfahrt, genannt *„Propp-Maschine"*. Dieses vor einigen Jahren von Jim Propp vorgeschlagene Modell besitzt an jedem Kreuzungspunkt einen Zeiger. Richtungsentscheidungen werden nun deterministisch getroffen, indem der Weg der Zeigerrichtung folgt. Um dennoch alle Richtungen gleichmäßig zu bedienen, wird ein passierter Zeiger in einer vorgeschriebenen Reihenfolge weiter gedreht.

Besetzte Felder nach einer zufälligen Irrfahrt von 1600 Partikeln um einen Kondensationspunkt

Ein-Knoten-Diskrepanz

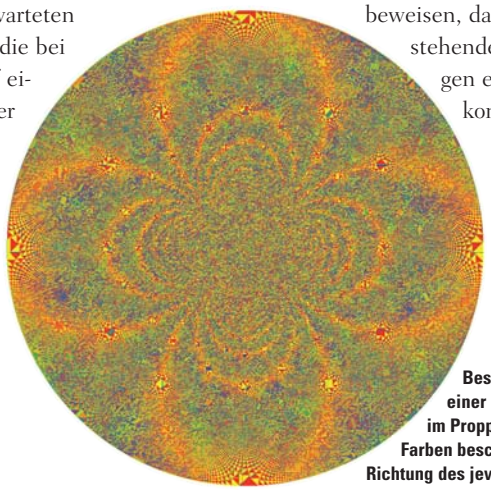
Dieses deterministische Modell ähnelt der bekannten zufälligen Irrfahrt in überraschender Weise. Als Beispiel betrachten wir ein unendlich großes Schachbrett, auf dem Spielsteine sich in jedem Schritt zu einem der vier benachbarten Felder bewegen. Wir haben nun untersucht, wie sich beide Modelle ver- halten, wenn man eine große Zahl Steine auf den schwarzen Feldern des Schachbretts verteilt und sie eine feste Anzahl von Schritten laufen lässt. Wir konnten zeigen, dass der Unterschied zwischen der erwarteten Anzahl der Spielsteine, die bei zufälliger Bewegung auf einem Feld enden, und der Anzahl der Spielsteine, die die Propp-Maschine dahin befördert, maximal acht ist. Diese Schranke ist unabhängig von der Gesamtzahl der Spielsteine, der Laufzeit und der anfänglichen Zeiger-Ausrichtung.

Wachstumsprozess

Zufällige Irrfahrten beschreiben auch physikalische Wachstumsprozesse wie die Kondensation oder die Entwicklung eines Blitzes. Wir betrachten ein ähnliches Modell, bei dem Partikel aus einem einzelnen Kondensationskeim austreten und nach einer zufälligen Irrfahrt an der ersten freien Position, die sie erreichen, liegen bleiben. Dadurch entsteht eine ausgefranste runde Fläche von Partikeln um den Kondensationskeim. Es ist bekannt, dass diese Form gegen einen Kreis konvergiert und dass die erwartete Fluktuation höchstens etwa mit der sechsten Wurzel der Anzahl der Partikel zunimmt.

Propp-Kreis

Analog zu dem beschriebenen Kondensationsprozess der zufälligen Irrfahrt kann man die Partikel auch nach den Zeigern der Propp-Maschine lenken. Für dieses Modell konnte man ebenso beweisen, dass die entstehende Form gegen einen Kreis konvergiert.



Besetzte Felder von einer Million Partikeln im Propp-Modell. Die Farben beschreiben die letzte Richtung des jeweiligen Zeigers.

Experimentell zeigt sich jedoch, dass der Unterschied zwischen Inkreis und Umkreis nur eine sehr kleine Konstante zu sein scheint. Es ist überraschend, dass dieses einfache Modell eine so komplexe und zugleich regelmäßige Struktur erzeugen kann. :::

KONTAKT

Benjamin Doerr
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-104
Email doerr@mpi-inf.mpg.de

Tobias Friedrich
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-126
Email tfried@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/departments/d1/quasirandom/

Bewegungsanalyse bekleideter Personen aus Videodaten: Ein Röntgenblick durch Kleidungsstücke

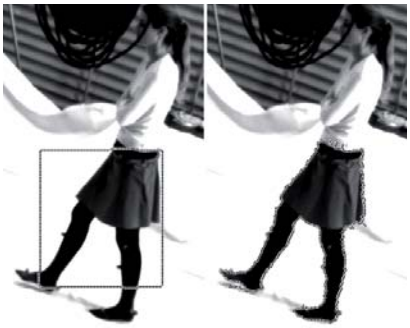
Markerfreie menschliche Bewegungsanalyse

Motion Capture bezeichnet die Aufnahme menschlicher Bewegungen und deren Analyse. Anwendungsbeispiele findet man in der Medizin. Hier wird Motion Capture eingesetzt, um orthopädische Krankheiten zu diagnostizieren oder den Heilungsverlauf zu dokumentieren. In der Sportwissenschaft nutzt man Motion Capture, um die Leistung von Profi-Sportlern zu verbessern, in der Filmindustrie, um Avatare zu animieren.

Zurzeit verwendet man meist so genannte markerbasierte Verfahren: An einer Person werden zunächst (reflektierende) Marker angebracht. Kameras verfolgen dann die markierten Personen in speziellen Aufnahmevorrichtungen (beispielsweise unter Stroboskoplicht).

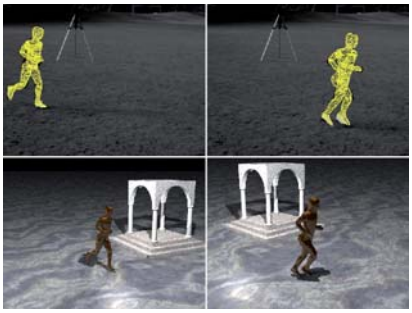
Ein dreidimensionaler Scan einer Versuchsperson

Doch für viele Anwendungen wie zum Beispiel im Außenbereich, im Schwimmbad oder in hochdynamischen Szenen sind solche Marker ungeeignet. Gerade ältere Patienten empfinden derartige Marker zudem als unangenehm. Folglich bewegen sie sich bei Aufzeichnungen unnatürlich. Deshalb ist die markerfreie Verfolgung von Personen in Bildsequenzen ein aktuelles und attraktives Forschungsgebiet, das am Max-Planck-Institut für Informatik neben anderen offenen Problemstellungen bearbeitet wird. In diesem Projekt werden zunächst Personen in einem Laserscanner rekonstruiert. In das Modell wird ein virtuelles Skelett mit Gelenken eingebaut, sodass es animiert werden kann. Eine Person wird anschließend von verschiedenen synchronisierten Kameras beobachtet. Um das Modell an die Bilddaten anzupassen, wird zunächst das Bild segmentiert.



Segmentierung des Bildes durch Evolution einer Levelsetfunktion

Hierzu werden so genannte Level-set-Funktionen eingesetzt, die mit einem 3D-Shape-Prior versehen werden. Anschließend wird die Oberfläche mit einem Registrierungsverfahren an die Bildkonturen angepasst. In einer Reihe von Experimenten konnte gezeigt werden, dass unser markerfreies Motion-Capture-System die gleiche Fehlertoleranz aufweist wie markerbasierte Verfahren. Zusätzlich haben wir die Möglichkeit, auch Personen im Außenbereich zu verfolgen.



Personenverfolgung im Außenbereich

Bewegungsanalyse bekleideter Personen

Das bisherige Verfahren erfordert, dass die Personen enge Ganzkörperanzüge tragen. Nur so lassen sich die Körperkonturen gut an die Bilddaten anpassen. Dies ist insofern ungünstig, als bei der Analyse von Sportbewegungen (beispielsweise von Fußballspielern) Trikots getragen werden. Und ältere Personen fühlen sich in Ganzkörperanzügen ähnlich unwohl wie mit angebrachten Markern. Deshalb wird in einem der Forschungsprojekte eine Kleidungssimulation in den Optimierungsprozess integriert: So lässt sich aufgrund der Kleidung und der sichtbaren Körperteile auf die verdeckten Bereiche schließen.

Die verfolgte Person sowohl mit Rock als auch mit Shorts

Dabei werden sowohl geometrische als auch physikalische Kleidungssimulationen eingesetzt (über Mass-spring-Modelle). Mit dem Optimierungsverfahren kann man also quasi durch die Kleidung blicken und so auf die Konfiguration der Gelenke schließen. :::

KONTAKT

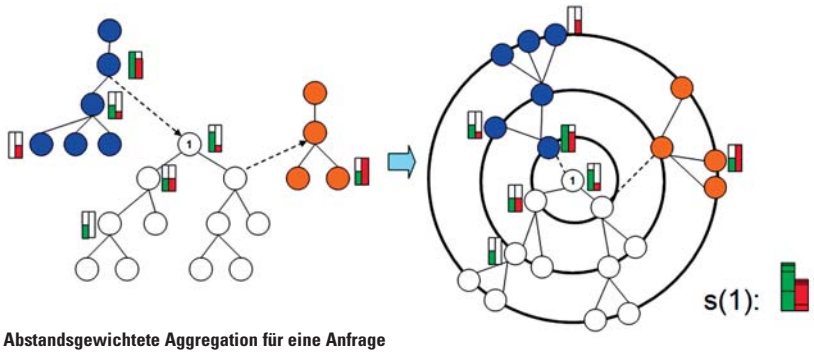
Bodo Rosenhahn
ABT. 4 Computergraphik
Telefon +49 681 9325-417
Email rosenhahn@mpi-inf.mpg.de

Intelligente und effiziente Suche auf XML-Daten

Im Web der Zukunft wird Information nicht mehr in Formaten wie HTML vorliegen, da diese kaum strukturiert und vorwiegend zur Präsentation gedacht sind. Ausdrucksstärkere Formate wie XML *(eXtensible Markup Language)* werden sie ablösen. Mit dieser Metasprache kann man Information strukturiert in einer graphartigen Form repräsentieren und mit Semantik annotieren. Dabei ist es allein wegen der schieren Größe des Webs unwahrscheinlich, dass Information überall gleichartig dargestellt wird. Stattdessen gibt es eine bunte Mischung aus den verschiedensten Annotationen. Dokumente, die durch Links verbundenen sind, machen das Web zu einem engmaschigen Informationsnetz.

Sobald zusätzliche Annotationen ausgenutzt werden sollen, stellt die Suche auf großen, heterogenen Sammlungen von XML-Dokumenten, für die das Web der Zukunft neben Intranets oder Online-Bibliotheken nur ein Beispiel ist, große Herausforderungen an Suchmaschinen. Einfache Wortanfragen, wie sie im heutigen Web benutzt werden, können dies nicht leisten. Stattdessen sind neuartige Anfragesprachen erforderlich, die Bedingungen an die Struktur und Annotation von Daten stellen. Benutzer, die Anfragen formulieren, können wegen der Heterogenität der Datensammlungen allerdings immer nur vermuten, wie die Datenstruktur letztlich aussieht. Die Suchmaschine muss versuchen, Trefferseiten zu finden, die eine größtmögliche Übereinstimmung mit der Anfrage sowohl in der dargestellten Information als auch in der Struktur aufweisen. Um dieses schwer zu lösende Problem automatisch bearbeiten zu können, muss man wissen, welche Annotationen ähnliche Daten beschreiben. Unsere Suchmaschinen TopX und SphereSearch verwenden dazu Ontologien, die Beziehungen zwischen sprachlichen Konzepten wiedergeben, beispielsweise Ober- und Unterbegriffe sowie Synonyme. Ein Ähnlichkeitsmaß für die Ontologie, das auf der Korrelation von Begriffen in einem großen Korpus basiert, hilft, auch Annotationen zu finden, die in der Ontologie innerhalb eines kleinen Abstands um das Gesuchte liegen. Die XPath-artige Anfrage *//article[//~person Max Planck]*, die nach Artikeln über die Person Max Planck sucht, findet so zum Beispiel auch Dokumente, in denen Max Planck als Wissenschaftler oder Physiker annotiert ist, nicht aber Dokumente, in denen von Max-Planck-Instituten die Rede ist. Um die Wichtigkeit eines potenziellen Ergebnisses für eine Anfrage zu ermitteln, spielt neben der semantischen Ähnlichkeit der Annotationen auch der textuelle Inhalt eine Rolle. Dabei wird jeder Teil des Dokumentes (also zum Beispiel Abschnitte) numerisch bewertet – und zwar auf Basis der Häufigkeit der Anfrageterme und der Länge des Dokumentteils. Die Ergebnisse werden dann zusätzlich entlang der Graphstruktur der Dokumentsammlung abstandsgewichtet aggregiert. Rückmeldungen des Benutzers über die Güte von Suchergebnissen werden zusätzlich genutzt, um Anfragen zu verfeinern und so die Ergebnisqualität zu verbessern.

Da die zugrunde liegenden Datensammlungen sehr groß sein können, spielt die Effizienz der eingesetzten Algorithmen eine große Rolle. Der Wunsch ist es, zügig die Dokumente mit richtiger Struktur zu finden. Dies lässt sich durch Algorithmen auf Graphen lösen. Dazu haben wir zwei Indexstrukturen entwickelt, die auf dieses Problem spezialisiert sind. Der HOPI-Index speichert kompakt die Menge aller Verbindungen in Graphen und erlaubt effiziente Zugriffe auf einzelne Pfade; das FliX-Framework errechnet die gleiche Information mit etwas höheren Kosten zur Laufzeit, kann dafür aber auf aufwändige Vorberechnungen wie beim HOPI-Index weitgehend verzichten. Zusätzlich gilt es, die Berechnung unnötiger Zwischenergebnisse zu vermeiden und die besten Treffer möglichst rasch zu erhalten. Wir erweitern dazu etablierte Ansätze für unstrukturierte Daten so, dass sie auch Strukturbedingungen effizient unterstützen. :::



Abstandsgewichtete Aggregation für eine Anfrage aus zwei Wörtern

KONTAKT

Ralf Schenkel
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-504
Email schenkel@mpi-inf.mpg.de

SOFTWARE

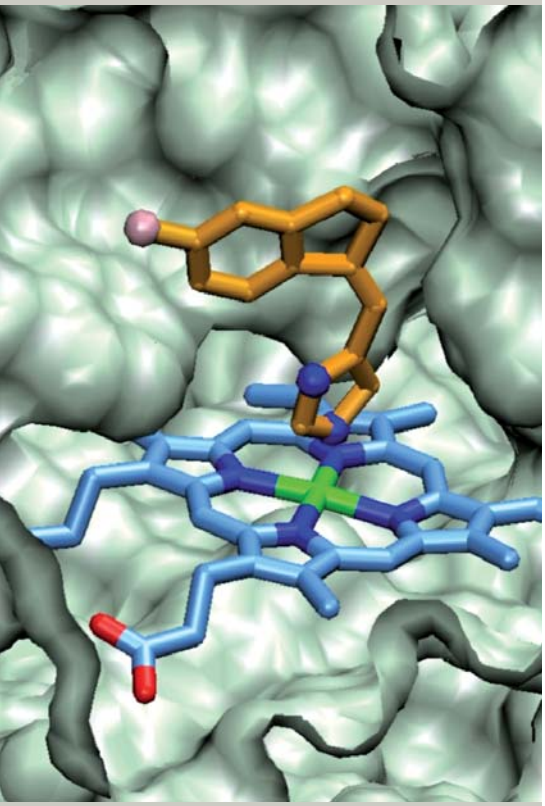
Informatik ist einerseits eine Grundlagenwissenschaft, die sich mit universellen Berechnungs- und Problemlösungsmethoden und deren fundamentalen Eigenschaften wie Korrektheit und Komplexität beschäftigt. Andererseits hat sie aber auch den Charakter einer Ingenieurwissenschaft und lebt von den vielfältigen Berührungspunkten mit verschiedensten Anwendungen. Die Grundlagenforschung trägt entscheidend zur Entwicklung von mathematischen Modellen und neuen Algorithmen für einsatzfähige Softwaresysteme bei. Progammbibliotheken und -systeme, die als Open-Source-Software mit kostenfreien Lizenzen Forschern und Anwendern zur Verfügung gestellt werden, bringen enormen Nutzen für andere Wissenschaftler, beeinflussen die langfristige Entwicklung der IT-Industrie und liefern Feedback für die weitere Weichenstellung der eigenen Forschungsarbeiten.

Am Max-Planck-Institut für Informatik wird diese Philosophie seit Gründung des Instituts mit großem Erfolg verfolgt. Alle Abteilungen und Gruppen arbeiten daran, die Ergebnisse ihrer Grundlagenarbeiten in praxisrelevante Softwaresysteme umzusetzen und für Wissenschaft und Industrie verfügbar zu machen. Es gibt eine beachtliche Anzahl am Institut entwickelter Prototypsysteme aus allen Bereichen vom Theorembeweisen und der Algorithmik bis zur Bioinformatik, Sicherheit, Visualisierung und Web-Suche, die ihren Weg in die Wissenschaftsgemeinde gefunden haben und an vielen Orten in der Welt für Forschungsarbeiten benutzt werden. Dabei handelt es sich überwiegend um kostenfreie Open-Source-Software, in einigen Fällen wurden Startup-Firmen gegründet, die die Software weiter entwickeln und vertreiben. Beispiele für eine kommerzielle Nutzung unserer Software sind die LEDA Bibliothek für effiziente Algorithmen aus der Abteilung Mehlhorn, der BiQ-DNA-Methylation-Analysator aus der Abteilung Lengauer, Sicherheitssoftwarekomponenten aus der Forschungsgruppe Scheffer und der Waldmeister-Gleichheitsbeweiser aus der Forschungsgruppe Weidenbach.

Die geschickte Umsetzung mathematischer Modelle und Algorithmen in lauffähige Software ist zudem selbst ein wichtiger Forschungsgegenstand. Algorithmen, die abstrakt sehr gute, mathematisch analysierbare Laufzeit- und Speicherplatzeigenschaften haben, so genannte asymptotische Komplexitätsmaße, sind in der Implementierung auf modernen Rechnern und verteilten Cluster-Systemen nicht automatisch effizient. Eigenschaften von Prozessoren, Magnetplatten und Netzwerken sowie die Charakteristika realer Daten müssen im Engineering geeignet berücksichtigt werden, um einsatztaugliche Softwaresysteme zu bauen.

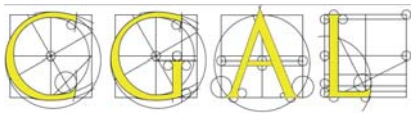
Das Zusammenspiel vieler Algorithmen in einem kompletten System und die Erweiterbarkeit und Selbstorganisation von Software werfen selbst schwierige wissenschaftliche Fragestellungen auf, die teilweise am Max-Planck-Institut für Informatik untersucht werden, teilweise aber über den Themenbereich des Instituts hinausgehen und innerhalb der Max-Planck-Gesellschaft vom neuen Max-Planck-Institut für Softwaresysteme weiter verfolgt werden. :::

BEITRÄGE



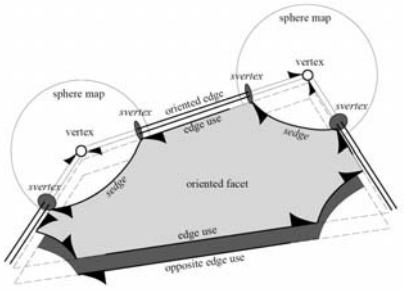
CGAL: Algorithmen für geometrische Probleme

CGAL ist eine hoch-modulare C++-Bibliothek für geometrische Datenstrukturen und Algorithmen. Dazu zählen beispielsweise verschiedene 2D- und 3D-Triangulierungen, Voronoidiagramme, Mengenoperationen auf Polygonen und Polyedern, konvexe Hüllen und polyedrische Flächenrekonstruktion. Es ist prinzipiell schwierig, geometrische Algorithmen korrekt zu implementieren. Ein Hauptgrund ist, dass Computer nur mit großem Aufwand wirklich exakt rechnen. Eine Folge sind die üblichen Rundungsfehler, die die Korrektheitsbeweise der geometrischen Algorithmen ungültig machen. Dass es sich hier nicht um ein rein theoretisches Problem handelt, zeigen immer wieder falsche Resultate, Crashs oder nicht-terminierende Programme.



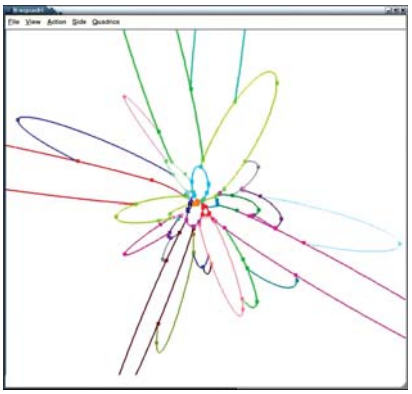
Um auf die Probleme hinzuweisen, veröffentlichte die einflussreiche Computational-Geometry-Impact-Task-Force von Bernard Chazelle und anderen Wissenschaftlern 1996 ihren Report *„Advances in Discrete and Computational Geometry"*, der von der American Mathematical Society herausgegeben worden ist. Der Bericht fordert in seiner ersten Empfehlung die Entwicklung und Verbreitung eines geometrischen Programmcodes. CGAL wurde bereits 1995 ganz in diesem Sinne initiiert und ist seitdem höchst erfolgreich.

CGAL wird aktuell vom CGAL-Open-Source-Projekt entwickelt, dem verschiedene Forschungsinstitute weltweit angehören. Unser Ziel ist es, den Benutzern in der Industrie und der Lehre die wichtigsten Lösungen und Methoden aus der Algorithmischen Geometrie zur Verfügung zu stellen. Wir legen insbesondere Wert auf Korrektheit, Vollständigkeit und Effizienz – sowohl theoretisch als auch praktisch. CGAL ist als Open-Source lizensiert. Zusätzlich sind kommerzielle Lizenzen der GeometryFactory erhältlich, einer Startup-Firma, die im Januar 2003 aus dem Projekt heraus für Support und Lizensierung gegründet wurde. Beispiele für kommerzielle Anwendungsbereiche von CGAL sind Luftbildverarbeitung (BAE Systems und Leica Geosystems, USA), Bildverarbeitung (Toshiba, Japan) und strukturelle Geologie (Midland Valley, UK, Agip, Italien), CAD/CAM, Biochemie und andere. Das Max-Planck-Institut für Informatik ist einer der CGAL-Entwickler. In diesem Rahmen wurde es, zunächst mit europäischer Förderung durch das gleichnamige Projekt unterstützt. Seit Projektabschluss werden die Arbeiten unter anderem in EU-Projekten weitergeführt



Nachbarschaften entlang einer Kante eines 3D-Nef-Polyeders

Wir haben am ersten geometrischen Kernel und vielen grundlegenden Entwurfsentscheidungen mitgearbeitet, die in der ersten Phase getroffen wurden. Danach lagen unsere Schwerpunkte auf dem d-dimensionalen geometrischen Kernel und Algorithmen für konvexe Hüllen in 2-, 3- und d-Dimensionen. Darüber hinaus entwickelten wir Methoden zur Polygonzerlegungen in konvexe Teilpolygone und für Mengenoperationen auf 2D-Nef-Polygonen und 3D-Nef-Polyedern.



Lower Envelope von 200 Ellipsoiden

Aktuell beteiligen wir uns an der Generalisierung von 2-dimensionalen Arrangements. Darüber hinaus portieren und vervollständigen wir die in Exacus entwickelten algebraischen Grundlagen nach CGAL. Dadurch wird es möglich, die große Sammlung an effizienten Methoden aus Exacus ebenfalls in CGAL zu integrieren. Die Methoden hatten sich bereits im Vorfeld bei der Untersuchung von Kurven (Algebraische Kurven beliebigen Grades) und gekrümmten Flächen (Quadriken) als erfolgreich erwiesen. Die Sammlung kann somit der großen Nutzergemeinde unmittelbar zur Verfügung gestellt werden. In Zusammenarbeit mit der Universität Tel-Aviv (Israel) entstand die exakte und effiziente Berechnung von *„Lower Envelopes"* von Quadriken. Die Entwicklung der algebraischen und geometrischen Kernels erfolgt in enger Kooperation mit unseren Projektpartnern aus Frankreich (Inria) und Griechenland (Universität Athen). :::

KONTAKT

Eric Berberich
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-100
Email eric@mpi-inf.mpg.de

Michael Sagraloff
ABT. 1 Algorithmen und Komplexität
Telefon +49 681 9325-100
Email msagralo@mpi-inf.mpg.de
Internet http://www.cgal.org

pfstools – Bearbeitung von HDR-Bildern und Video

Die meisten traditionellen Bildverarbeitungsbibliotheken speichern Bildpixel mit eingeschränkter Genauigkeit. Zusätzlich können sie nur begrenzt Farben kalibrieren. Um diese Probleme zu beheben, entwickelten wir ein High-Dynamic-Range-(HDR)-Bildverarbeitungs-System, das aus einem Paket von verschiedenen Kommandozeilenprogrammen besteht. Dieses System ermöglicht es, HDR-Bilder und -Videos zu lesen, zu schreiben und zu verändern. Da das Programmpaket unsere aktuellen Forschungsprobleme lösen sollte, standen während der Entwicklung Schlichtheit und Flexibilität im Vordergrund. Da die Software für zahlreiche Forschungsprojekte erfolgreich angewendet werden konnte, beschlossen wir, sie als ein Open-Source-Projekt unter der Lizenz (GPL) der Allgemeinheit zugänglich zu machen.

Die Hauptfunktion der Software ist, verschiedene Bildverarbeitungs- und Bildformatsbibliotheken, wie zum Beispiel *ImageMagick*, *OpenEXR* und *NetPBM*, in ein Gesamtsystem zu integrieren. Damit das Anwendungsspektrum möglichst flexibel bleibt, konstruierten wir *pfstools* auf Basis folgender Konzepte:

- Bilder/Videobilder sollen eine beliebige Anzahl von Bildkanälen besitzen dürfen; neben Farbe sollen zusätzlich auch Tiefe, Transparenz und Textur werte gespeichert werden können.

- Jeder Bildkanal soll mit einer großen Genauigkeit gespeichert werden, die durch die Verwendung von *floating point*-Nummern gewährleistet wird. Wenn möglich, sollten die Daten farbkalibriert sein und eine Genauigkeit erlauben, die größer als die menschliche Wahrnehmung ist.

- Helligkeit soll in der physikalischen Einheit cd/m² gespeichert werden, um sowohl das menschliche Tag- als auch das Nacht-Sehen simulieren zu können.

- Der Benutzer soll die Möglichkeit haben, zusätzliche anwendungsabhängige Information (beispielsweise die Farbkoordinaten des Weißpunkts) speichern zu können.

pfstools besteht fast ausschließlich aus Kommandozeilen-Programmen und bietet nahezu keine graphische Benutzeroberfläche. Die Hauptkomponenten sind Programme zum Lesen und Schreiben von Bildern in allen bedeutenden HDR- und LDR-Formaten (zum Beispiel OpenEXR, Radiance's RGBE, logLuv TIFF, 16-bit TIFF, PFM, JPEG, PNG und anderen), Programme für einfache Bildmanipulation (Rotation, Skalierung oder Ausschneiden), ein HDR-Bildbetrachtungsprogramm sowie eine Bibliothek zum Vereinfachen von Lese- und Schreib-Routinen in C++. Das Paket enthält außerdem eine Schnittstelle für *GNU Octave* und *matlab*. Die typische Verwendung von *pfstools* beinhaltet die Ausführung verschiedener Programme, die durch die UNIX-Pipeline verbunden sind. Das Bild/Videobild wird dabei von einem Programm zum nächsten weitergereicht. Das letzte Programm sollte das Bild entweder darstellen oder speichern. Die Pipeline-Architektur erhöht die Flexibilität der Software und erlaubt die parallele Ausführung der Pipelinekomponenten auf Multiprozessor-Computern.

pfstools
For High Dynamic Range Images and Video

pfstools ist eine Zusammenstellung von Basisprogrammen, die einfach erweitert und in andere Softwarepakete integriert werden kann. Während der Prototyp-Entwicklung unseres Kompressions-Algorithmus wurde *pfstools* zum Beispiel für das Lesen, Schreiben und Konvertieren von Bildern und Videobildern verwendet. HDR-Bilder können auf Standard-Bildschirmen mit einem *tone mapping*-Algorithmus dargestellt werden. Das auf *pfstools* aufgebaute Programmpaket *pfstmo* enthält Implementierungen von *tone mapping*-Algorithmen. Mit dem auf *pfstools* aufgebauten Paket *pfscalibration* können Kameras sowie Bilder in physikalischen oder kolorimetrischen Einheiten kalibriert werden. In der Bibliothek *HDR-VDP* ist ein Algorithmus implementiert, der das menschliche visuelle System simuliert. Dieser Algorithmus verwendet *pfstools*, um verschiedene Bildformate zu verarbeiten. :::

KONTAKT

Rafal Mantiuk
ABT. 4 Computergraphik
Telefon +49 681 9325-427
Email mantiuk@mpi-inf.mpg.de

Grzegorz Krawczyk
ABT. 4 Computergraphik
Telefon +49 681 9325-427
Email krawczyk@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/resources/pfstools/

Die TopX-Suchmaschine

Die Suchmaschine TopX recherchiert effizient und effektiv Daten vielfältigen Typs und Ursprungs: von strukturierten Daten aus Datenbanken wie Produktkatalogen über Musik- und Filmdatenbanken bis hin zu unstrukturierten Texten und Webseiten. Ein besonderer Schwerpunkt liegt auf semistrukturierten Daten, die im weit verbreiteten Austauschformat XML vorliegen. Dieses Format kombiniert stark strukturierte, semantisch annotierte Inhalte wie Autoreninformationen mit weitgehend unstrukturierten textuellen Informationen.

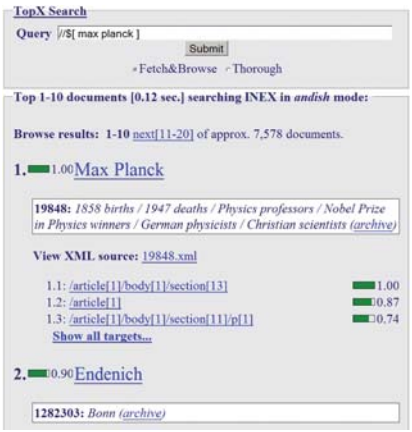
TopX vereint aktuelle Forschungsergebnisse aus den Bereichen Datenbanksysteme und Information Retrieval in einem frei verfügbaren Prototypsystem. So errechnen neuartige Bewertungsfunktionen die erwartete Qualität eines Dokumentfragmentes für eine gegebene Anfrage. Um mit der inhärenten sprachlichen Vielfalt und Heterogenität von verschiedenen Datenquellen – im Extremfall aus dem ganzen Web – umgehen zu können, kann TopX Anfragen optional um semantisch verwandte Begriffe erweitern. Diese Begriffe werden aus Ontologien und Thesauri wie zum Beispiel Wordnet oder Yago (siehe auch *„Automatische Erstellung von Ontologien“*, Seite 81) abgeleitet. Effiziente Auswertealgorithmen bestimmen auch für sehr große Datensammlungen in kurzer Zeit die besten „Treffer“, also die Ergebnisse mit der höchsten Bewertung. Probabilistische Varianten dieser Methoden beschleunigen die Ergebnisberechnung und nehmen dabei eine kleine, quantitativ kontrol-lierbare Verschlechterung der Ergebnisgüte in Kauf.

Der Benutzer interagiert mit TopX über ein Webseiteninterface, wie man es von existierenden Websuchmaschinen kennt. Es bietet aber zusätzliche Optionen zur automatischen Expansion von Anfragen und zur Darstellung von Teilergebnissen innerhalb von Dokumenten. Neben reinen Schlüsselwortanfragen können auch komplexere Anfragen formuliert werden, die Inhaltsbedingungen mit Anforderungen an die Struktur von Ergebnissen kombinieren. Bei der Suche in einer Literaturdatenbank kann ein Benutzer mit einer XPath-artigen, strukturierten Anfrage wie

```
//publication[//author Gerhard Weikum]
[//year~2000]//*[XML retrieval]
```

so nicht nur den Inhalt einer Veröffentlichung spezifizieren, sondern auch den Autor oder das Erscheinungsjahr eines Dokuments einschränken.

Dabei sind auch unscharfe Bedingungen möglich, wie beim Beispiel für das Erscheinungsjahr. Exakte Treffer erzielen hier eine höhere Bewertung als Treffer, die der Anfrage nur ähnlich sind. Das Anfrageergebnis sind in der Regel keine vollständigen Dokumente, sondern Dokumentfragmente wie Abschnitte oder Sätze, die besonders relevant für die Anfrage sind. TopX wertet explizite und implizite Rückmeldungen des Benutzers über einzelne Ergebnisse aus. Auf diesem Weg lassen sich die Anfragen verfeinern und vor allem automatisch Anforderungen an die Struktur von Ergebnissen generieren. Ein zusätzliches automatisiertes Interface ermöglicht die Einbindung von TopX in Web 2.0-Anwendungen.



Das Interface der TopX-Suchmaschine

Die prototypische Implementierung von TopX auf der Basis von Java ist als Open-Source-Software frei verfügbar. Sie wird als Referenz-Suchmaschine für eine Sammlung von gut 600.000 semistrukturierten Wikipedia-Dokumenten innerhalb des INEX-Projektes eingesetzt. Dabei handelt es sich um eine Initiative von etwa 60 internationalen Forschergruppen auf dem Gebiet der Suche nach semistrukturierten Daten. :::

KONTAKT

Ralf Schenkel
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-504
Email schenkel@mpi-inf.mpg.de
Internet http://topx.sourceforge.net

Minerva

Minerva ist eine Prototypimplementierung einer verteilten Peer-to-Peer-Websuchmaschine. Sie beruht auf einer unbeschränkten Anzahl einzelner, autonomer Suchmaschinen, die in kollektiver Weise zusammenarbeiten. Diese autonomen Suchmaschinen werden als Peers bezeichnet. Sie stellen jeweils eine eigene Suchfunktionalität über Dokumente (Webseiten) zur Verfügung, die lokal auf einem Computer gespeichert sind.

Der Benutzer eines Peers kann frei entscheiden, welche Dokumente für andere Peers sichtbar sein sollen. Kompakte Statistiken über alle Peer-Dokumente werden in einem gemeinsamen Datenverzeichnis zusammengefasst, das verteilt auf alle Peers des Systems gespeichert wird. Fällt ein einzelner Peer aus, gibt es so keine Datenverluste, und Zugriffe auf das Datenverzeichnis werden gleichmäßig auf alle Peers verteilt.

Sind die Ergebnisse einer lokalen Suche nicht zufriedenstellend, kann jeder Peer anhand der Statistiken im Datenverzeichnis weitere Peers als viel versprechendere Datenquellen identifizieren. Auf diese Weise lässt sich die Resultatsgüte erhöhen. Die Suchanfrage wird anschließend an die ausgewählten Peers weitergeleitet und von deren lokaler Suchmaschine bearbeitet. Anschließend werden die Ergebnisse auf direktem Weg zur anfragenden Suchmaschine zurückgeleitet. Senden mehrere Peers lokale Ergebnisse zurück, werden diese entsprechend ihrer Güte zu einer gemeinsamem Ergebnisliste zusammengefügt und für den Benutzer aufbereitet. Eine Herausforderung ist dabei, innerhalb eines a priori unbegrenzt großen P2P-Verbundes geeignete Peers schnell und effizient aufzufinden. Minerva unterhält zu diesem Zweck im gemeinsamen Datenverzeichnis Listen mit Peer-Adressen für alle aussagekräftigen und charakteristischen Dokumentenbegriffe aus dem gesamten Netzwerk. Auf diese Weise lassen sich schnell all jene Peers für einen Suchbegriff bestimmen, die überhaupt Dokumente für diesen Begriff besitzen. Bezieht man weitere statistische Daten mit ein, können so die aussichtsreichsten Peers in diesen Listen für jeden Suchbegriff effizient bestimmt und die Suchanfrage dorthin weitergeleitet werden.

Minerva fördert dank der Möglichkeit, eigene oder gefundene Webseiten zu annotieren *(Social Tagging)* sowie eine Suche auf Annotationen zu erweitern, die Kooperation zwischen allen Peers und ihren Benutzern. Das Einbeziehen von Annotationen in die Websuche bedeutet, dass nicht nur der Dokumenteninhalt zur Resultatsgüte beiträgt, sondern auch die Benutzermeinung – beispielsweise über die Qualität der Dokumente. Suchanfrage-Ergebnisse lassen sich auf diese Weise unmittelbar durch die intellektuelle Einflussnahme der Benutzer verbessern.

Um die lokalen Datenbestände eines jeden Peers zu erstellen, wurden die P2P-Suchmaschine Minerva und der fokussierte Web-Crawler Bingo! unter einer gemeinsamen Benutzeroberfläche zusammengeführt. Bingo! imitiert das Websurf-Verhalten eines menschlichen Benutzers und durchstreift *(„crawlt")* das Internet nach Webseiten, die für den Benutzer relevant und interessant sind. Bingo! erlernt die Benutzerinteressen im voraus anhand thematisch gruppierter Webseiten, die der Benutzer über eine Lesezeichendatei seines bevorzugten Webbrowsers festlegt. Bingo! ist im Anschluss in der Lage, aufgespürte Dokumente im Internet zu klassifizieren und anhand der vorgegebenen Gruppierungen thematisch zu kategorisieren. Der Crawl-Vorgang findet im Hintergrund und stets fokussiert auf die Benutzerinteressen statt, so dass der lokale Datenbestand eines Peers fortlaufend um relevante Daten erweitert werden kann.

Minerva ermöglicht es in Zusammenarbeit mit Bingo!, viele spezialisierte Datenbestände zu erstellen und diese im Internet verteilt zu hinterlegen. Jeder Benutzer kann auf diese Weise bequem auf eine riesige Datenmenge zugreifen. Minerva bietet Benutzern eine einfache Suchumgebung, die es ihnen erlaubt, vom gemeinschaftlichem Nutzen und den Vorteilen der Informationssuche in Peer-to-Peer-Systemen (siehe auch *„Informationssuche in Peer-to-Peer-Systemen"*, Seite 42) zu profitieren. :::

KONTAKT

Tom Crecelius
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-506
Email tcrecel@mpi-inf.mpg.de
Internet http://www.minerva-project.org

STATISTISCHES LERNEN

Die Fähigkeit des Menschen, auf der Grundlage von Beobachtungen Schlussfolgerungen über die zugrundeliegende Realität zu ziehen, berührt den Kern unseres Intelligenzbegriffs. Auf dem Forschungsgebiet statistisches Lernen gehen Wissenschaftler der Frage nach, wie sich Softwaresysteme konstruieren lassen, die Daten analysieren und aus ihnen Modelle der zugrundeliegenden Systeme gewinnen. Statistische Lernalgorithmen können betrügerische Phishing-Emails erkennen, das im Internet vorhandene Wissen nutzen um bei der Suche nach Informationen zu helfen, biologische Prozesse in Tumorzellen verstehen und menschliche Gesichtsausdrücke modellieren.

Computer sind uns im Bezug auf die Fähigkeit des Lernens aus Beobachtungen weit unterlegen. Uns Menschen fällt es leichter, ein Modell zu finden, dass das Verhalten eines beobachteten Systems beschreibt. Mit Hilfe eines solchen Modells können wir Schlussfolgerungen ziehen und das weitere Verhalten des Systems vorhersagen.

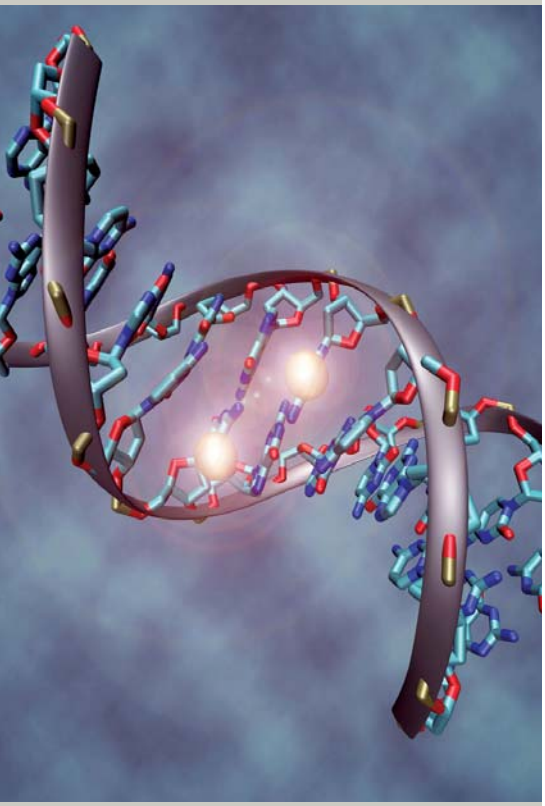
Das Forschungsgebiet statistisches Lernen beschäftigt sich mit der Konstruktion technischer Systeme, die Modelle aus Daten gewinnen. Die Anwendungsgebiete des statistischen Lernens sind vielfältig. In vielen Gebieten beschäftigen sich Wissenschaftler mit komplexen Systemen, deren Zusammenhänge noch nicht verstanden sind, über die aber sehr viel Datenmaterial vorliegt. Lernalgorithmen können helfen, aus diesen Daten Wissen zu gewinnen.

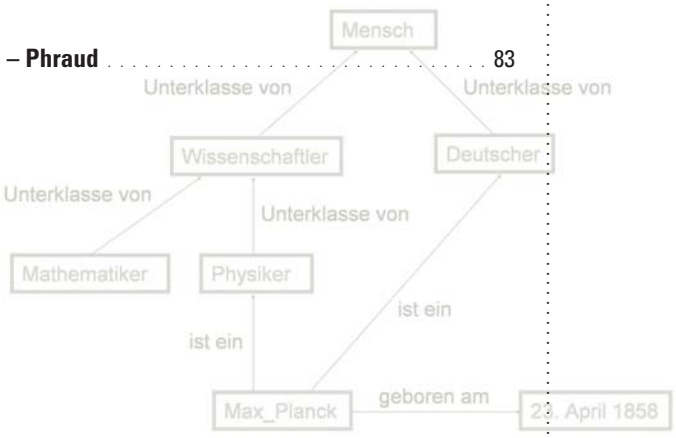
Am Max-Planck-Institut für Informatik werden Verfahren des statistischen Lernens entwickelt und in unterschiedlichen Kontexten angewendet. Die Forschungsgruppe Maschinelles Lernen geht grundlegenden Fragestellungen nach und untersucht die Anwendung von Lernalgorithmen im Bereich der Sicherheit gegen Spam, Phishing und Viren. Basierend auf einer großen Datenbasis lernen Softwaresysteme, anhand welcher Merkmale betrügerische Nachrichten und Webseiten sich von glaubwürdigen Nachrichten unterscheiden. Die Forschungsgruppe Automatisierung der Logik und die Abteilung Datenbanken und Informationssysteme untersuchen, wie Ontologien – Begriffshierarchien – aus Texten im Internet gewonnen werden können, und wie sich Schlussfolgerungen aus ihnen ziehen lassen.

In der Abteilung Bioinformatik und angewandte Algorithmik nutzen Forscher Lernalgorithmen, um den Zusammenhang zwischen genetischen Eigenschaften einzelner Patienten und der Wirksamkeit bestimmter Medikamente zu verstehen. Die Ergebnisse dieser Arbeiten könnten dazu beitragen, dass Patienten zukünftig individuell auf ihre genetische Eigenschaften abgestimmte Medikamente erhalten. Auch für ein besseres Verständnis der Entwicklung von Tumorzellen werden Modelle algorithmisch aus Daten gewonnen. In der Epigenetik untersuchen Wissenschaftler die Funktion von Methylanlagerungen am Erbgut. Bei einigen Krebserkrankungen spielt diese Methylierung eine wichtige Rolle. Da sich derartige Schäden grundsätzlich sogar umkehren lassen, besteht die Hoffnung auf neue Krebsmedikamente – bessere Modelle der Funktion der Methylierung vorausgesetzt.

Die Abteilung Computergraphik verwendet Lernalgorithmen um Eigenschaften von Bilddaten mit Gesichtern zu analysieren. Eines der Ergebnisse dieser Arbeiten ist ein Modell, das die Erzeugung von Gesichtern mit genau definierten Ausdrucksparametern zu visualisieren. Das System kann beispielsweise zur Erzeugung von Phantombildern und für Trickfilmgrafiken eingesetzt werden. :::

BEITRÄGE





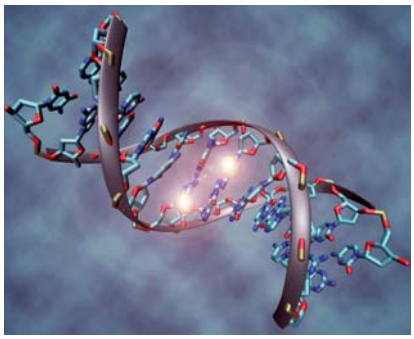
Bioinformatische Epigenetik: Bioinformatik für neue Wege in der Krebsforschung

Nach dem klassischen Verständnis von Krebs können Erbgut-Veränderungen zur Bildung von Tumoren führen. Im Zellerbgut werden einzelne DNA-Bausteine bis hin zu ganzen Erbgutabschnitten auf den Chromosomen ausgetauscht, gelöscht oder vervielfältigt. Da diese Schäden unumkehrbar sind, zielen chirurgische Eingriffe und Chemotherapie darauf ab, bei einem Patienten alle Krebszellen zu entfernen oder zu zerstören. Problematisch ist, dass genetische Veränderungen oft erst in einem späten Stadium der Krankheit festgestellt werden können, sodass eine langfristig erfolgreiche Behandlung oft nicht möglich ist.

Neue Forschungsergebnisse zeigen, dass Krebszellen in vielen Fällen bereits epigenetisch geschädigt sind, bevor sich in der DNA selbst Fehler anhäufen. Als epigenetische Veränderungen des Erbguts bezeichnet man vererbbare Modifikationen, die nicht das Genom selbst, also die Reihenfolge der DNA-Bausteine, verändern. Ein Beispiel ist die Markierung einzelner DNA-Bausteine durch eine zusätzliche Methylgruppe [siehe Abbildung]. Solche Veränderungen erfüllen in gesunden Zellen lebenswichtige Aufgaben. Zum Beispiel steuert die Methylierung das Andocken von anderen Molekülen an die DNA. So schützt sie die Zelle vor fremder DNA, hilft dabei, Fehler bei der DNA-Neubildung zu korrigieren und die Aktivität von Genen zu steuern. In vielen Krebszellen ist die DNA-Methylierung gestört, sodass DNA- Bereiche methyliert werden, die normalerweise nicht davon betroffen sein sollten. Dadurch können wichtige krebsunterdrückende Gene nicht mehr abgelesen werden. Die Folge: Diese Zellen vermehren sich übermäßig.

Die Tatsache, dass Krebszellen oft epigenetische Schäden aufweisen, bevor sich ein sichtbarer Tumor entwickelt, kann für eine verbesserte Frühdiagnose verwendet werden. Darüber hinaus ergibt sich ein weiterer vielversprechender Behandlungsansatz: Epigenetische Modifikationen von Krebszellen sind prinzipiell umkehrbar. Deshalb sollte es möglich sein, Tumore mit neuen Medikamenten in einen harmlosen Zustand zurückzuverwandeln, anstatt sie abzutöten oder zu entfernen. Mehrere Laboratorien und Pharmaunternehmen haben bereits begonnen, epigenetische Krebsmedikamente zu entwickeln. Diese Medikamente verändern die DNA-Methylierung von Krebszellen. Allerdings machen sie dabei nicht nur die epigenetischen Veränderungen in den Tumorzellen rückgängig. Sie beeinflussen auch natürliche DNA-Methylierungen, die für die normale Zellentwicklung notwendig sind. Deshalb haben epigenetische Medikamente bisher schwere Nebenwir-kungen und können – wie auch die klassische Chemotherapie – zu Schäden bei Nachkommen der Patienten führen. In einer Kooperation zwischen dem Max-Planck-Institut für Informatik (Abteilung 3: Bioinformatik) und der Universität des Saarlandes (Lehrstuhl für Genetik, Prof. Jörn Walter) versuchen Saarbrücker Wissenschaftler, dieses Problem mithilfe von Bioinformatik zu lösen. Die Grundidee ist, dass ein intelligentes epigenetisches Krebsmedikament nur die fehlerhaften DNA-Methylierungen in einer Krebszelle rückgängig machen soll. Die Bioinformatik soll helfen zu verstehen, inwieweit sich fehlerhafte DNA-Methylierungen von biologisch notwendigen unterscheiden.

Im ersten Schritt wurde eine Software entwickelt, mit der experimentell ermittelte DNA-Methylierungsdaten auf ihre Richtigkeit überprüft werden können. Dieses Programm unterstützt die Arbeit im Labor, bei der Tumorproben auf epigenetische Veränderungen hin untersucht werden. Ferner hilft es dabei, einen einheitlichen Qualitätsstandard für die Analyse zu etablieren.



Ein methyliertes DNA-Molekül. Fehlerhafte Methylierungen im menschlichen Erbgut können Krebs verursachen.

Darauf aufbauend wurden DNA-Methylierungsmuster im Blut von gesunden Patienten mit verschiedenen Informationen über das menschliche Erbgut verglichen. Mit Hilfe der Data-Mining-Methoden wurden drei Eigenschafts-Gruppen menschlicher DNA identifiziert, die für die normale DNA-Methylierung entscheidend sind: 1. die DNA-Sequenz, 2. sich wiederholende DNA-Abschnitte und 3. die dreidimensionale Struktur der DNA. Mit dieser Erkenntnis lässt sich mit 90-prozentiger Genauigkeit die Verteilung der Methylierung im Erbgut gesunder Zellen vorhersagen. Dies ist insofern wichtig, als trotz Entschlüsselung des menschlichen Genoms bisher keine genomweiten DNA-Methylierungsdaten von ausreichender Genauigkeit zur Verfügung stehen.

Aus dem Vergleich der Methylierungsmuster von gesundem Gewebe und Krebszellen sollen künftig Konzepte für verträglichere Medikamente gegen Krebs entwickelt werden. Darüber hinaus wird untersucht, wie sich ein epigenetisches Krebsmedikament, das bereits erprobt wird, auf die DNA-Methylierung im gesamten Erbgut auswirkt. Die Ergebnisse könnten einen konkreten Startpunkt für die Optimierung einer epigenetisch wirkenden Chemotherapie bieten. :::

KONTAKT

Christoph Bock
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-322
Email cbock@mpi-inf.mpg.de
Internet http://computational-epigenetics.mpi-inf.mpg.de

Statistische Analyse bei medizinischer Diagnose und Prognose

Personalisierte Medizin aus dem Genom

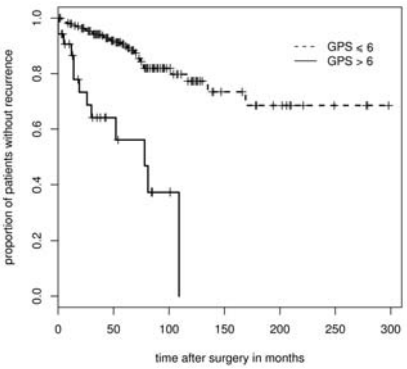
Menschen erkranken an derselben Krankheit - die Verläufe aber unterscheiden sich bisweilen sehr. In der modernen Massenmedizin diagnostizieren und prognostizieren Ärzte Krankheiten und deren Verläufe anhand von klinischen und histopathologische Messungen am Patienten. Es ist jedoch bekannt, dass bei verschiedenen Patienten mit gleichen klinischen Merkmalen die Krankheit unterschiedlich verlaufen kann, und die Betroffenen folglich auf die Therapie unterschiedlich ansprechen. Durchschnittlich reagieren etwa zwei Drittel in gewünschter Weise auf die heutigen Massenmedikamente, bei dem übrigen Drittel ist entweder die Dosis zu gering oder es treten unerwünschte Nebenwirkungen auf.

Für diese unterschiedlichen Verläufe und Reaktionen gibt es zwei Gründe. Zum einen variieren die genetischen Merkmale der Patienten. Der andere Grund ist die Variabilität im Krankheitsbild, die selbst dann vorliegt, wenn sich die Messergebnisse klinisch und genetisch ähneln. Um beide Probleme angemessen zu berücksichtigen, setzt die Forschungsgruppe statistische Methoden ein. Das erste Ziel ist, Patientengruppen anhand ihres Genoms besser zu charakterisieren und damit Diagnose und Prognose individuell zu verbessern. Das zweite Ziel ist, die Varianz im Krankheitsbild bei gleichen genetischen und klinischen Messungen zu quantifizieren.

Eine weitere Herausforderung für die moderne Genomforschung ist die explosionsartig zunehmende Zahl der zur Verfügung stehenden genetischen Messergebnisse. Mit der Microarray-Technologie etwa wird gleichzeitig die Aktivität von zirka 30.000 Genen in einem bestimmten Gewebe gemessen. Auch hier sind statistische Verfahren nötig, um die Vielfalt der in den Daten enthaltenen Information über die beteiligten Gene auf das biologisch Relevante zu reduzieren.

Klinische Prognose aus genetischen Tumordaten

Um Krebspatienten optimal zu therapieren, ist die Abschätzung der Zeit zwischen Behandlung und Rückfall beziehungsweise Tod wichtig. Nur so lässt sich die geeignete Therapie wählen, zum Beispiel in Bezug auf deren Aggressivität. Mit einer neuen Modellklasse von Mischungen von Bäumen schätzen wir tumorspezifische Krankheitsverläufe. Der Fortschritt der Krankheit ist durch die Reihenfolge permanenter genetischer Veränderungen in den Tumorzellen gekennzeichnet. Von diesen Modellen leiten wir den genetischen Progressions-Score (GPS) ab, der den genetischen Status eines Tumors statistisch quantifiziert. Wir konnten zeigen, dass der GPS ein medizinisch relevanter prognostischer Faktor ist. Bei mehreren Krebsarten, insbesondere bei verschiedenen Hirntumoren und beim Prostatakrebs, können Patientengruppen über den GPS mit verschiedenen klinischen Verläufen differenziert werden; und das, obwohl sich die Tumoren gemäß der klassischen Einteilung nach pathohistologischen Ergebnissen nicht unterschieden hätten [siehe Abbildung].



Kaplan-Meier-Kurven für Zeiten bis zum Rückfall nach operativer Entfernung eines Meningioms. Auf der x-Achse ist die Zeit bis zur Wiederkehr des Tumors aufgetragen, auf der y-Achse der relative Anteil der Patienten ohne Rückfall bis zu diesem Zeitpunkt. Die untere Kurve beschreibt Patienten mit genetisch fortgeschrittenen Tumoren und zeigt eine deutlich schlechtere klinische Prognose.

Wirkung einer Chemotherapie

Wir haben die Relevanz von genetischen Veränderungen in Tumorzellen von Glioblastompatienten (Hirntumor) für eine Chemotherapie mit dem Medikament Temodal untersucht. Mit multivariaten Modellen der medizinischen Statistik wurde gezeigt, dass für Patienten ohne begleitende Therapie der Verlust von Sequenzen auf den Chromosomen 9 und 10 in den Tumorzellen die Überlebenszeit verkürzt. Es zeigt sich jedoch, dass gerade Patienten mit solchen Verlusten von einer Behandlung mit Temodal profitieren.

Identifizierung wichtiger biologischer Prozesse in verschiedenen Krebssorten

Mit Microarray-Experimenten wird die Expression (Aktivität) von Tausenden Genen in einem Gewebe gleichzeitig gemessen. Für viele Krebsarten gibt es typische Muster von Genen, die in dem kranken Gewebe besonders stark exprimiert sind. Oft kann man anhand dieser Muster Krebsgewebe von gesundem Gewebe sowie biologisch definierte Untergruppen einer Krebsart zuverlässig unterscheiden. Im Hinblick auf eine geeignete Therapie will man nun die biologischen Prozesse verstehen, die für die Unterschiede verantwortlich sind. Für viele Gene ist bekannt, an welchen biologischen Prozessen sie beteiligt sind. Wir haben Algorithmen entwickelt, die für Patientengruppen mit verschiedenen Krankheiten die wichtigsten unterschiedlichen biologischen Prozesse anhand von Microarray-Experimenten bestimmen. Eine Schwierigkeit ist dabei, dass viele Gene an mehreren Prozessen gleichzeitig beteiligt sind. Wir haben gezeigt, dass die neuen Algorithmen nachweisbar mehr biologisch relevante Prozesse identifizieren als klassische etablierte Verfahren. Die Methoden wurden erfolgreich auf Prostatakrebsdaten angewendet. :::

KONTAKT

Jörg Rahnenführer
ABT. 3 Bioinformatik und Angewandte Algorithmik
Telefon +49 681 9325-300
Email rahnenfj@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/departments/d3/projects.html

Lernbasierte Modellierung dreidimensionaler Objekte

Lernbasierte Objektmodellierung

Detailreicher und realistischer – so lässt sich der stetig wachsende Anspruch an synthetische Bilder in computeranimierten Spielfilmen und anderen Medien formulieren. Der damit einhergehende Zuwachs an Komplexität stellt die Produktionsstudios vor erhebliche personelle Herausforderungen, da die Künstler heute noch überwiegend manuell modellieren und animieren. Alternativ dazu lassen sich in manchen Anwendungen reale Objekte dreidimensional einscannen und in die virtuelle Szene einbauen. Mit einem geringen Arbeitsaufwand lassen sich auf diese Weise qualitativ hochwertige Ergebnisse erzielen. Die Methode bildet allerdings ausschließlich reale Gegebenheiten ab.

Ziel des Forschungsprojektes *Lernbasierte Objektmodellierung* ist es, aus Beispieldaten automatisch die wesentlichen Eigenschaften einer Objektklasse zu extrahieren und anschließend gezielt zu verändern. Mitglieder der Arbeitsgruppe von Volker Blanz, der während des Berichtszeitraums einen Ruf an die Universität Siegen angenommen hat, berechneten beispielsweise aus dreidimensionalen Gesichtsdaten die Unterschiede zwischen männlichen und weiblichen Gesichtern oder zwischen schlanken und übergewichtigen Personen. Diese Daten stehen dem Benutzer nun in Form von Schiebereglern als übergeordnete, semantisch bedeutungsvolle Steuergrößen zur Verfügung. Eine Beispielanwendung ist die Generierung von Phantombildern, die in der Arbeitsgruppe realisiert und gemeinsam mit dem saarländischen Landeskriminalamt erfolgreich getestet wurde.

Eine weitere gesichtsspezifische Eigenschaft, die in der Arbeitsgruppe untersucht wurde, ist das Alter einer Person. Dazu nahmen Wissenschaftler vom Max-Planck-Institut Gesichts-Scans von Jugendlichen dreidimensional auf. Ergänzend wurden 3D-Daten von Babygesichtern aus einer medizinischen Beobachtungsreihe des Universitätsklinikums Tübingen genutzt. Die Daten wurden statistisch analysiert und in Altersklassen unterteilt. Aus den Ergebnissen ließen sich altersbezogene Mittelwertsgesichter berechnen [Abbildung 1].

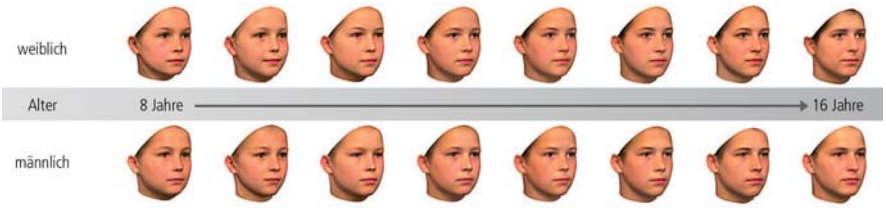


Abbildung 1: Altersbezogene Mittelwertsgesichter von Jugendlichen

Mit Hilfe dieser Mittelwerte wurden die Merkmale einzelner Gesichter im Vergleich zu Gleichaltrigen extrahiert und Wachstumsprognosen erstellt. Dazu wurden die persönlichen Gesichtsmerkmale entlang einer Wachstumsfunktion verschoben, die zuvor aus den Mittelwerten berechnet wurde.

Ein potenzielles Anwendungsgebiet ist die Phantombilderstellung. Werden Kinder beispielsweise über einen längeren Zeitraum vermisst, ist es notwendig, Phantombilder auf einem möglichst aktuellen Stand zu halten. Veränderungen durch das Größenwachstum müssen also widergespiegelt werden. Bislang zeichnete speziell ausgebildetes Personal die Phantombilder mit zeitaufwendiger Retuschearbeit von Hand. Eine automatisierte Methode erspart dem Anwender somit langwierige und detailreiche Handarbeit. Zudem kann ungeschultes Personal die Technik nutzen.

Schließlich wurden mit einem 3D-Scanner während des Sprechens Gesichtsbewegungen aufgenommen und statistisch ausgewertet [Abbildung 2]. Mit Hilfe dieser Untersuchungen sollen die Mundbewegungen virtueller Sprecher realistischer wirken. Gleichzeitig soll die mühsame, schrittweise Animationsmethode der Trickfilmer durch automatische Verfahren ersetzt werden. Dank dieser Arbeitserleichterung können sich die Gestalter von Computeranimationen künftig auf die künstlerische Ausgestaltung der Mimik beim Sprechen konzentrieren und damit ausdrucksvollere, realistischere sowie nuancenreichere Inhalte schaffen als bisher. :::

Abbildung 2: Dreidimensionale Scans von verschiedenen Mundstellungen, aufgenommen mit einem dynamischen 3D Scanner

KONTAKT

Kristina Scherbaum
ABT . 4 Computergraphik
Telefon +49 681 9325-424
Email scherbaum@mpi-inf.mpg.de

Volker Blanz
ABT . 4 Computergraphik
Telefon +49 681 9325-400
Email blanz@mpi-inf.mpg.de

Automatische Erstellung von Ontologien

Die Suche im Internet mit Suchmaschinen

Welche Physiker wurden im selben Jahr wie Max Planck geboren? Wer diese Frage im Internet recherchiert, gerät schnell an die Grenzen der Technik: Für Suchmaschinen wie Google ist dieser Wissenswunsch zu komplex. Alle Anfragen nach *„Physiker, geboren, Jahr, Max Planck“* geben lediglich Max Planck selbst zurück. Google kann nur diejenigen Fragen beantworten, zu denen es im Internet bereits eine vorgefertigte Antwort auf einer Webseite gibt. Um komplexere Fragen beantworten zu können, müsste dem Computer das Wissen dieser Welt in einer gigantischen Wissensstruktur zur Verfügung stehen.

Wissensrepräsentation in Ontologien

Eine solche strukturierte Wissenssammlung heißt **Ontologie**. In einer Ontologie sind Personen, Dinge und Daten durch Beziehungen miteinander verbunden. Beispielsweise steht die Person *„Max Planck“* mit dem Datum *„23. April 1858“* in der Beziehung *„geboren am“*, denn Max Planck wurde am 23. April 1858 geboren. Ebenso steht die Person „Max Planck“ mit der Klasse „Physiker“ in der Beziehung *„ist ein“*, denn Max Planck gehört zur Klasse der Physiker. Letztendlich ist eine Ontologie also eine große, netzartige Wissensstruktur.

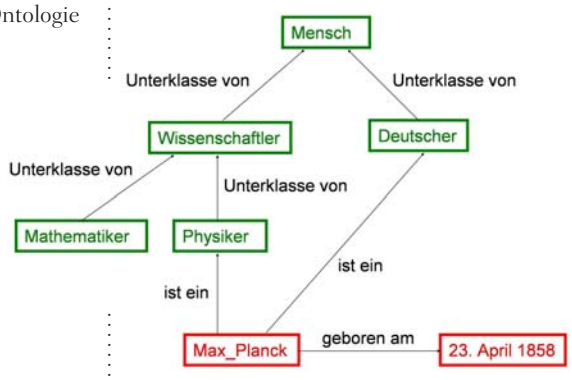
Automatische Konstruktion und Pflege von Ontologien

Um eine solche Ontologie automatisch mit Wissen zu füllen, nutzen wir die Online-Enzyklopädie Wikipedia. Wikipedia enthält unzählige Artikel über Persönlichkeiten, Produkte, Begriffe und Organisationen. Jeder dieser Artikel ist bestimmten Kategorien zugeordnet. So befindet sich beispielsweise der Artikel über Max Planck in den Kategorien *„Deutscher“*, *„Physiker“* und *„Geboren 1858“*. Diese Information nutzen wir, um die Klassenzugehörigkeit und das Geburtsdatum von Max Planck in der Ontologie zu vermerken. Um auch Informationen aus anderen Internetseiten zu sammeln, verwenden wir einen Ansatz namens **Pattern Matching**. Um beispielsweise neue Geburtsdaten in die Ontologie einzufügen, finden wir zunächst automatisch anhand bekannter Geburtsdaten heraus, nach welchem Muster Geburtsdaten häufig auf Webseiten genannt werden. Ein sehr gängiges Muster für Geburtsdaten ist z.B. *„X wurde am Y geboren“ („Max Planck wurde am 23. April 1858 geboren“)*. Durchsucht man nun das Internet nach weiteren Vorkommnissen dieses Musters, so werden andere Paare aus Personen und Geburtsdaten zu Tage gefördert. Diese lassen sich dann in die Ontologie eintragen.

Yago

Da wir beide Techniken kombiniert haben, ist es uns gelungen, eine sehr große Ontologie herzustellen: Yago *(Yet another Great Ontology)*. Yago kennt momentan fast eine Million Begriffe und hält passend dazu rund 6 Millionen Fakten bereit. Yago ist online verfügbar und kann über eine spezielle Abfragesprache Anfragen beantworten. So nennt Yago beispielsweise auf die eingangs gestellte Frage *„Welche Physiker wurden in selben Jahr geboren wie Max Planck?“* mehrere Dutzend andere Physikern. Diese Daten lassen sich nicht nur abfragen, sondern auch benutzen, um logische Schlussfolgerungen daraus zu ziehen (siehe auch *„Entscheidungsverfahren für Ontologien“*, Seite 82).

Diese Methoden zur ontologiegestützten Wissenssuche können auch in künftige Suchmaschinen integriert werden. Derartige Suchmaschinen wären ein Durchbruch für den Schritt von der fortgeschrittenen Informationsgesellschaft zu einer modernen Wissensgesellschaft, in der das Wissen der Menschheit nicht nur im Internet verfügbar ist, sondern auch effektiv genutzt werden kann. :::



Ausschnitt aus einer Ontologie

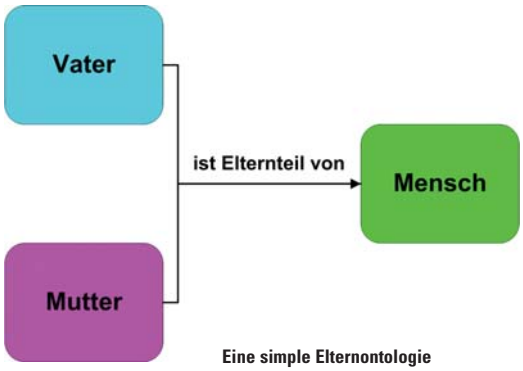
KONTAKT

Fabian M. Suchanek
ABT. 5 Datenbanken und Informationssysteme
Telefon +49 681 9325-509
Email suchanek@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/yago/

Entscheidungsverfahren für Ontologien

Viele Themen lassen sich heute hervorragend im Internet recherchieren. Oft genügt es, eine Phrase bei einer gängigen Suchmaschine wie GOOGLE einzugeben, um an die gewünschten Informationen zu gelangen. Es ist erstaunlich, wie erfolgreich Suchmaschinen mit Mehrdeutigkeiten der natürlichen Sprache umgehen können (siehe auch die Beiträge zum Thema *„Internet“*, ab Seite 56). Geht es allerdings um die Bedeutungen von Wörtern, die sich etwa durch Synonyme ausdrücken, sind den hier verwendeten Suchmethoden Grenzen gesetzt. Zum Beispiel sind für uns die Bezeichnungen *„Vater“* und *„männlicher Elternteil“* äquivalent. Die heute gängigen, wortorientierten Suchmaschinen werden für die beiden Bezeichnungen nicht unbedingt die gleichen Informationen liefern, möglicherweise sogar nicht einmal die gewünschten Ergebnisse.

Folgerichtig ist die Idee, die heutige Technologie um solches Wissen zu erweitern, um so genannte Ontologien. Dabei ergeben sich zwei Probleme: die Erstellung von Ontologien, die im Forschungsthema *„Automatische Erstellung von Ontologien“*, Seite 81 behandelt wird, sowie das Schlussfolgern in Ontologien, zum Beispiel, dass *„Vater“* und *„männlicher Elternteil“* dasselbe bedeuten.



Eine simple Elternontologie

Dazu muss die Theorie über Väter und Eltern beschrieben werden. Wir könnten beispielsweise sagen, dass jeder Mensch genau einen Vater und eine Mutter hat, die wieder Menschen sind; dass Väter Männer sind und Mütter Frauen; dass die Elternteile genau aus Vater und Mutter bestehen und dass das Geschlecht eines Menschen entweder männlich oder weiblich ist. Diese Aussagen reichen aus, um zu schlussfolgern, dass die Bedeutungen von *„Vater“* und *„männlicher Elternteil“* gleich sind.

Die Herausforderung liegt nun darin, dies nicht nur für die *„Elterntheorie“* zu tun, sondern für einen möglichst großen Teil unseres Wissens über Kategorien und Relationen von Objekten. Im Allgemeinen sind Aussagen über dieses Wissensgebiet nicht mehr entscheidbar, aber Entscheidbarkeit ist unerlässlich, um die eingangs erwähnten Anfragen effektiv auszuwerten. Deshalb ist die Erforschung entscheidbarer Fragmente ein wichtiges Thema.

Eine Familie solcher Fragmente, die sich besonders gut eignet, um konzeptuelles Wissen darzustellen, sind die so genannten Beschreibungslogiken. Im einfachsten Fall erhält man sie aus der Prädikatenlogik, indem man auf Funktionen und Gleichheit verzichtet. Unäre Prädikate firmieren als Konzepte; im Beispiel etwa Mann und Frau. Binäre Prädikate (Rollen) und Quantoren treten nur gemeinsam auf, wobei der Gültigkeitsbereich des Quantors durch ein Konzept beschränkt ist. So existiert für jede Mutter ein Mensch, zu dem sie in der Relation *ist-Elternteil-von* steht. Höherstellige Prädikate gibt es nicht. Die Fundierung von Ontologien in solchen Beschreibungslogiken gibt ihnen eine mathematisch präzise erklärte Semantik. Die Entscheidbarkeit der Logik(en) macht es möglich, Ontologien auf Redundanzen oder auf etwaige Widersprüche zu testen und Suchanfragen semantisch zu bearbeiten.

Naheliegend ist die Frage, wie weit die logischen Ausdrucksmittel erweitert werden können, ohne dass die Entscheidbarkeit verloren geht. Ihre Untersuchung hat in den vergangenen Jahren zu einer fast unüberschaubaren Vielzahl von Beschreibungslogiken geführt, für die jeweils eigene Entscheidungsverfahren entwickelt und implementiert wurden. Interessanterweise lassen sich viele der Entscheidbarkeitsaussagen auf Sonderfälle eines Semi-Entscheidungsverfahrens der Prädikatenlogik zurückführen: der Resolution. Mit dem Beweiser SPASS steht an dieser Stelle ein mächtiges Werkzeug zur Verfügung. Eine naheliegende Erweiterung ist die Negation von Rollen. Mit der Negation der Rolle *ist-Elternteil-von* kann dann zum Beispiel Kinderlosigkeit ausgedrückt werden. Für das Schlussfolgern in solchen Erweiterungen ist SPASS gegenwärtig das einzige Werkzeug. :::

KONTAKT

Thomas Hillenbrand
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-217
Email hillen@mpi-inf.mpg.de

Christoph Weidenbach
FG. 1 Automatisierung der Logik
Telefon +49 681 9325-900
Email weidenbach@mpi-inf.mpg.de

Phishing – Pharming – Phraud

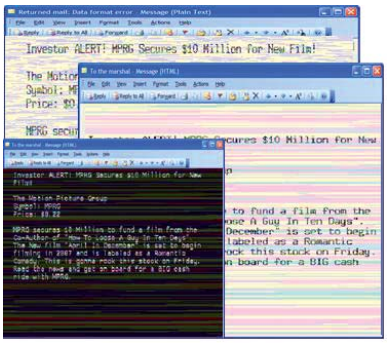
Täglich werden mehrere Milliarden Spam-Emails versendet. Weitgehend harmlose Werbung für Medikamente, Universitätsabschlüsse, Software-Raubkopien oder „Original"-Rolex-Uhren macht nur einen Teil davon aus. Viele der Spam-Emails haben einen betrügerischen Hintergrund. Ein Beispiel: die „Pump-and-Dump"-Kampagne. Kriminelle kaufen Aktien unbekannter Firmen und versenden Mails mit gefälschten Informationen über dieses Unternehmen mit dem Ziel, den Aktienkurs zu verändern (Pump) und die Aktien dann mit Gewinn zu verkaufen (Dump). Mithilfe von Computern und dem Internet ist es für den Spam-Versender überhaupt kein Problem, die entsprechenden Kontaktdaten zu verzerren. Untersuchungen zeigen, dass der Spammer mit solchen Kampagnen hohe Gewinne erzielt. Sieben Prozent des eingesetzten Kapitals sind durchaus realistisch.

Die so genannten 419-Scam-Emails – benannt nach dem Paragraphen über „Advance Fee Fraud" im nigerianischen Strafgesetzbuch – verlocken viele Empfänger dazu, immer höhere „Gebühren" vorzustrecken, um am Ende einen besseren Gewinn zu machen. Die Folgen können tödlich sein: In den vergangenen zehn Jahren wurden mindestens fünfzehn Personen erschossen, die ihre betrügerischen Geschäftspartner persönlich besucht hatten, als Geisel genommen wurden und nicht rechtzeitig freigekauft werden konnten.

Mit Phishing-Emails entlocken Betrüger ihren Opfern Kreditkartendaten und Zugangsdaten zu Banken, Bezahlsystemen und Online-Shops. Professionelle Phishing-Software findet individuelle Anknüpfungspunkte, die der Email Glaubwürdigkeit verleihen. Ebay-Phishing-Mails beziehen sich häufig auf Artikel, die das Phishing-Opfer tatsächlich zum Kauf angeboten hat; folgt der Anbieter dem Link zum Beantworten der vorgeblichen Anfrage, gibt er seine Zugangsdaten preis.

Mithilfe von Viren und manipulierten Domain-Name-Servern leiten Betrüger Zugriffe auf bestimmte Webadressen wie Bank-Webseiten oder PayPal auf einen eigenen Seiten-Nachbau um. Der Nutzer bleibt ahnungslos. Mit diesen „Pharming"-Attacken können massenhaft Zugangsdaten gesammelt werden.

Spam-Versender vermeiden es, identische Kopien derselben Nachricht zu verschicken, denn sobald eine Nachricht als Spam oder Virus bekannt ist, könnten alle später versendeten Kopien einfach gelöscht werden. Stattdessen werden die Nachrichten mit probabilistischen kontextfreien Grammatiken individuell generiert und von vielen, mit Viren befallenen Rechnern aus versendet. Auch Bilder-Spams werden mit Grammatiken einzeln erzeugt; selbst für den Source-Code von Viren werden probabilistische Grammatiken verwendet, weshalb diese schwer wiederzuerkennen sind.



Es ist nun eine spannende Forschungsaufgabe, Algorithmen zu entwikkeln, die solche Text- und Bild-Nachrichten in einem Nachrichtenstrom erkennen, die von einem Generator nach demselben Muster erzeugt worden sind.

Die Entwicklung von Technologien, die diese Betrugsarten verhindern, ist ein Wettlauf zwischen Spammern und Entwicklern von Gegenmaßnahmen. Häufig haben Betrüger schon wieder neue Techniken entwickelt, gegen die der bisherige Schutz nicht mehr reicht. Die Filter-Entwickler müssen dann zügig aufholen. Robuste Algorithmen zu entwickeln, die den nächsten Zug der Spam-Versender automatisch vorwegnehmen und sich nicht durch Weiterentwicklungen der Spammer aushebeln lassen, gehört somit zu den spannendsten Forschungsthemen im Brennpunkt von maschinellem Lernen und Spieltheorie. :::

KONTAKT

Tobias Scheffer

FG. 2 Maschinelles Lernen

Telefon +49 681 9325-502

Email scheffer@mpi-inf.mpg.de

VISUALISIERUNG

Bilder sind der schnellste Weg zum menschlichen Bewusstsein. So nimmt die Entwicklung von Algorithmen zur geeigneten Visualisierung digitaler Information auch eine besondere Rolle in der Informatik ein. Insbesondere die Unterhaltungselektronik stellt immer höhere Anforderungen an die Visualisierungsalgorithmen: Immer realistischer und schneller müssen künstliche und natürliche Welten dargestellt werden können – in Flugsimulatoren, chirurgischen Operationsplanungssystemen oder Computerspielen. In den Natur- und Ingenieurwissenschaften kommen mehr und mehr Visualisierungsverfahren zum Einsatz, um komplizierte Fragestellungen effizient lösen zu können.

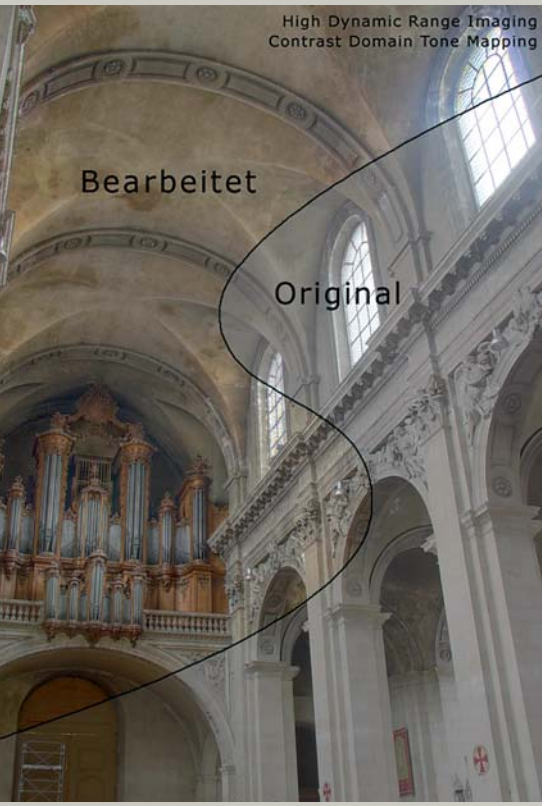
Um qualitativ hochwertige Bilder zu erzeugen, sind viele Faktoren zu berücksichtigen. Basis dafür ist die akkurate Szenenmodellierung. Am Max-Planck-Institut für Informatik werden daher automatisierte Methoden zur Digitalisierung realer Gegenstände erforscht, die neben der eigentlichen Szenengeometrie auch die Beleuchtung und die Reflexionseigenschaften exakt vermessen. Die Akquisitionsverfahren sind an die Größe der Objekte angepasst – kleinere Kunstgegenstände, Straßenzüge oder auch astronomische Nebel. Die digitalisierten Modelle erlauben es, den Gegenstand aus beliebigen Perspektiven unter beliebiger, virtueller Beleuchtung exakt wiederzugeben und damit nahtlos in eine virtuelle Welt einzubetten. Ein weiterer Forschungsschwerpunkt liegt in der Rekonstruktion dynamischer Szenen aus Videoströmen mit dem Ziel, die nächste Generation von 3D-Video und 3DTV-Anwendungen zu entwickeln. Auch hier ist die exakte Wiedergabe der sich verändernden Form, der Textur- und der Reflexionseigenschaften von entscheidende Bedeutung. Inzwischen ist es beispielsweise möglich, die Bewegung von Schauspielern mitsamt ihrer Kleidung aufzunehmen und aus neuen Blickwinkeln wiederzugeben.

Die Aufbereitung von Datensätzen anderer wissenschaftlicher Disziplinen ist ein weiterer Kernbereich der Visualisierung. Hauptaufgabe ist es hier, wesentliche Eigenschaften eines komplexen Datensatzes visuell hervorzuheben. Erforscht wurden Techniken, die die zeitliche Entwicklung von topologisch interessanten Punkten in zeitabhängigen Volumendatensätzen sichtbar machen, beispielsweise die Bahn eines Tornado-Auges. Andere Visualisierungsverfahren wurden entwickelt, um die Verbindung zwischen den verschiedenen physikalischen Größen wie etwa Druck, Temperatur und Geschwindigkeit effektiv analysieren zu können.

Um virtuelle Welten naturgetreu erscheinen zu lassen, wird am Max-Planck-Institut für Informatik an Methoden zur Simulation der Lichtausbreitung in Szenen, der so genannten globalen Beleuchtung, geforscht. Bei der globalen Beleuchtung simuliert man neben dem einfallenden Licht, das direkt von den Objekten reflektiert wird, vor allem das indirekte Licht, das mehrfach von verschiedenen Oberflächen hin- und hergeworfen wird,bevor es den Betrachter erreicht. Das indirekte Licht erreicht fast jeden Winkel und lässt auch Schattenbereiche nicht völlig dunkel erscheinen. Bisher galt die Simulation der indirekten Lichtausbreitung als sehr zeitintensiv. Am Max-Planck-Institut werden deshalb zwei Hauptrichtungen verfolgt, um effizientere Algorithmen zu entwickeln: zum einen die Berechnung der globalen Beleuchtung auf schnelleren Graphikkarten (GPU) sowie die beschleunigte Simulation für Bildsequenzen, bei der wir die zeitliche Kohärenz und die Beschränkungen der menschlichen Wahrnehmung (Perzeption) ausnutzen.

Die durch die Simulation oder durch Aufnahme entstandenen Bilder umfassen typischerweise einen Helligkeitsbereich und einen Farbumfang, die nahe an die reale Welt heranreichen. Dieser große Dynamikbereich (HDR) erfordert spezielle Methoden zur Bildverarbeitung und zur Darstellung auf heutigen Monitoren, deren Wiedergabemöglichkeiten sowohl bezüglich der Helligkeit als auch der Farben deutlich limitiert sind. Am Max-Planck-Institut für Informatik werden Werkzeuge entwickelt, die den Umgang mit HDR-Bildern und -Videos und deren möglichst naturgetreue Wiedergabe deutlich vereinfachen.

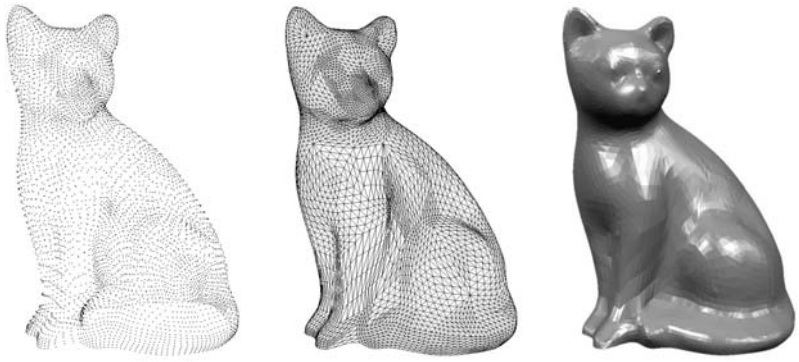
BEITRÄGE



Oberflächenrekonstruktion

Digitale Modelle

Digitale Modelle physikalischer Objekte aus unserer dreidimensionalen Welt stehen nach wie vor hoch im Kurs: Kultur-Denkmäler wie die Freiheitsstatue in New York sollen beispielsweise digital erhalten werden, um sie im Falle eines Terroranschlags rekonstruieren zu können. Auch Anwendungen in der Unterhaltungsindustrie sind unter wirtschaftlichen Gesichtspunkten von großer Bedeutung.



Punktwolke, rekonstruiertes Dreiecksnetz und digitales Modell

Um ein digitales Modell herzustellen, tastet in der Regel ein Laser die Oberfläche eines dreidimensionalen Objektes ab. Das Ergebnis des Abtastvorgangs ist eine Punktwolke der Oberfläche des Objekts. Oftmals erhält man aber nicht eine Punktwolke, die die ganze Oberfläche repräsentiert, sondern Punktwolken von verschiedenen, überlappenden Oberflächenteilen. Das Zusammenfügen dieser Teilpunktwolken zu einer Punktwolke des gesamten Objekts stellt durchaus ein Problem dar. Ist aber schließlich eine Punktwolke für die gesamte Oberfläche entstanden, lässt sich das gewünschte digitale Modell mit einem Algorithmus für Oberflächenrekonstruktion erstellen.

Rekonstruktionsalgorithmen

Heute existieren diverse Rekonstruktionsalgorithmen, die eine Punktwolke in ein digitales Oberflächenmodell überführen. Viele davon wurden am Max-Planck-Institut für Informatik entwickelt. Die Mathematik dieser Algorithmen reicht von algebraischer Topologie, jenem Zweig der Mathematik, der es erlaubt beispielsweise den Unterschied zwischen einem Ball und einem Rettungsring zu erfassen, bis hin zu partiellen Differentialgleichungen – dem Handwerkszeug der theoretischen Physiker. Nach wie vor werden verschiedene mathematische Techniken benutzt und entwickelt, um Flächenrekonstruktionsalgorithmen zu formulieren, denn das Rekonstruktionsproblem ist sowohl theoretisch als auch praktisch noch nicht vollkommen befriedigend gelöst. Auf abgetasteten Punktwolken funktionieren die meisten Algorithmen sehr gut, enthält die Punktwolke aber Ausreißer, ist sie verrauscht oder stammt sie von einer scharfkantigen Oberfläche, haben fast alle Algorithmen Probleme.

Die allgemein akzeptierte Abtasttheorie für Flächen ohne scharfe Kanten erlaubt es, theoretische Garantien für Rekonstruktionsalgorithmen zu geben: Ist die Punktwolke dicht genug im Sinne der Abtasttheorie, lässt sich für einige Algorithmen mathematisch beweisen, dass das Ergebnis des Algorithmus eine Fläche ist, die sehr viele Eigenschaften mit der abgetasteten Fläche teilt. Mittlerweile existieren auch erste Ansätze für eine Abtasttheorie allgemeiner Flächen. Doch ihre algorithmischen Konsequenzen sind noch weitgehend unerforscht.

Für verrauschte Daten ist die Situation ähnlich. Es gibt erste theoretische Arbeiten, die Rauschen im Abtastprozess berücksichtigen. Das hat auch in der Praxis zu besseren Rekonstruktionsalgorithmen geführt. Der Nachteil: Diese Arbeiten benutzen ein sehr eingeschränktes Rauschmodell, das mit dem Rauschen in der Praxis nur bedingt etwas zu tun hat. Überhaupt ist fraglich, ob es eine allgemeine Rausch-Theorie im Abtastprozess geben kann, da das Rauschen stark vom Abtasten abhängt: Tastet ein Laser ein Objekt ab, gibt es andere Rauschcharakteristika als bei einem Abtastprozess mit strukturiertem Licht oder bei einem mechanischen Abtastprozess.

Wie geht es weiter?

Punktwolken und ihre Eigenschaften bleiben also ein spannendes Forschungsfeld. Interessante und tiefe Mathematik lässt sich hier in einer konkreten Anwendung einsetzen und begreifen. Vor allem diese macht den Reiz dieses Gebiets aus. Obwohl noch längst nicht alle Probleme für Punktwolken von Oberflächen gelöst sind, gibt es schon neue Herausforderungen. In Anwendungen wie der Bioinformatik und der Robotik müssen immer häufiger Punktwolken in hochdimensionalen Räumen analysiert werden. Viele Methoden, mit denen sich niedrigdimensionale Punktwolken analysieren lassen, können aber nicht in hohen Dimensionen verwendet werden. Neue Ideen sind also gefragt. Punktwolken werden uns auch künftig beschäftigen und hoffentlich zur Entdeckung besserer Rekonstruktionsalgorithmen und interessanter Mathematik führen. :::

KONTAKT

Joachim Giesen
ABT . 1 Algorithmen und Komplexität
Telefon +49 681 9325-105
Email jgiesen@mpi-inf.mpg.de

Videobasierte Rekonstruktion dynamischer Szenen

Das Ziel dieses Forschungsschwerpunktes ist es, Verfahren zu entwickeln, mit deren Hilfe dreidimensionale Modelle von bewegten Szenen, die von verschiedenen Kameras aufgezeichnet wurden, rekonstruieren kann. Modelle, die so berechnet werden, ermöglichen es, dynamische Szenen im Computer aus beliebigen neuen Blickwinkeln darzustellen. Ein Beispiel: Eine Frau überquert im wehenden Mantel eine Straße – mehrere Videokameras filmen sie aus verschiedenen Blickwinkeln. Damit das Bild möglichst realistisch aussieht, müssen mehrere Aspekte der Szene detailgetreu aus den Videodaten rekonstruiert werden. Hierzu zählen die Bewegung, die Geometrie, die Materialeigenschaften von Objekten sowie die Beleuchtungssituation im Raum. Um den technologischen Herausforderungen wie Szenenaufzeichnung, -berechnung und -wiedergabe gerecht zu werden, müssen neue Algorithmen der Computergraphik und der maschinellen Bilderkennung entwickelt werden. Die hier gewonnenen Erkenntnisse bilden die Grundlage für die Entwicklung der nächsten Generation visueller Medien, insbesondere 3D-Video und 3DTV.
Die Europäische Union fördert anlässlich des 3DTV-Network of Excellence die Forschung zu diesem Thema am Max-Planck-Institut für Informatik. Zwei der Projekte aus dem Forschungsschwerpunkt werden im Folgenden näher betrachtet.

3D-Videos von virtuellen Schauspielern

Menschen stehen in den meisten Szenen der realen Welt im Mittelpunkt. Menschen aus Videoaufnahmen als realistisches Modell zu rekonstruieren, zählt aber zu den schwierigsten Aufgaben. Die Forschungsgruppe hat einen modellbasierten Ansatz entwickelt, der dreidimensionale Videos von Schauspielern aus lediglich acht Videoströmen errechnet. Zu diesem Zweck wird ein Standard-Menschmodell zunächst so verformt, dass es so aussieht wie der Schauspieler. Mit Hilfe des geformten Modells wird anschließend die Bewegung der Person aus den unveränderten Videodaten gemessen. Es sind also keine optischen Markierungen in der Szene erforderlich. Weiterhin werden sowohl die dynamischen Textur- als auch die dynamischen Reflexionseigenschaften der Szenenoberflächen geschätzt. Der Schauspieler kann nun in Echtzeit aus neuen Blickwinkeln und unter neuen simulierten Beleuchtungssituationen gerendert werden [Abbildung 1].

Abbildung 1: Berechnetes und neu beleuchtetes 3D-Video eines Schauspielers

Dynamische Szenenrekonstruktion durch Deformation von Laser-Scans

Abbildung 2: Der Laser-Scan einer Frau im Kimono (rechts) bewegt sich wie die echte Frau in den Videoaufnahmen (links). Die Bewegungsinformation wurde ohne Hilfe eines Skelettmodells oder optischer Markierungen direkt aus den Videodaten errechnet.

Modellbasierte Verfahren erlauben die Rekonstruktion hochwertiger 3D-Videos mit nur wenigen Kameras. Ein Nachteil ist allerdings, dass für jedes Objekt zunächst ein eigenes Standardmodell erzeugt werden muss und dass Aspekte wie die Bewegung weiter Kleidung nicht leicht modellbasiert gemessen werden können. Die Wissenschaftler haben daher neue Algorithmen entwikkelt, mit deren Hilfe ausschließlich die Videodaten analysiert werden. Auf diese Weise lässt sich ein detaillierter statischer Laser-Scan einer Person so bewegen, wie es die Person in der realen Welt tut. Auch die Bewegung von Schauspielern in beliebiger Kleidung (wie etwa einem Kimono, [Abbildung 2]) oder die Bewegung anderer gescannter Objekte lässt sich so detailgetreu messen. :::

KONTAKT

Christian Theobalt
ABT. 4 Computergraphik
Telefon +49 681 9325-419
Email theobalt@mpi-inf.mpg.de

Globale Beleuchtungsberechnung und Bilderzeugung mittels GPU

Eines der bekanntesten Teilprobleme der Computergraphik ist das **Rendering** – die Erstellung möglichst realistisch wirkender Bilder aus dreidimensionalen Szenenbeschreibungen. Die hier auftretenden Probleme sind aus mehreren Gründen sehr komplex. Erstens bestehen typische Szenenbeschreibungen oft aus vielen Tausenden bis Millionen von Primitiven, die effizient gehandhabt werden müssen. Zweitens verlangt **realistisches** Aussehen oft die Berechnung von komplexen **globalen** Effekten wie Schatten, Reflexionen, realistischen Oberflächenbeschreibungen und globalen Beleuchtungsberechnungen (Lichtsimulation). In diesem Zusammenhang bedeutet „global", dass prinzipiell jeder Punkt der Szene jeden anderen Punkt beeinflussen kann. Daraus resultiert ein hoch-dimensionales, kontinuierliches, nicht-stetiges und rekursiv definiertes Problem. Drittens erfordern viele praktische Anwendungen „interaktive" oder „realtime" Bildwiederholungsraten. Das bedeutet, dass alle oben genannten Berechnungen für ein Bild in Sekundenbruchteilen auszuführen sind.

Effiziente und korrekte Bilderzeugung mittels Graphik-Hardware

Eines der Projekte, das am Max-Planck-Institut für Informatik in Kooperation mit der Universität des Saarlandes entwickelt wurde, beschäftigt sich mit der effizienten Umsetzung des Ray Tracing Verfahrens auf Graphik-Hardware. Normalerweise berechnen Graphikkarten ein Bild im Standardverfahren. Ray Tracing hingegen ermöglicht die physikalisch korrekte Simulation der Lichtausbreitung und somit beispielsweise auch korrekte Schatten und Reflexionen [Abbildung 1]. Erste Forschungsergebnisse zeigen, dass nun mit herkömmlicher Graphik-Hardware Ray Tracing in **Echtzeit** realisiert werden kann.

Abbildung 1: Mit dem Ray Tracing Verfahren können nun auch auf herkömmlichen Graphikkarten physikalisch korrekte Schatten und Reflexionen in Echtzeit berechnet werden.

Hochqualitative globale Beleuchtungsberechnung für Animationen

Da die globale Beleuchtungsberechnung hochkomplex ist, müssen bei allen Formen der interaktiven Darstellung noch Kompromisse bei der Simulationsqualität gemacht werden. Daher werden für viele Anwendungen in der Praxis weiterhin „offline"-Animationen berechnet. Um diesen Vorgang sowohl so realistisch als auch so effizient wie möglich zu gestalten, werden am Max-Planck-Institut für Informatik zwei Ansätze verfolgt: die Nutzung temporaler Kohärenz sowie perzeptions-basierte Verfahren. Bei der temporalen Kohärenz wird das Wissen um die Ähnlichkeit aufeinander folgender Bilder einer Animation ausgenutzt: Ähnliche Bilder führen im Allgemeinen auch ähnliche Berechnungen aus, die dann zusammengefasst über mehrere Bilder eine Zeitersparnis bringen können. Da die Realitätsnähe einer Animationssequenz letztlich von einem Menschen beurteilt wird, wird bei perzeptions-basierten Verfahren zusätzlich zur temporalen Kohärenz noch das Wissen über die Funktionsweise der menschlichen Wahrnehmung ausgenutzt. Auf diese Weise lässt sich der Genauigkeitsgrad der Simulation optimal steuern. Damit konzentrieren sich die Berechnungen effizient auf die sichtbaren Aspekte, wodurch deutliche Geschwindigkeitssteigerungen erreicht werden. Durch die Kombination beider am Max-Planck-Institut für Informatik entwickelter Verfahren können Animationen sowohl sehr effizient als auch hochgradig realistisch berechnet werden [Abbildung 2]. :::

Abbildung 2: Die globale Beleuchtungssimulation hat die photorealistische Darstellung computergenerierter Welten zum Ziel.

KONTAKT

Karol Myszkowski
ABT. 4 Computergraphik
Telefon +49 681 9325-429
Email karol@mpi-inf.mpg.de

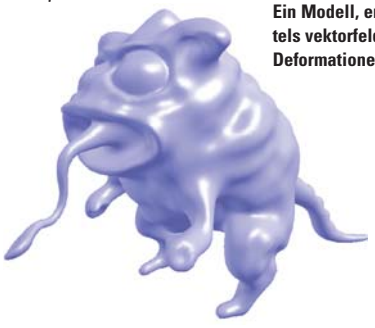
Johannes Günther
ABT. 4 Computergraphik
Telefon +49 681 9325-651
Email guenther@mpi-inf.mpg.de

Topologie-orientierte Verarbeitung von Vektorfeldern

Die Strömungsvisualisierung hat sich in den letzten Jahren zu einem der wichtigsten Teilgebiete der wissenschaftlichen Visualisierung entwickelt. Ein populärer Ansatz für die visuelle Analyse von Strömungsdaten, die oftmals als Vektorfelder gegeben sind, ist die Extraktion von relevanten Features. Von besonderer Bedeutung sind topologische Features. Mit ihrer Hilfe ist es möglich, selbst komplexe Strömungsfelder mit einer recht geringen Anzahl von graphischen Primitiven darzustellen.

Obwohl topologische Methoden seit zirka 15 Jahren zur visuellen Analyse von Strömungsfeldern eingesetzt werden, ist in den letzten Jahren ein neuer Aufschwung dieser Techniken zu verzeichnen. Der Grund hierfür ist zum einen die ständig wachsende Komplexität der zu visualisierenden Daten, zum anderen die schnelle Entwicklung der Hardware, welche es mehr und mehr ermöglicht, rechen- und speicherintensive topologische Analysen großer Datensätze durchzuführen.

In unserer Gruppe wurden eine Reihe von Beiträgen und Anwendungen von topologischen Methoden zur Strömungsvisualisierung erforscht. Schwerpunkte waren bislang die Extraktion globaler topologischer Features in zeitabhängigen Vektorfeldern sowie die Extraktion lokaler topologischer Features in zwei-Parameter-abhängigen Feldern. Gleichzeitig wurden Anwendungen topologischer Methoden in anderen Bereichen der Computergraphik und Visualisierung behandelt: der Einsatz zu Steuerung von Shape-deformations in der Modellierung, die Kompression und Simplifikation von Vektorfeldern sowie die Erfassung des dynamischen Verhaltens von Wirbelkern-Linien.



Ein Modell, erstellt mittels vektorfeld-basierter Deformationen

Vektorfeld-basierte Deformationen

Ziel des Forschungsprojektes war es, neue Ansätze zur Deformation von bestimmten Flächen zu entwickeln. Diese erhalten das Volumen der von der Fläche umschlossenen Körper, schließen Selbstüberschneidungen der Fläche (sowohl lokal als auch global) aus und sind intuitiv und schnell. Die Grundidee war es, derartige Deformationen mithilfe 3D-divergenzfreier Vektorfelder zu beschreiben. Wird die Deformation einer Fläche als Stromobjekt-Integration eines solchen Vektorfeldes interpretiert, so sind die ersten beiden Punkte automatisch erfüllt.

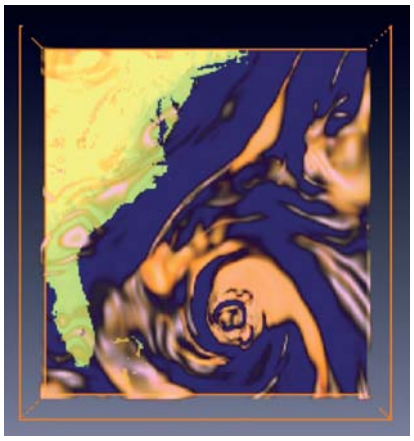
Pfadlinien-orientierte Topologie von zeitabhängigen Vektorfeldern

Ziel der Extraktion und Visualisierung des topologischen Skeletts von Vektorfeldern ist die Separierung von Bereichen mit ähnlichem Strömungsverhalten. Für zeitabhängige Felder kann dieses Verhalten entweder anhand der Stromlinien oder der Pfadlinien betrachtet werden. Entsprechend kann man für zeitabhängige Vektorfelder zwischen einer stromlinien- und einer pfadlinien-orientierten Topologie unterscheiden. Während für die stromlinien-orientierte Topologie bereits eine Anzahl von Ansätzen existiert, ist die Extraktion einer pfadlinien-orientierten Topologie bislang wenig erforscht. Der Hauptgrund hierfür ist, dass topologische Skelette im Allgemeinen von kritischen Punkten ausgehend konstruiert werden, diese jedoch bei Pfadlinien nicht existieren. Ziel des Forschungsschwerpunktes ist es, einen pfadlinien-orientierten Topologie-Ansatz für eine spezielle Klasse von Strömungsfeldern zu entwickeln: die periodischen Strömungsfelder. Grundidee ist, dass für diese Klasse spezielle Pfadlinien existieren, die den kritischen Punkten in stromlinien-orientierten Ansätzen entsprechen. Die Extraktion und Visualisierung dieser speziellen kritischen Pfadlinien bildet die Grundlage des pfadlinien-orientierten topologischen Skeletts.

Multifield-Visualisierung von 3D Skalarfeldern

Die meisten existierenden Visualisierungstechniken für Skalar-/Vektorfelder zielen auf die Analyse der Eigenschaften eines einzelnen Feldes ab. Das Ergebnis vieler Simulationen ist jedoch eine Menge von Feldern über demselben Definitionsbereich.

Ziel des Projektes war es, Ansätze zur visuellen Analyse der Korrelation verschiedener 3D-Skalarfelder über dem gleichen Definitionsbereich zu entwikkeln. Ausgehend von der Tatsache, dass zur Visualisierung der Eigenschaften einzelner Felder eine Vielzahl von Ansätzen existiert, wurden Techniken entwickelt, die sich ausschließlich auf die Visualisierung der Korrelationen zwischen den Feldern konzentrieren. :::



Multifield-Analyse einer Hurrican-Simulation ***(Datensatz erstellt vom Weather Research and Forecast (WRF) model, courtesy of NCAR and the U.S. National Science Foundation (NSF))***

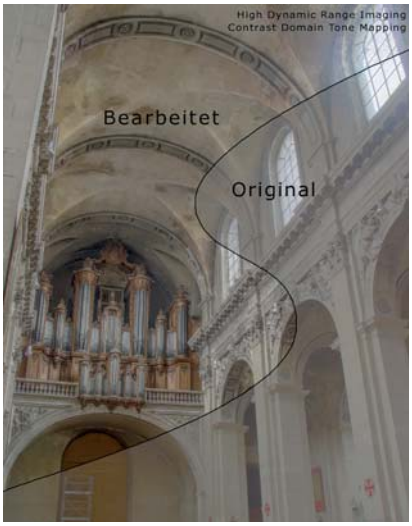
KONTAKT

Holger Theisel
ABT. 4 Computergraphik
Telefon +49 681 9325-400
Email theisel@mpi-inf.mpg.de

HDR – Bilder und Videos mit erhöhtem Kontrastumfang

Die meisten Bild- und Videokameras können nur einen gewissen Teil des Farb- und Helligkeitsspektrums *(high dynamic range)* speichern. Das Auge sieht deutlich mehr Farben und Kontraste. Außerdem sind viele dieser Bilder und Filme qualitativ nicht hochwertig genug, um auf den Bildschirmen der neuesten Generation dargestellt zu werden. So wurde zum Beispiel das weit verbreitete Bildformat JPEG aus Effizienz-Gründen entwickelt: Im JPEG werden nur so viele Informationen gespeichert, wie auf Standardbildschirmen und Ausgabegeräten, die zum Entwicklungszeitpunkt des JPEG-Formats existierten (Kathodenstrahlrohr-Bildschirme und Fernsehgeräte), wiedergegeben werden können. Diese Annahme ist aber nicht mehr aktuell. Heutige LCD- und Plasmabildschirme können eine viel größere Farbskala *(color gamut)* und einen größeren Helligkeitsbereich *(dynamic range)* darstellen als ihre Vorgänger.

Das Tone Mapping komprimiert die hohe Dynamik *(rechter Einsatz)*, indem es das Kontrastverhältnis zwischen den Zonen gleicher Belichtung im zerlegten HDR Bild *(linker Einsatz)* optimiert. Das Ergebnis zeichnet sich durch gute Tonabbildung in allen Bildbereichen aus.



Rendern von HDR-Bildern für Geräte mit limitiertem Kontrastumfang

Die „High Dynamic Range" *(HDR)* Bildverarbeitung *(HDRI)* überwindet die Grenzen der bisherigen Bildverarbeitung, indem alle Farboperationen sehr viel genauer ausgeführt werden. Selbst die Wahrnehmung des menschlichen Auges wird übertroffen. Mit HDRI lassen sich Bilder von natürlichen Szenen mit allen wahrnehmbaren Farben ohne Unter- oder Überbelichtung originalgetreu wiedergeben. HDRI arbeitet also nicht nur präziser, sondern schafft es auch, visuelle Signale in der menschlichen Perzeption zu synthetisieren oder zu visualisieren. So können im Gegensatz zu traditionellen Bildern die HDR-Bilder visuelle Phänomene wie selbstleuchtende Oberflächen (Sonne, Glühlampen), Glanzpunkte, Schatten sowie lebendige, stark gesättigte Farben darstellen.

In der Forschungsgruppe der Computergraphik haben wir neue Video- und Bildformate entwickelt, die natürliche Szenen äußerst genau kodieren. Um die größeren Datenmengen entsprechend verwalten zu können, erreichen diese Formate gute Kompressionsraten. Wir haben außerdem ein Softwarepaket entwickelt (siehe *„pfstools – Bearbeitung von HDR-Bildern und Video"*, Seite 73), das nützliche Programme für die HDR-Bildverarbeitung beinhaltet und für zukünftige Forschungsarbeiten gedacht ist. Mit unserer Arbeit haben wird versucht, das Softwarepaket möglichst unabhängig von spezieller Bildtechnologie zu gestalten. Auf diese Weise schränken uns nur die Fähigkeiten der menschlichen Perzeption ein. Bereits bestehende Bildverarbeitungssoftware und Hardware auf HDR-Daten umzugestalten, erfordert jedoch viel Aufwand. Auch waren einige neue Definitionen von Standards für die Bildverarbeitung notwendig. Im Wesentlichen wollen wir das HDR-Bildverarbeitungs-Konzept popularisieren, neue Standardwerkzeuge und Algorithmen für die Verarbeitung von HDR-Daten entwickeln und die Forschung im Bereich visueller Perzeption vorantreiben, da diese einen entscheidenden Einfluss auf die digitale Bildverarbeitung hat. :::

KONTAKT

Karol Myszkowski
ABT. 4 Computergraphik
Telefon +49 681 9325-429
Email karol@mpi-inf.mpg.de

Rafal Mantiuk
ABT. 4 Computergraphik
Telefon +49 681 9325-427
Email mantiuk@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de/resources/hdr/

Rechnergestützte 3D-Fotografie – Digitalisierung von Geometrie, Struktur und Materialien

In diesem Forschungsgebiet entwickeln wir rechnergestützte Aufnahmeverfahren, mit denen sich digitale Modelle der realen Welt herstellen lassen. Diese Verfahren sind in der Lage, unterschiedlich große Objekte zu vermessen und zu digitalisieren. So lassen sich Objekte aus neuen Blickwinkeln und in beliebigen virtuellen Umgebungen korrekt darstellen, auch ein besserer Überblick über eine Szene kann gegeben werden.

Ein Reflektanzfeld beschreibt die Lichtausbreitung in einer Szene so, dass sie aus beliebigen Perspektiven und unter neuartiger Beleuchtung korrekt wiedergegeben wird.

Nach der Rekonstruktion der 3D-Struktur des Reflexionsnebels NGC 7023 können Ansichten aus verschiedenen Perspektiven gerendert werden.

Akquisition von Reflektanzfeldern

Um ein reales Objekt oder eine Szene fotorealistisch am Rechner wiedergeben zu können, müssen neben der 3D-Geometrie vor allem die Reflexionseigenschaften der beteiligten Materialien erfasst werden. Das Erscheinungsbild lässt sich vollständig durch ein so genanntes Reflektanzfeld beschreiben: Es gibt für jeden einfallenden Lichtstrahl die Intensität des reflektierten Lichts für ausgehende Lichtstrahlen an. Mit einem Reflektanzfeld können alle Lichttransportwege korrekt erfasst und neu synthetisiert werden. Dazu zählen beispielsweise spekulare und diffuse Reflexionen, aber auch globale Beleuchtungseffekte wie Interreflexionen, Brechungen und Kaustiken. Digitale Kameras und Videoprojektoren nehmen die Bilder des Objekts unter strukturierten Lichtmustern auf, die sich wiederum an das Objekt anpassen. Die digitalisierte Szene kann schließlich virtuell mit beliebigen Lichtmustern beleuchtet werden oder aus verschiedenen, nicht aufgenommenen Perspektiven gezeigt werden. Bei gleichen Bedingungen ist das Erscheinungsbild kaum vom Original zu unterscheiden.

Digitalisierung von Straßenzügen

Damit Benutzer sich beim Navigieren in virtuellen Stadtplänen besser orientieren können und gleichzeitig einen Eindruck der Lokalität erhalten, werden Verfahren zu effizienten Akquisition und Darstellung von Straßenzügen entwickelt. Ausgehend von einem groben 3D-Scan und einem Videostrom, den ein vorbeifahrendes Auto aufnimmt, werden längere Straßenabschnitte in einem einzigen Bild zusammengefasst. Das so erzeugte Bild wird glatt aus den Video-Einzelbildern zusammengesetzt und enthält mehrere Perspektiven gleichzeitig. Um ein optimales Bild zu erhalten, wurden die Bilder verschiedener Perspektiven so zusammengefügt, dass es möglichst wenige Verzerrungen gibt. Das Bild unterscheidet sich deutlich von einem natürlichen Foto mit nur einem Blickpunkt: Es gibt das Aussehen der gesamten Straße intuitiv wieder. Der Vorteil dieses Verfahrens liegt auf der Hand: Gegenüber einer Einzelbilder-Kollektion oder dem Videostrom enthält es keine Redundanzen und ist somit deutlich kompakter zu speichern und einfacher zu übertragen. Auch ist die Navigation in einem Bild deutlich intuitiver.

3D-Rekonstruktion von Reflexionsnebeln

Will man astronomische Objekte wie Reflexionsnebel vermessen und digitalisieren, steht vor allem die Rekonstruktion der 3D-Struktur im Vordergrund. Aufgrund der Entfernung stehen nur zweidimensionale Messungen aus einer einzigen Perspektive, von der Erde aus, zur Verfügung. Um trotzdem 3D-Informationen für den beispielsweise oben erwähnten Reflexionsnebel zu gewinnen, wird angenommen, dass der Nebel eine Achsensymmetrie aufweist. Diese wird häufig hervorgerufen, wenn Nebel entstehen. Die Achsensymmetrie reduziert das Problem auf die Bestimmung einer 2D-Dichteverteilung, die in einem nicht-linearen Optimierungsverfahren errechnet wird. Rotiert man die 2D-Verteilung um die Symmetrieachse, erhält man schließlich die 3D-Struktur des Nebels. Daraus lassen sich wiederum mit Volumenrendering 2D-Bilder berechnen. In einem inversen Renderingverfahren wird die 2D-Dichteverteilung so angepasst, dass das synthetisierte Bild möglichst exakt den aufgenommenen Bildern des Nebels entspricht. :::

Multiperspektivisches Bild eines Straßenzuges

KONTAKT

Hendrik Lensch
ABT . 4 Computergraphik
Telefon +49 681 9325-428
Email lensch@mpi-inf.mpg.de

International Max Planck Research School for Computer Science (IMPRS-CS)

Die Ausbildung des wissenschaftlichen Nachwuchses ist von elementarer Bedeutung für die Zukunft von Wissenschaft, Forschung und Innovation in Deutschland. Die Max-Planck-Gesellschaft hat daher gemeinsam mit der Hochschulrektorenkonferenz eine Initiative zur Nachwuchsförderung ins Leben gerufen: die International Max Planck Research Schools (IMPRS). Diese bieten besonders begabten deutschen und ausländischen Studierenden die Möglichkeit, sich im Rahmen einer strukturierten Ausbildung unter exzellenten Forschungsbedingungen auf die Promotion vorzubereiten. Auf diese Weise sollen verstärkt junge Wissenschaftler angeworben und ausgebildet werden.

IMPRS-CS

Förderung des Wissenschaftlichen Nachwuchs

Die IMPRS-CS ist ein Angebot für Nachwuchswissenschaftlerinnen und -wissenschaftler, die zwischen dem ersten berufsqualifizierenden Abschluss, dem Bachelorabschluss, und der Promotion stehen. Wir wollen durch ein erstklassiges, interdisziplinäres Ausbildungsangebot, wissenschaftliche Schwerpunktbildung, thematische Verzahnung der einzelnen Promotionen, die enge Zusammenarbeit von Doktoranden und ihrer Betreuer und nicht zuletzt durch den geförderten sozialen Zusammenhalt aller an der Schule Beteiligten, einen Mehrwert erzeugen.

Ein Schwerpunkt liegt auf der internationalen Zusammenarbeit: Die IMPRS-CS will insbesondere ausländische Bewerberinnen und Bewerber für eine Promotion in Deutschland gewinnen, sie mit den Forschungseinrichtungen vertraut machen und ihr Interesse für eine spätere Tätigkeit in oder in Kooperation mit deutschen Forschungseinrichtungen wecken. Fast 60 Prozent unserer aktuellen Studierenden stammen aus dem Ausland, wobei Bulgarien, Rumänien, China und Russland zu den am stärksten vertretenen Herkunftsländern zählen.

Programme der IMPRS-CS

Die IMPRS-CS bietet in Zusammenarbeit mit der Universität des Saarlandes Programme für den Masterabschluss und die Promotion.

Alle Graduiertenprogramme werden in enger Kooperation mit dem Max-Planck-Institut für Informatik, dem Max-Planck-Institut für Softwaresysteme und dem Fachbereich Informatik der Universität des Saarlandes angeboten. Die drei genannten Institute befinden sich in benachbarten Gebäuden auf dem Universitätscampus. Die Projekte werden gemeinsam von den Wissenschaftlern der Max-Planck-Institute und deren Kollegen aus dem Fachbereich Informatik der Universität betreut. Hervorragende Englischkenntnisse sind für alle Bewerber unerlässlich.

Finanzielle Unterstützung

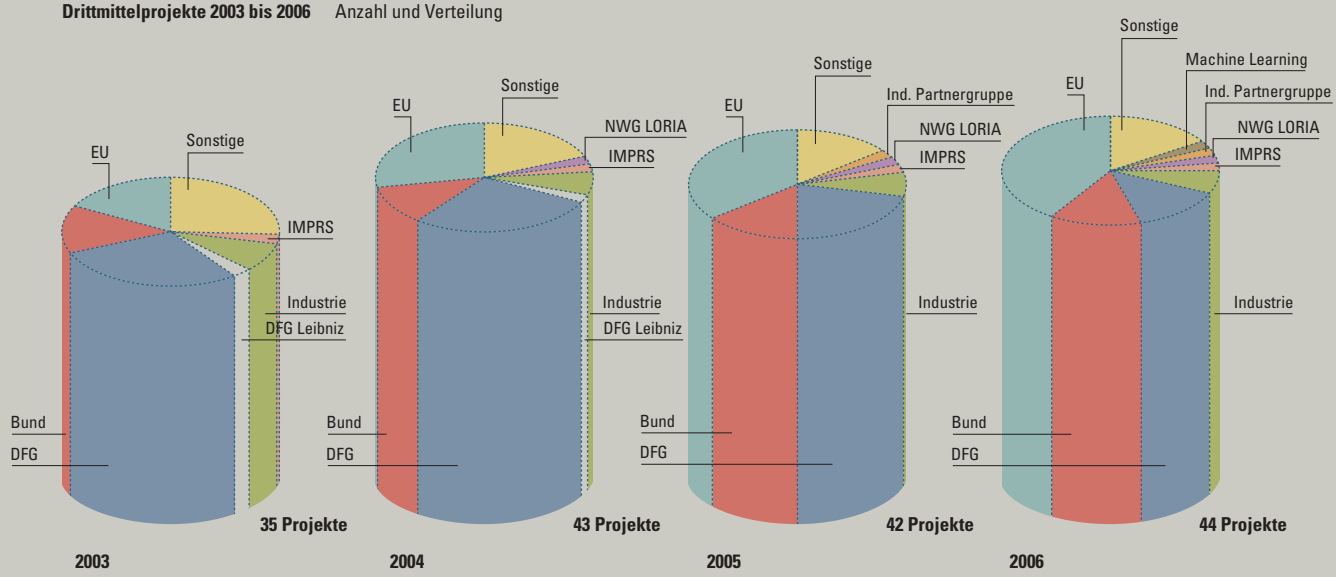
Die zur IMPRS-CS zugelassenen Studierenden erhalten ein Stipendium, das Gebühren, Lebenshaltungskosten und Krankenversicherungskosten sowohl der Studierenden als auch gegebenenfalls ihrer Ehepartner oder Kinder abdeckt. Außerdem helfen wir unseren Stipendiaten bei der Wohnungssuche und organisatorischen Problemen aller Art, bieten Englisch- und Deutschkurse auf mehreren Niveaus, Freizeitaktivitäten und Exkursionen an. :::

KONTAKT

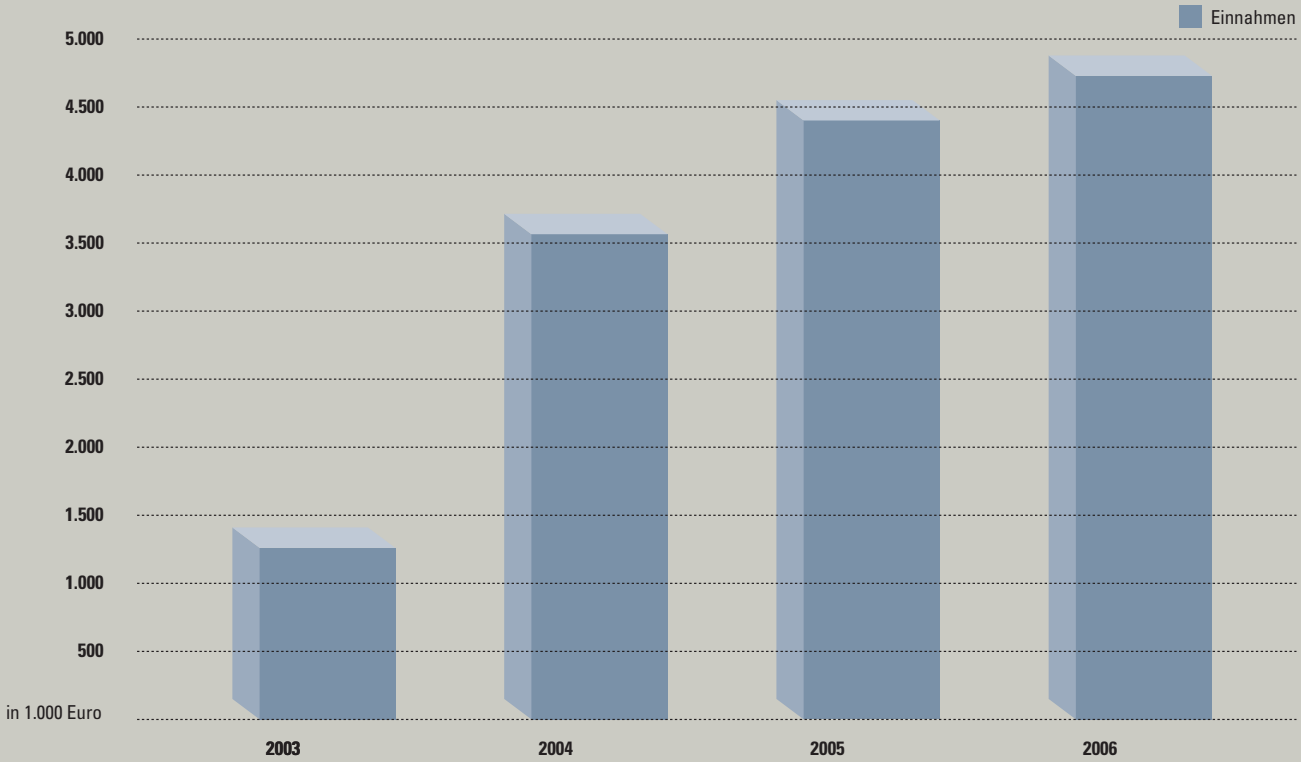
Kerstin Kathy Meyer-Ross
IMPRS-CS
Telefon +49 681 9325-226
Email kmeyer@mpi-inf.mpg.de
Internet http://www.imprs-cs.de

Das Institut in Zahlen

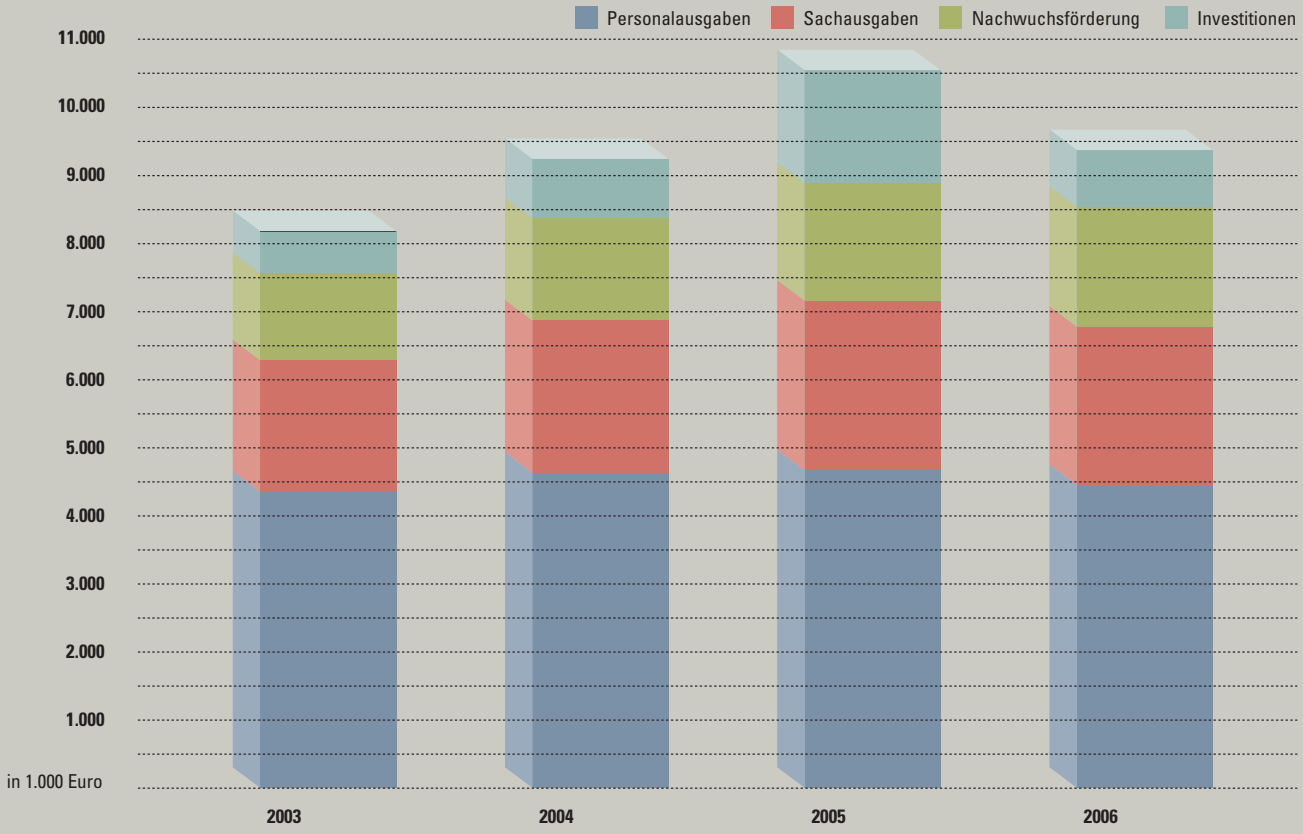
Drittmittelprojekte 2003 bis 2006 Anzahl und Verteilung



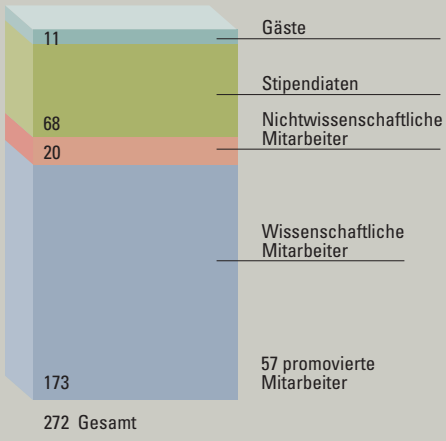
Drittmittelprojekte 2003 bis 2006



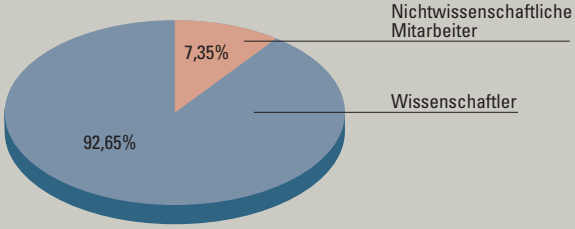
Betriebsmittel 2003 bis 2006



Mitarbeiter am Institut Stand 1.2.2007



Verhältnis von Wissenschaftlern zu nichtwissenschaftlichen Mitarbeitern am Institut Stand 1.2.2007



KONTAKT

Volker Geiß
Gemeinsame Verwaltung
Telefon +49 681 9325-700
Email geiss@mpi-inf.mpg.de

Rechnerbetrieb

Ungehinderte weltweite Kooperation und Kommunikation in einem motivierenden Umfeld bilden die Basis für ein Institut mit dem Anspruch, erstklassige Forschungsergebnisse zu erbringen. Flexibilität, Qualität und Zuverlässigkeit der Ausstattung sowie ihre einfache Nutzbarkeit leisten dazu einen entscheidenden Beitrag.

Dieser Anspruch lässt sich auf unsere Rechnerinfrastruktur übertragen: Wir betreiben ein vielfältiges, sich schnell änderndes System, das sich dem Anwender einheitlich und verlässlich präsentiert, und das trotz der notwendigen Offenheit zur Unterstützung internationaler Kooperationen die Sicherheit nicht vernachlässigt.

Vielfalt der Werkzeuge

Wir setzen Systeme unterschiedlicher Hersteller ein. Die Auswahl wird durch die Anforderungen der Projekte bestimmt. Die zunehmende Verfügbarkeit und preisliche Attraktivität von 64-Bit-Systemen auf der Basis von AMD oder Intel CPUs hat andere Systeme weitgehend verdrängt. Lediglich bei Datenbank- oder File-Server-Systemen setzen wir auf Grund der Architektur-Vorteile noch SUN-Ultra-SPARC-Systeme ein. Bei den Betriebssystemen verhält es sich ähnlich: Linux und Windows dominieren sowohl den Notebook- und Workstation- als auch den Server-Einsatz.

Dynamik und Innovation

Forschung an vorderster Front heißt auch, immer wieder innovative Technologien einzusetzen. Vor allem in der Computer-Grafik implizieren die möglichen Performance-Steigerungen und die Erweiterung des Funktionsumfangs durch den Einsatz neuester Hardware nicht selten die Notwendigkeit des Einsatzes von Prototypen.

Weitgehend einheitliche Nutzung

Die Forschungsprojekte sind häufig plattformübergreifend angelegt, da sie entweder für einen heterogenen Verbund gedacht sind, oder aber die Vorzüge der verschiedenen Rechner- und Betriebssystemarchitekturen ausnutzen müssen.

Mit Ausnahme weniger Spezialsysteme sind daher alle Maschinen für jeden unserer ca. 500 Anwender (Mitarbeiter, Studenten, Projektpartner) ohne besonderen Aufwand direkt nutzbar – die User-Daten und die wichtigsten Softwarepakete sind plattformunabhängig verfügbar. Diese Homogenität erleichtert den Umgang mit dem Gesamtsystem erheblich und fördert seine Akzeptanz.

Zuverlässigkeit der Installationen

Ständige Updates und Upgrades bei Soft- und Hardware in Verbindung mit dem Wunsch, plattformübergreifend zu arbeiten, stellen hohe Anforderungen an die Zuverlässigkeit der Installations- und Administrationsvorgänge. Sie müssen reproduzierbar und verlässlich sein. Nach einer Umstellung oder Neuinstallation eines Systems sollte niemand gezwungen sein, seine Experimente oder gar seine Arbeitsweise anzupassen.

Kooperation und Kommunikation

Unser Netzwerk ist nach organisatorischen und sicherheitsrelevanten Gesichtspunkten in verschiedene Bereiche aufgeteilt. Die Endgeräte in den Teilnetzen der einzelnen Bereiche werden über ihren Etagen-Switch mit 10/100/1000-Mbit-Ethernet versorgt. Die Switches sind ausfallsicher mit 10-Gigabit-Ethernet an die Zentrale (zwei redundante Backbone-Switches) angeschlossen, die auch zentrale Server mit dieser Bandbreite versorgt. Server-Farmen und Compute-Cluster sind über eigene Switches redundant mit der Zentrale verbunden.

Der externe Bereich umfasst die Verbindungen zur Universität des Saarlandes und zum Internet. Der Internet-Anschluss wird über einen mit der Universität gemeinsam genutzten 1,3 Gigabit-XWIN-Anschluss des DFN-Vereins realisiert. Extern zugreifbare aber anonyme Dienste (DNS (Internet-Adressbuch), WWW, FTP (Datentransfer), SMTP (E-Mail) etc.) werden an der Firewall des Institutes in einer entmilitarisierten Zone (DMZ) zusammengefasst.

Logisch ist das Netzwerk so strukturiert, dass die Integration von Gastwissenschaftlern und Studenten durch die Möglichkeit gefördert wird, mitgebrachte Notebooks ohne zusätzliche Softwareinstallation anschließen oder über WLAN betreiben zu können. Sie werden unabhängig von ihrem Standort in speziell dafür vorgesehene Teilnetze gelenkt. Auch diese Netze werden in einer entmilitarisierten Zone zusammengefasst.

Die internationale Kooperation verlangt externe Zugriffsmöglichkeiten auf interne Ressourcen der Infrastruktur (Intranet). Hier bieten wir unter anderem einen Terminal-Zugang und den Zugriff auf E-Mail und andere wichtige Datenbanken und Dienste an. Die Kooperation bei der Softwareentwicklung wird durch einen geschützten Zugang zu einer Versionsdatenbank unterstützt (Software-Repository: Subversion).

Sicherheit und Schutz

Pauschale Schutzmechanismen gegen Sabotage und Spionage sind in offenen Systemen nicht möglich. Sie schränken ihre Nutzbarkeit zu sehr ein. Die Sicherheitsrichtlinien können deshalb nur ein Kompromiss sein, der flexibel den Anforderungen folgt.

Zwar können die Strukturierung des Netzwerks, die Firewall, die Verschlüsselung für externe Zugriffe oder der Virenscanner im Mail-Server einige direkte Gefahren abwenden. Indirekte Gefahren, die unter anderem durch den Anschluss virenverseuchter Rechner im Intranet oder durch Fehler in extern agierenden Software-Systemen entstehen, können nur durch aktuelle Softwarestände und kontinuierlich aktualisierte Virenscanner auf den Maschinen verhindert werden.

Automatisierung als Garantie

Betrachtet man die zuvor beschriebenen Merkmale der Infrastruktur, so ist ihre ständige Weiterentwicklung und Anpassung an die neusten Anforderungen der Forschungsprojekte und des Sicherheitskonzeptes unter Beibehaltung der Homogenität und Verlässlichkeit unsere zentrale Aufgabe.

Bei der Vielzahl der eingesetzten Hard- und Softwarekomponenten, sind es die weitgehend von uns konzipierten und weiterentwickelten automatisierten Abläufe, die es uns ermöglichen, die notwendigen Arbeiten in angemessener Zeit auf den betroffenen Systemen zu erledigen.

In einer einheitlichen Hardwareumgebung ließe sich die Automatisierung in vielen Fällen durch Vervielfältigung besonders gepflegter Installationen durchführen. Unser Umfeld dagegen ist zu dynamisch und zu heterogen für diese Vorgehensweise.

Aus diesem Grund favorisieren wir Betriebssysteme, deren Installationsmechanismen paketorientiert sind. Solaris, aber vor allem Debian im Linux-Umfeld setzen auf Paketsysteme, die Abhängigkeiten zwischen den installierten Paketen berücksichtigen. Für die Windows-Plattformen haben wir ein solches Paket-System mit Hilfe zusätzlicher Software eingeführt (netinstall).

Dem Beispiel von Solaris folgend, haben wir für die aktuellen Betriebssystemversionen von Debian und Windows ein Installations- und Administrationssystem implementiert. Durch die Wartung und Weiterentwicklung dieses Systems administrieren wir automatisch alle betreuten Systeme und steuern Neuinstallationen. Einmal erzielte Ergebnisse sind beliebig wiederholbar und sehr schnell auf die ganze Infrastruktur anzuwenden. Diese Implementierungsarbeit kostet allerdings Zeit und verlangsamt in manchen Fällen die Reaktionszeiten. Die Vorteile für die Betriebssicherheit wiegen diesen Nachteil eindeutig auf.

Das System ist so flexibel gestaltet, dass auch Sonderfälle durch spezifische Erweiterungen schnell realisiert werden können. Nicht sinnvoll wäre allerdings die Integration kurzlebiger Spezialinstallationen.

Compute-Service

Neben Workstations, Notebooks und einigen kleineren Servern betreiben wir eine F15000 (SUN) mit 72 eng gekoppelten CPUs und 216 GB Hauptspeicher für Applikationen, die eine hohe Parallelität und den uniformen Zugriff auf großen Hauptspeicher benötigen. Zum Zeitpunkt der Beschaffung im Jahre 2002 verfehlte dieses System nur knapp den Einzug in die Liste der 500 weltweit leistungsfähigsten Rechner (TOP 500).

Im Jahre 2003 wurde ein mit Gigabit-Ethernet vernetztes Dual-Xeon-Cluster aus 30 DELL-Systemen aufgebaut. Solche mit verhältnismäßig geringer Bandbreite gekoppelten Systeme eignen sich schlecht für parallele Berechnungen, die mit gemeinsamem Speicher arbeiten. Um diesen Nachteil zu mildern, wurden die Clustersysteme zusätzlich mit einem speziellen Netzwerk (Scalable Coherent Interface) verbunden, das durch seine kurzen Übertragungszeiten in der Lage ist, gemeinsamen Speicher für die beteiligten Systeme zu simulieren. Hier werden parallele Berechnungen, insbesondere Molekular- und Proteindockingverfahren aus dem Bereich der Bioinformatik durchgeführt.

Die größte aggregierte Rechenleistung wird durch unser 2005 in Betrieb genommenes Dual-Opteron-Cluster (96 Systeme) mit Gigabit-Netzwerk zur Verfügung gestellt, das unter der Grid-Engine betrieben wird. Durch die automatische Verteilung der Prozesse auf die einzelnen Rechner des Clusters erreichen wir eine hohe Auslastung des Gesamtsystems.

File-Service und Datensicherung

Die Daten des Instituts werden über mehrere File-Server per NFS und CIFS (SMB) zur Verfügung gestellt. Die mittlerweile ca. 30 TB zentralen Daten sind auf mehrere RAID-Systeme verteilt, die über ein „Storage Area Network“ (SAN) angeschlossen sind. Alle RAID-Systeme werden paarweise gespiegelt betrieben, um auch gegen den Ausfall eines ganzen RAID-Systems geschützt zu sein.

Unsere Datensicherung basiert auf zwei unterschiedlichen Systemen: einem konventionellen Band-Backup, das die Daten direkt mittels Legato-Networker auf zwei Band-Roboter sichert und einem Online-Disk-Backup-System, das mit Hilfe von Datenvergleichen den Platzbedarf minimiert (Open-Source-System „backuppc“) und die gesicherten Daten immer Online bereit hält. Die Zahlen dieser Online-Datensicherung sind beeindruckend. Da keine Daten in diesem System wirklich doppelt gehalten werden, können ca. 71 Terabyte Brutto-Daten der verschiedenen Backupläufe auf etwa 11 Terabyte untergebracht werden. Um die Vorteile von Disk- und

Tapetechnologien zu vereinen, werden wir dieses System in Zukunft mit dem Bandroboter kombinieren.

Spezialsysteme

Für spezielle Forschungsaufgaben, insbesondere aus dem Bereich der Computergrafik, werden diverse Spezialsysteme benötigt. Es stehen u.a. ein digitales Videoschnittsystem, mehrere 3D-Scanner, Multivideoaufnahmesysteme, Videokonferenzsysteme und 3D-Projektionssysteme zur Verfügung.

Betriebssicherheit

Die Maßnahmen zur Sicherheit (Firewall, VPN), die automatisierten Installations- und Konfigurationswerkzeuge, der Betrieb ausfallsicherer Disksysteme, der Betrieb eines ausfallsicheren Netzwerk-Backbones und die Datensicherung werden durch ein Überwachungssystem ergänzt, das über kritische Zustände der Serversysteme und des Netzwerks aber auch über Fehlfunktionen komplexer Prozesse per E-Mail und SMS informiert.

Stromversorgung und Kühlung sind so konzipiert, dass der Serverbetrieb auch bei Stromausfällen aufrechterhalten werden kann. Eine doppelte Absicherung über Generatorbetrieb ist in Vorbereitung und garantiert in Zukunft den durchgängigen Betrieb aller zentralen Systeme. Zurzeit sind nur die wichtigsten Systeme auf diese Art abgesichert.

Zuständigkeiten

Einkauf, Installation, Administration und Fortschreibung der beschriebenen Systeme und Techniken sind Aufgabe der Rechnerbetriebsgruppe des Instituts. Neben Leitung, Einkauf und Sekretariat (zwei Stellen) umfasst das Team fünf wissenschaftliche Mitarbeiter, einen Techniker und drei studentische Hilfskräfte. Wir ergänzen uns mit dem Team des MPI für Softwaresysteme, das inklusive der Leitung vier wissenschaftliche Mitarbeiter mit einbringt.

Als Kommunikationsschnittstelle für die Anwender bilden sechs weitere studentische Hilfskräfte einen Helpdesk, der einerseits virtuell per Mail oder Web-Interface, zu Bürozeiten aber auch persönlich erreichbar ist. Neben der Bearbeitung von Fragen zur Benutzung der Infrastruktur werden auch Informationssysteme gepflegt. Dazu gehört eine Zusammenfassung der interessanten Fragen und Antworten (FAQ) und ein Bulletin-Board. :::

KONTAKT

Jörg Herrmann
IT-Abteilung
Telefon +49 681 9325-800
Email jh@mpi-sws.mpg.de
Internet http://www.mpi-inf.mpg.de/services/computer/

Kooperationen

BIOINFORMATIK

::: Biochemisches Institut, Universität des Saarlandes, *Saarbrücken, Deutschland*

::: BiosolveIT, *Sankt Augustin, Deutschland*

::: British Columbia Centre for Excellence in HIV/AIDS, *Vancouver, Kanada*

::: Christian-Albrechts-Universität Kiel, *Kiel, Deutschland*

::: European Bioinformatics Institute, *Cambridge, Großbritannien*

::: Freie Universität Berlin, *Berlin, Deutschland*

::: Harvard Universität, *Cambridge, UK*

::: Institut für Medizinische Mikrobiologie und Immunologie, Rheinische Friedrich-Wilhelms-Universität Bonn, *Bonn, Deutschland*

::: Institut für Mikrobiologie, Universität des Saarlandes, *Saarbrücken, Deutschland*

::: Institut für Pharmazeutische Chemie, Universität des Saarlandes, *Saarbrücken, Deutschland*

::: Institut für Virologie, Universität Köln, *Köln, Deutschland*

::: Institut Pasteur, *Paris, Frankreich*

::: Johann-Wolfgang-Goethe-Universität Frankfurt, *Frankfurt, Deutschland*

::: John Hopkins University, *Baltimore, USA*

::: Kaiser-Permanente Medical Care Program Northern California, *San Francisco, USA*

::: Max-Delbrück-Centrum für Molekulare Medizin, *Berlin, Deutschland*

::: Max-Planck-Institut für Genetik, *Berlin, Deutschland*

::: Ruprechts-Karls-Universität Heidelberg, *Heidelberg, Deutschland*

::: Stanford University, *Stanford, USA*

::: Wellcome Trust Sanger Institute, *Cambridge, Großbritannien*

::: Università di Roma „Tor Vergata“, *Rom, Italien*

::: Universität des Saarlandes, *Saarbrücken, Deutschland*

::: Universität Dortmund, *Dortmund, Deutschland*

::: Universitätsklinikum des Saarlandes, *Homburg, Deutschland*

::: Universitat Pompeu Fabra, *Barcelona, Spanien*

:: Universiteit van Amsterdam *Amsterdam, Niederlande*

::: University of Kansas, *Kansas, USA*

::: Zentrum für Bioinformatik, Universität des Saarlandes, *Saarbrücken, Deutschland*

GEOMETRIE

::: Ben Gurion University Tel Aviv, *Tel Aviv, Israel*

::: IMATI, *Genua, Italien*

::: INRIA, *Sophia Antipolis, Frankreich*

::: LORIA/INRIA Lorraine, *Nancy, Frankreich*

::: Ohio State University, *Ohio, USA*

::: RIKEN, *Tokio, Japan*

::: Stanford University, *Stanford, USA*

::: Technion-Institute of Technology, *Haifa, Israel*

::: Technische Universität Berlin, *Berlin, Deutschland*

::: Universität des Saarlandes, *Saarbrücken, Deutschland*

::: Universität Mannheim, *Mannheim, Deutschland*

GARANTIEN

::: Albert-Ludwigs-Universität Freiburg, *Freiburg, Deutschland*

::: Carl von Ossietzky-Universität Oldenburg, *Oldenburg, Deutschland*

::: Indian Institute of Technology, *New Delhi, Indien*

::: Masaryk University of Brno, *Brno, Tschechische Republik*

::: Microsoft Research, *Cambridge, UK*

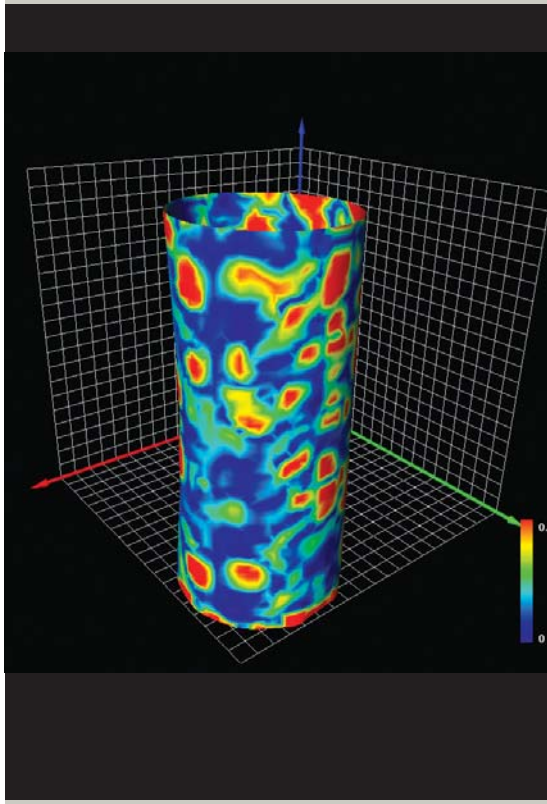
::: Research Academic Computer Technology Institute, *Patras, Griechenland*

::: Rice University Houston, *Houston, USA*

::: Telenor, *Fornebu, Norwegen*

::: Università di Bologna, *Bologna, Italien*

::: Università di Roma „La Sapienza“, *Rom, Italien*



::: Universität des Saarlandes, *Saarbrücken, Deutschland*

::: Universität Duisburg-Essen, *Duisburg, Deutschland*

::: Universität Karlsruhe, *Karlsruhe, Deutschland*

::: Universität Paderborn, *Paderborn, Deutschland*

::: Universität Tel Aviv, *Tel Aviv, Israel*

::: Universitat Polytecnica de Catalunya, *Barcelona, Spanien*

INTERNET

::: Deutsche Telekom Laboratories, *Berlin, Deutschland*

::: Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), *Saarbrücken, Deutschland*

::: Internationales Begegnungs- und Forschungszentrum Schloss Dagstuhl, *Dagstuhl, Deutschland*

::: SAP Research Sophia Antipolis, *Sophia Antipolis, Frankreich*

::: Technische Universität München, *München, Deutschland*

::: Telenor, *Fornebu, Norwegen*

::: Università di Roma „La Sapienza“, *Rom, Italien*

::: Universität Koblenz, *Koblenz-Landau, Deutschland*

::: Yahoo! Research, *Barcelona, Spanien*

::: Microsoft Research, *Redmond, USA*

OPTIMIERUNG

::: Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, *Budapest, Ungarn*

::: Athens University of Economics & Business, *Athen, Griechenland*

::: Christian-Albrechts-Universität Kiel, *Kiel, Deutschland*

::: Cornell University, *Ithaca, USA*

::: Courant Institute, Universität New York, *New York, USA*

::: ETH Zürich, *Zürich, Schweiz*

::: Friedrich-Schiller-Universität Jena, *Jena, Deutschland*

::: INRIA Futurs, *Paris, Frankreich*

::: Institute of Mathematics, Universität Budapest, *Budapest, Ungarn*

::: Stanford University, *Stanford, USA*

::: Technische Universität Carolo-Wilhelmina zu Braunschweig, *Braunschweig, Deutschland*

::: Università di Roma „La Sapienza“, *Rom, Italien*

::: Università di Roma „Tor Vergata“, *Rom, Italien*

::: Universität Aarhus, *Aarhus, Dänemark*

::: Universität Auckland, *Auckland, Neuseeland*

::: Universität Birmingham, *Birmingham, UK*

::: Universität Bonn, *Bonn, Deutschland*

::: Universität Dortmund, *Dortmund, Deutschland*

::: Universität Duisburg-Essen, *Duisburg, Deutschland*

::: Universität Karlsruhe, *Karlsruhe, Deutschland*

::: Universität Paderborn, *Paderborn, Deutschland*

::: Universität Siegen, *Siegen, Deutschland*

::: Universität Tel Aviv, *Tel Aviv, Israel*

::: University College London, *London, UK*

::: University of Halifax, *Halifax, Kanada*

::: University of Southern Denmark, *Odense, Dänemark*

::: University of Washington, *Seattle, USA*

::: University of Wisconsin, *Madison, USA*

SOFTWARE

::: Centrum voor Wiskunde en Informatica (CWI), *Amsterdam, Niederlande*

::: Queen Mary University, *London, Großbritannien*

::: Royal School of Library and Information Science, *Kopenhagen, Dänemark*

::: Universität Duisburg-Essen, *Duisburg, Deutschland*

::: University of Cambridge, *Cambridge, UK*

STATISCHES LERNEN

::: DaimlerChrysler AG, Ulm, *Ulm, Deutschland*

::: Deutsches Krebsforschungszentrum Heidelberg, *Heidelberg, Deutschland*

::: Friedrich-Alexander-Universität Erlangen-Nürnberg, *Erlangen, Deutschland*

::: Harvard University, *Cambridge, USA*

::: Heinrich-Heine-Universität Düsseldorf, *Düsseldorf, Deutschland*

::: Institute for Cancer Research, Dept. of Tumor Biology, Rikshospitalet-Radiumhospitalet Medical Center, *Oslo, Norwegen*

::: Max-Planck-Institut für biologische Kybernetik, *Tübingen, Deutschland*

::: nugg-.ad AG, *Berlin, Deutschland*

::: Ruprecht-Karls-Universität, *Heidelberg, Deutschland*

::: Strato AG, *Berlin, Deutschland*

::: Universität Bonn, FB Neuropathology, *Bonn, Deutschland*

::: Universität des Saarlandes, FR Biowissenschaften, Genetik/ Epigenetik, *Saarbrücken, Deutschland*

::: Universitätsklinikum des Saarlandes, *Homburg, Deutschland*

::: Zentrum für Innovative Genetische Diagnostik, *Homburg, Deutschland*

VISUALISIERUNG

::: Eidgenössische Technische Hochschule Zürich, *Zürich, Schweiz*

EMPA Eidgenössische Materialprüfanstalt, Abteilung Hochleistungskeramik, *Dübendorf, Schweiz*

::: Institut National de Recherche en Informatique et en Automatique, *Sophia Antipolis, Frankreich*

::: Daimler Chrysler, Virtual Reality Center, *Sindelfingen, Deutschland*

::: BrightSide Technologies Inc., *Vancouver, Kanada*

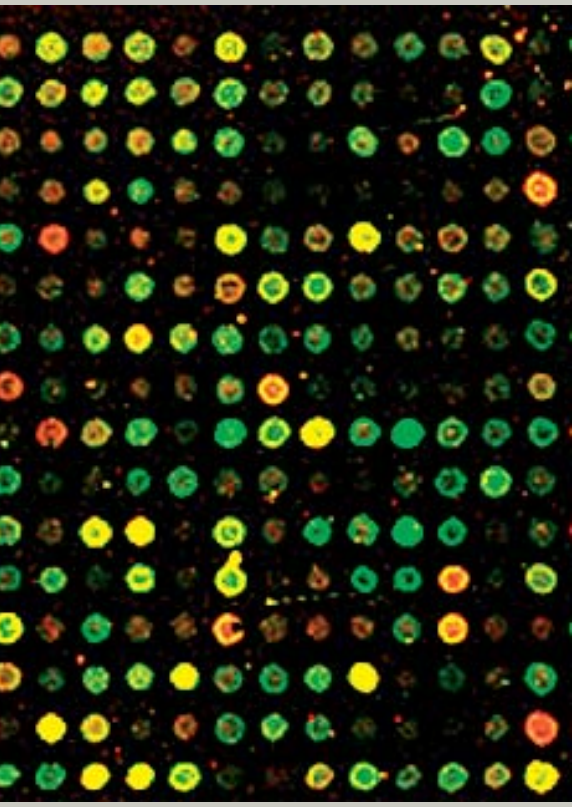
::: Zuse-Institut Berlin, *Berlin, Deutschland*

::: VRVis Institut für Virtual Reality und Visualisierung, *Wien, Österreich*

::: Universität Bielefeld, *Bielefeld, Deutschland*

::: ICIMAF Havanna, *Havanna, Kuba*

Ausgewählte Publikationen 2005 | 2006



[1] D. Ajwani, R. Dementiev and U. Meyer. A computational study of external memory BFS algorithms. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-06),* Miami, USA, 2006, pp. 601–610. ACM / Siam.

[2] I. Albrecht, M. Schröder, J. Haber and H.-P. Seidel. Mixed feelings: Expression of non-basic emotions in a muscle-based talking head. *Virtual Reality,* 8(4):201–212, September 2005.

[3] M. Albrecht. LRRK2 mutations and parkinsonism. *The Lancet,* 365(9466):1230–1230, 2005.

[4] M. Albrecht, C. Huthmacher, S. C. Tosatto and T. Lengauer. Decomposing protein networks into domain-domain interactions. *Bioinformatics,* 21(Suppl. 2):ii220–ii221, 2005.

[5] A. Alexa, J. Rahnenführer and T. Lengauer. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics,* 22(13):1600–1607, 2006.

[6] E. Althaus and R. Naujoks. Computing steiner minimal trees in hamming metric. In *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-06),* Miami, USA, 2006, pp. 172–181. ACM / Siam.

[7] R. Angelova and G. Weikum. Graph-based text classification: Learn from your neighbors. In E. N. Efthimiadis, S. T. Dumais, D. Hawking and K. Jaervelin, eds., *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Seattle, Washington, 2006, pp. 485–492. ACM. Acceptance ratio 1:5 (74 of 399).

[8] I. Antes, C. Merkwirth and T. Lengauer. Poem: Parameter optimization using ensemble methods: Application to target specific scoring functions. *Journal of Chemical Information and Modeling,* 45(5):1291–1302, 2005.

[9] R. Bargmann, V. Blanz and H.-P. Seidel. Learning-based facial rearticulation using streams of 3D scans. In B.-Y. Chen, ed., *The 14th Pacific Conference on Computer Graphics and Applications,* Taipei, Taiwan, October 2006, pp. 232–241. National Taiwan University.

[10] H. Bast, G. Dupret, D. Majumdar and B. Piwowarski. Discovering a term taxonomy from term similarities using principal component analysis. In M. Ackermann, B. Berendt, M. Grobelnik and V. Svatek, eds., *Semantics, Web and Mining 2005, Lecture Notes in Artificial Intelligence (LNAI),* vol. 4289, pp. 103–120. Springer, Berlin Heidelberg, 2006.

[11] H. Bast, S. Funke and D. Matijevic. Transit: Ultrafast shortest-path queries with linear-time preprocessing. In C. Demetrescu, A. Goldberg and D. Johnson, eds., *9th DIMACS Implementation Challenge – Shortest Path,* 2006.

[12] H. Bast, S. Funke, D. Matijevic, P. Sanders and D. Schultes. In transit to constant time shortest-path queries in road networks. In D. Applegate and G. Brodal, eds., *9th Workshop on Algorithm Enginneering and Experiments (ALENEX'07),* New Orleans, USA, 2007. SIAM.

[13] H. Bast, D. Majumdar, R. Schenkel, M. Theobald and G. Weikum. IO-Top-k: Index-access optimized top-k query processing. In U. Dayal, K.-Y. Whang, D. B. Lomet, G. Alonso, G. M. Lohman, M. L. Kersten, S. K. Cha and Y.-K. Kim, eds., *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006),* Seoul, Korea, 2006, pp. 475–486. ACM. Acceptance ratio 1:7.

[14] H. Bast and I. Weber. Type less, find more: Fast autocompletion search with a succinct index. In E. N. Efthimiadis, S. Dumais, D. Hawking and K. Järvellin, eds., *SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Seattle, USA, 2006, pp. 364–371. ACM.

[15] S. Baswana, V. Goyal and S. Sen. All-pairs nearly 2-approximate shortest paths in $o(n^2 \text{polylog} n)$ time. In V. Diekert and B. Durand, eds., *STACS 2005, 22nd Annual Symposium on Theoretical Aspects of Computer Science,* Stuttgart, Germany, 2005, *LNCS 3404,* pp. 666–679. Springer.

[16] P. Baumgartner, A. Fuchs and C. Tinelli. Implementing the model evolution calculus. *International Journal on Artificial Intelligence Tools,* 15(1):21–52, 2006.

[17] P. Baumgartner and C. Tinelli. The model evolution calculus with equality. In R. Nieuwenhuis, ed., *Automated deduction – CADE-20, 20th International Conference on Automated Deduction,* Tallinn, Estonia, 2005, *LNAI 3632,* pp. 392–408. Springer.

[18] N. Beerenwinkel, M. Däumer, T. Sing, J. Rahnenführer, T. Lengauer, J. Selbig, D. Hoffmann and R. Kaiser. Estimating HIV evolutionary pathways and the genetic barrier to drug resistance. *The Journal of Infectious Diseases,* 191(11):1953–1960, 2005.

[19] N. Beerenwinkel, J. Rahnenführer, M. Däumer, D. Hoffmann, R. Kaiser, J. Selbig and T. Lengauer. Learning multiple evolutionary pathways from cross-sectional data. *Journal of Computational Biology,* 12(6):584–598, 2005.

[20] R. Beier, A. Czumaj, P. Krysta and B. Vöcking. Computing equilibria for a service provider game with (im)perfect information. *ACM Transactions on Algorithms,* 2(4):679–706, October 2006.

[21] R. Beier and B. Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM Journal on Computing,* 35(4):855–881, February 2006.

[22] A. Belyaev. On transfinite barycentric coordinates. In D. W. Fellner, S. N. Spencer, A. Sheffer and K. Polthier, eds., *SGP 2006: Fourth Eurographics/ ACM SIGGRAPH Symposium on Geometry Pro-*

cessing, Cagliari, Sardinia, Italy, 2006, pp. 89–99. Eurographics.

[23] M. Bender, S. Michel, P. Triantafillou and G. Weikum. Global document frequency estimation in peer-to-peer web search. In D. Zhou, ed., *9th Internation Workshop on the Web and Databases (WebDB 2006) @ SIGMOD2006,* n/a, 2007, pp. 69–74. n/a. Acceptance Ratio 12:48.

[24] M. Bender, S. Michel, P. Triantafillou, G. Weikum and C. Zimmer. Improving collection selection with overlap-awareness. In R. A. Baeza-Yates, N. Ziviani, G. Marchionini, A. Moffat and J. Tait, eds., *SIGIR 2005, Proceedings of the Twenty-Eighth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '05),* Salvador, Brazil, 2005, pp. 67–74. ACM. Acceptance ratio 1:5 (71 of 368).

[25] M. Bender, S. Michel, P. Triantafillou, G. Weikum and C. Zimmer. MINERVA: Collaborative P2P search (demo). In K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-A. Larson and B. C. Ooi, eds., *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005),* Trondheim, Norway, 2005, pp. 1263–1266. ACM. Acceptance Ratio: 1:2.3 (Demo Track) 1:6.6 (Research Track).

[26] M. Bender, S. Michel, P. Triantafillou, G. Weikum and C. Zimmer. "to infinity and beyond": P2P web search with Minerva and Minerva1. In R. Baldoni, G. Cortese, F. Davide and A. Melpignano, eds., *Global Data Management,* vol. 8, ch. Applications, pp. 301–323. IOSPress, Amsterdam, The Netherlands, 2006.

[27] M. Bender, S. Michel, G. Weikum and C. Zimmer. Das MINERVA-Projekt: Datenbankselektion für Peer-to-Peer-Websuche. *Informatik – Forschung und Entwicklung,* 20(3):152 – 166, December 2005.

[28] E. Berberich, A. Eigenwillig, M. Hemmer, S. Hert, L. Kettner, K. Mehlhorn, J. Reichel, S. Schmitt, E. Schömer and N. Wolpert. Exacus: Efficient and exact algorithms for curves and surfaces. In G. S. Brodal and S. Leonardi, eds., *13th Annual European Symposium on Algorithms (ESA 2005),* Palma de Mallorca, Spain, October 2005, LNCS 3669, pp. 155–166. Springer.

[29] K. Berberich, S. J. Bedathur and G. Weikum. Rank synopses for efficient time travel on the web graph. In P. S. Yu, V. J. Tsotras, E. A. Fox and B. Liu, eds., *ACM Fifteenth Conference on Information and Knowledge Management (CIKM2006),* Arlington, USA, 2006, pp. 864–865. ACM.

[30] K. Berberich, M. Vazirgiannis and G. Weikum. Time-aware authority ranking. *Internet Mathematics,* 2(3):301–332, 2006.

[31] S. Bickel and T. Scheffer. Dirichlet-enhanced spam filtering based on biased samples. In B. Schölkopf, J. Platt and T. Hoffman, eds., *Advances in Neural Information Processing Systems,* Cambridge, USA, 2007, vol. 19. MIT Press.

[32] B. Blanchet and A. Podelski. Verification of cryptographic protocols: Tagging enforces termination. *Theoretical Computer Science,* 333(1-2): 67–90, March 2005.

[33] V. Blanz, I. Albrecht, J. Haber and H.-P. Seidel. Creating face models from vague mental images. In L. Szirmay-Kalos and E. Gröller, eds., *Eurographics 2006 (EG'06),* Vienna, Austria, September 2006, Computer Graphics Forum, vol. 25, pp. 645–654. Blackwell.

[34] C. Bock, T. Lengauer, S. Reither, T. Mikeska, M. Paulsen and J. Walter. Biq analyzer: visualization and quality control for dna methylation data from bisulfite sequencing. *Bioinformatics,* 21(21):4067–4068, 2005.

[35] C. Bock, M. Paulsen, S. Tierling, T. Mikeska, T. Lengauer and J. Walter. Cpg island methylation in human lymphocytes is highly correlated with dna sequence, repeats and predicted dna structure. *PLoS Genetics,* 2(3):0243–0252, 2006.

[36] J. Bojunga, C. Welsch, I. Antes, M. Albrecht, T. Lengauer and S. Zeuzem. Structural and functional analysis of a novel mutation of CYP21B in a heterozygote carrier of 21-hydroxylase deficiency. *Human Genetics,* 117(6): 558–564, 2005.

[37] S. Castellano, A. V. Lobanov, C. Chapple, S. V. Novoselov, M. Albrecht, D. Hua, A. Lescure, T. Lengauer, A. Krol, V. N. Gladyshev and R. Guigó. Diversity and functional plasticity of eukaryotic selenoproteins: Identification and characterization of the SelJ family. *Proceedings of the National Academy of Sciences,* 102(45):16188–16193, 2005.

[38] S. Chaudhuri, G. Das, V. Hristidis and G. Weikum. Probabilistic information retrieval approach for ranking of database query results. *ACM Transactions on Database Systems,* 31(3): 1134–1168, September 2006.

[39] S. Chaudhuri, R. Ramakrishnan and G. Weikum. Integrating DB and IR technologies: What is the sound of one hand clapping? In M. Stonebraker, G. Weikum and D. DeWitt, eds., *Proceedings of the Second Biennial Conference on Innovative Data Systems Research (CIDR 05),* Asilomar, CA, USA, 2005, pp. 1–12. VLDB. Acceptance ratio 1:3.

[40] T. Chen, M. Goesele and H.-P. Seidel. Mesostructure from specularity. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006),* Los Alamitos, CA, USA, 2006, vol. 2, pp. 1825–1832. IEEE.

[41] B. S. Chlebus and D. Kowalski. Cooperative asynchronous update of shared memory. In H. N. Gabow and R. Fagin, eds., *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC 2005),* Baltimore, USA, 2005, pp. 733–739. ACM.

[42] M. Chrobak, C. Dürr, W. Jawor, L. Kowalik and M. Kurowski. A note on scheduling equallength jobs to maximize throughput. *Journal of Scheduling,* 9(1):71–73, 2006.

[43] B. Cook, A. Podelski and A. Rybalchenko. Abstraction-refinement for termination. In C. Hankin and I. Siveroni, eds., *Static analysis, 12th International Symposium, SAS 2005,* London, UK, September 2005, LNCS 3672, pp. 87–101. Springer.

[44] W. Damm, S. Disch, H. Hungar, J. Pang, F. Pigorsch, C. Scholl, U. Waldmann and B. Wirtz. Automatic verification of hybrid systems with large discrete state space. In S. Graf and W. Zhang, eds., *Automated Technology for Verification and Analysis, 4th International Symposium, ATVA 2006,* Beijing, China, 2006, LNCS 4218, pp. 276–291. Springer.

[45] H. de Nivelle. Translation of resolution proofs into short first-order proofs without choice axioms. *Information and Computation,* 199(1): 24–54, April 2005.

[46] H. de Nivelle and S. Demri. Deciding regular grammar logics with converse through first-order logic. *Journal of Logic, Language and Information,* 14(3):289–329, June 2005.

[47] H. de Nivelle and J. Meng. Geometric resolution: A proof procedure based on finite model search. In U. Furbach and N. Shankar, eds., *Automated reasoning: Third International Joint Conference, IJCAR 2006,* Seattle, WA, USA, August 2006, LNAI 4130, pp. 303–317. Springer.

[48] R. Dementiev, J. Kärkkäinen, J. Mehnert and P. Sanders. Better external memory suffix array construction. In C. Demetrescu, R. Sedgewick and R. Tamassia, eds., *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics (ALENEX/ANALCO 2005),* Vancouver, British Columbia, Canada, January 2005, pp. 86–97. SIAM.

[49] R. Dementiev, L. Kettner and P. Sanders. STXXL: Standard template library for XXL data sets. In G. S. Brodal and S. Leonardi, eds., *Algorithms - ESA 2005, 13th Annual European Symposium (ESA 2005),* Palma de Mallorca, Spain, October 2005, LNCS 3669, pp. 640–651. Springer.

[50] B. Doerr. Generating randomized roundings with cardinality constraints and derandomizations. In B. Durand and W. Thomas, eds., *STACS 2006, 23rd Annual Symposium on Theoretical Aspects of Computer Science,* Marseille, France, 2006, LNCS 3884, pp. 571–583. Springer.

[51] B. Doerr and T. Friedrich. Deterministic random walks on the two-dimensional grid. In T. Asano, ed., *Algorithms and Computation: 17th International Symposium, ISAAC 2006,* Kolkata, India, December 2006, LNCS 4288, pp. 474–483. Springer.

[52] B. Doerr, M. Gnewuch and A. Srivastav. Bounds and constructions for the star-discrepancy via delta-covers. *Journal of Complexity,* 21(5): 691–709, October 2005.

[53] B. Doerr, N. Hebbinghaus and F. Neumann. Speeding up evolutionary algorithms through restricted mutation operators. In T. P. Runarsson, H. G. Beyer, E. Burke, J. J. Merelo-Guervös, L. D. Whitley and X. Yao, eds., *Parallel Problem Solving from Nature - PPSN IX, 9th International Conference,* Reykjavik, Iceland, October 2006, LNCS 4193, pp. 978–987. Springer.

[54] B. Doerr and C. Klein. Unbiased rounding of rational matrices. In S. Arun-Kumar and N. Garg, eds., *FSTTCS 2006: Foundations of Software Technology and Theoretical Computer Science : 26th International Conference,* Kolkata, India, 2006, *LNCS 4337,* pp. 200–211. Springer.

[55] B. Doerr, J. Lengler and D. Steurer. The interval liar game. In T. Asano, ed., *Algorithms and Computation: 17th International Symposium, ISAAC 2006, Kolkata,* India, 2006, *LNCS 4288,* pp. 318–327. Springer.

[56] F. Eisenbrand, S. Funke, A. Karrenbauer, J. Reichel and E. Schömer. Packing a trunk - now with a twist! In S. N. Spencer, ed., *Proceedings SPM 2005 ACM Symposium on Solid and Physical Modeling,* Cambridge, USA, June 2005, pp. 197–206. ACM.

[57] K. Elbassioni. On the complexity of the multiplication method for monotone cnf/dnf dualization. In Y. Azar and T. Erlebach, eds., *Algorithms - ESA 2006, 14th Annual European Symposium,* Zurich, Switzerland, 2006, *LNCS 4168,* pp. 340–351. Springer.

[58] K. Elbassioni, A. Fishkin, N. H. Mustafa and R. Sitters. Approximation algorithms for euclidean group tsp. In L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi and M. Yung, eds., *Automata, languages and programming, 32nd International Colloquim, ICALP 2005,* Lisbon, Portugal, 2005, *LNCS 3580,* pp. 1115–1126. Springer.

[59] J. Freiheit and F. Zangl. Model-based user-interface management for public services. In D. Remenyi, ed., *6th European Conference on e-Government,* Marburg, Germany, April 2006, pp. 141–151. Academic Conferences Limited.

[60] M. Fuchs, V. Blanz, H. P. A. Lensch and H.-P. Seidel. Reflectance from images: A model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics,* 11(3): 296–305, May 2005.

[61] M. Fuchs, V. Blanz and H.-P. Seidel. Bayesian relighting. In O. Deussen, A. Keller, K. Bala, P. Dutré, D. W. Fellner and S. N. Spencer, eds., *Rendering Techniques 2005: Eurographics Symposium on Rendering,* Konstanz, Germany, July 2005, Rendering Techniques, pp. 157–164. Eurographics.

[62] S. Funke and C. Klein. Hole detection or: "how much geometry hides in connectivity?". In N. Amenta and O. Cheong, eds., *Proceedings of the 22nd Annual Symposium on Computational Geometry (SCG-06),* Sedona, Arizona, USA, 2006, pp. 377–385. ACM.

[63] H. Ganzinger and K. Korovin. Theory instantiation. In M. Hermann and A. Voronkov, eds., *Logic for Programming, Artificial Intelligence and Reasoning, 13th International Conference (LPAR'06),* Phnom Penh, Cambodia, 2006, *LNCS 4246,* pp. 497–511. Springer.

[64] H. Ganzinger, V. Sofronie-Stokkermans and U. Waldmann. Modular proof systems for partial functions with Evans equality. *Information and Computation,* 204(10):1453–1492, October 2006.

[65] H. Ganzinger and J. Stuber. Superposition with equivalence reasoning and delayed clause normal form transformation. *Information and Computation,* 199(1-2):3–23, 2005.

[66] J. Graupmann, R. Schenkel and G. Weikum. The SphereSearch engine for unified ranked retrieval of heterogeneous XML and web documents. In K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-A. Larson and B. C. Ooi, eds., *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005),* Trondheim, Norway, 2005, pp. 529–540. ACM. Acceptance ratio 1:6.

[67] J. Günther, H. Friedrich, H.-P. Seidel and P. Slusallek. Interactive ray tracing of skinned animations. In *Pacific Graphics 2006,* Taipei, Taiwan, September 2006, *The Visual Computer,* vol. 22, pp. 785–792. Springer.

[68] J. Günther, H. Friedrich, I. Wald, H.-P. Seidel and P. Slusallek. Ray tracing animated scenes using motion decomposition. In L. Szirmay-Kalos and E. Gröller, eds., *Eurographics 2006 (EG'06),* Vienna, Austria, September 2006, *Computer Graphics Forum,* vol. 25, pp. 517–525. Blackwell.

[69] J. Haber, M. Magnor and H.-P. Seidel. Physically based simulation of twilight phenomena. *Transactions on Graphics,* 24(4):1353–1373, October 2005.

[70] R. Hariharan, K. Telikepalli and K. Mehlhorn. A faster deterministic algorithm for minimum cycle bases in directed graphs. In M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Part I,* Venice, Italy, 2006, LNCS 4051, pp. 250–261. Springer.

[71] V. Havran, J. Bittner, R. Herzog, and H.-P. Seidel. Ray maps for global illumination. In O. Deussen, A. Keller, K. Bala, P. Dutré, D. W. Fellner, and S. N. Spencer, eds., *Rendering Techniques 2005: Eurographics Symposium on Rendering,* Konstanz, Germany, June 2005, pp. 43–54, 311. Eurographics.

[72] V. Havran, R. Herzog and H.-P. Seidel. Fast final gathering via reverse photon mapping. In M. Alexa and J. Marks, eds., *The European Association for Computer Graphics 26th Annual Conference, EUROGRAPHICS 2005,* Dublin, Ireland, August 2005, Computer Graphics Forum, vol. 24, pp. 323–333. Blackwell.

[73] V. Havran, M. Smyk, G. Krawczyk, K. Myszkowski and H.-P. Seidel. Interactive system for dynamic scene lighting using captured video environment maps. In O. Deussen, A. Keller, K. Bala, P. Dutré, D. W. Fellner, and S. N. Spencer, eds., *Rendering Techniques 2005: Eurographics Symposium on Rendering,* Konstanz, Germany, June 2005, pp. 31–42,311. Eurographics.

[74] J. Hoffmann. The deterministic part of ipc-4: An overview. *Journal of Artificial Intelligence Research,* 24:519 – 579, October 2005.

[75] J. Hoffmann. In defense of pddl axioms. *Artificial Intelligence,* 168(1-2):38–69, 2005.

[76] J. Hoffmann. Where ignoring delete lists works: Local search topology in planning benchmarks. *Journal of Artificial Intelligence Research,* 24:685–758, November 2005.

[77] J. Hoffmann and R. I. Brafman. Conformant planning via heuristic forward search: A new approach. *Artificial Intelligence,* 170:507–541, 2006.

[78] G. Ifrim, M. Theobald and G. Weikum. Learning word-to-concept mappings for automatic text classification. In L. De Raedt and S. Wrobel, eds., *Proceedings of the 22nd International Conference on Machine Learning - Learning in Web Search (LWS 2005),* Bonn, Germany, 2005, pp. 18–26. ICMLW4-LWS2005.

[79] G. Ifrim and G. Weikum. Transductive learning for text classification using explicit knowledge models. In J. Fürnkranz, T. Scheffer, and M. Spiliopoulou, eds., *PKDD 2006: 10th European Conference on Principles and Practice of Knowledge Discovery in Databases,* Berlin, Germany, 2006, *LNAI 4213,* pp. 223–234. Springer. Acceptance ratio 1:7.

[80] R. Irving, T. Kavitha, K. Mehlhorn, D. Michail and K. Paluch. Rank-maximal matchings. *ACM Transactions on Algorithms,* 2(4):602–610, October 2006.

[81] I. Ivrissimtzis, R. Zayer and H.-P. Seidel. *Polygonal decompositions of quadrilateral subdivision meshes. Computer Graphics & Geometry,* 7(1): 16–30, 2005.

[82] S. Jacobs and V. Sofronie-Stokkermans. Applications of hierarchical reasoning in the verification of complex systems. *Electronic Notes in Theoretical Computer Science,* 2007.

[83] S. Jacobs and U. Waldmann. Comparing instance generation methods for automated reasoning. In B. Beckert, ed., *Automated reasoning with analytic tableaux and related methods, International Conference, TABLEAUX 2005,* Koblenz, Germany, 2005, *LNCS 3702,* pp. 153–168. Springer.

[84] S. Jacobs and U.Waldmann. Comparing instance generation methods for automated reasoning. *Journal of Automated Reasoning,* 38:57–78, 2007.

[85] C. Jones, D. Lomet, A. Romanovsky, G. Weikum, A. Fekete, M.-C. Gaudel, H. F. Korth, R. de Lemos, J. E. B. Moss, R. Rajwar, K. Ramamritham, B. Randell and L. Rodrigues. The atomic manifesto: a story in four quarks. *SIGMOD* Record, 34(1):63–69, March 2005.

[86] C. Jones, D. Lomet, A. Romanovsky, G. Weikum, A. Fekete, M.-C. Gaudel, H. F. Korth, R. de Lemos, J. E. B. Moss, R. Rajwar, K. Ramamritham, B. Randell and L. Rodrigues. The atomic manifesto: a story in four quarks.

SIGOPS Operating Systems Review, 39(2):41–46, April 2005.

[87] C. B. Jones, D. B. Lomet, A. B. Romanovsky and G. Weikum. The atomic manifesto. *Journal of Universal Computer Science,* 11(5):636–651, May 2005.

[88] K. Kaligosi, K. Mehlhorn, J. I. Munro and P. Sanders. Towards optimal multiple selection. In L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi and M. Yung, eds., *Automata, languages and programming, 32nd International Colloquim, ICALP 2005,* Lisbon, Portugal, 2005, *LNCS 3580,* pp. 103–114. Springer.

[89] N. Kammenhuber, J. Luxenburger, A. Feldmann and G. Weikum. Web search clickstreams. In J. M. Almeida, V. A. F. Almeida and P. Barford, eds., *Proceedings of the 6th ACM SIGCOMM on Internet measurement (IMC '06),* Rio de Janeiro, Brazil, 2006, pp. 245–250. ACM.

[90] A. Kaster, S. Siersdorfer and G. Weikum. Combining text and linguistic document representations for authorship attribution. In S. Argamon, J. Karlgren and J. G. Shanahan, eds., *Workshop Stylistic Analysis of Text for Information Access, 28th International SIGIR,* Salvador, Brazil, August 2005, vol. 1, pp. 27–35. ACM.

[91] K. Korovin and A. Voronkov. Knuth-Bendix constraint solving is NP-complete. *ACM Transactions on Computational Logic,* 6(2):361–388, 2005.

[92] A. Kovács. Fast monotone 3-approxima-tion algorithm for scheduling related machines. In G. S. Brodal and S. Leonardi, eds., *Algorithms - ESA 2005: 13th Annual European Symposium,* Mallorca, Spain, 2005, *LNCS 3669,* pp. 616–627. Springer.

[93] G. Krawczyk, K. Myszkowski and H.-P. Seidel. Lightness perception in tone reproduc-tion for high dynamic range images. In M. Alexa and J. Marks, eds., *The European Association for Computer Graphics 26th Annual Conference, EUROGRAPHICS 2005,* Dublin, Ireland, 2005, *Computer Graphics Forum,* vol. 24, pp. 635–645. Blackwell.

[94] M. Kutz. Computing shortest non-trivial cycles on orientable surfaces of bounded genus in almost linear time. In *Proceedings of the 22nd Annual Symposium on Computational Geometry (SCG06), Sedona,* Arizona, USA, 2006, pp. 430–437. ACM.

[95] T. Langer, A. Belyaev and H.-P. Seidel. Spherical barycentric coordinates. In D. W. Fellner, S. N. Spencer, A. Sheffer and K. Polthier, eds., *SGP 2006 : Fourth Eurographics/ACM SIGGRAPH Symposium on Geometry Processing,* Cagliari, Sardinia, Italy, June 2006, pp. 81–88. Eurographics.

[96] T. Lengauer, ed. *Bioinformatics - from genomes to therapies 1 : the building blocks: molecular sequences and structures.* Wiley-VCH, Weinheim, Germany, 2007.

[97] T. Lengauer and T. Sing. Bioinformatics-assisted anti-hiv therapy. *Nature Reviews Microbiology,* 4:790–797, October 2006.

[98] A. Linari and G. Weikum. Efficient peer-to-peer semantic overlay networks based on statistical language models. In *P2PIR '06: Proceedings of the international workshop on Information retrieval in peer-to-peer networks,* Arlington, Virginia, USA, 2006, pp. 9–16. ACM.

[99] D. B. Lomet, R. S. Barga, M. F. Mokbel, G. Shegalov, R. Wang and Y. Zhu. Transaction time support inside a database engine. In L. Liu, A. Reuter, K.-Y. Whang and J. Zhang, eds., *Proceedings of the 22nd International Conference on Data Engineering (ICDE 2006),* Los Alamitos, USA, 2006, pp. 1–12. IEEE.

[100] J. Luxenburger and G. Weikum. Exploiting community behavior for enhanced link analysis and web search. In D. Zhou, ed., *Proceedings of the 9th International Workshop on the Web and Databases (WebDB 2006),* 2006, pp. 14–19.

[101] A. Malkis, A. Podelski and A. Rybalchenko. Thread-modular verification is cartesian abstract interpretation. In K. Barkaoui, A. Cavalcanti and A. Cerone, eds., *Theoretical aspects of computing - ICTAC 2006: Third International Colloquium,* Tunis, Tunisia, 2006, LNCS 4281, pp. 183–197. Springer.

[102] R. Mantiuk, A. Efremov, K. Myszkowski and H.-P. Seidel. Backward compatible high dynamic range mpeg video compression. In J. Dorsey, ed., *Proceedings of ACM SIGGRAPH 2006,* Boston, MA, USA, July 2006, *ACM Transactions on Graphics,* vol. 25, pp. 713–723. ACM. Proc. of ACM SIGGRAPH '06.

[103] R. Mantiuk, K. Myszkowski and H.-P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception,* 3(3):286–308, July 2006. This is a revised and extended version of the publication of the same title in the Proceedings of Second Symposium on Applied Perception in Graphics and Visualization 2005.

[104] D. Mavroeidis, G. Tsatsaronis, M. Vazirgiannis, M. Theobald and G. Weikum. Word sense disambiguation for exploiting hierarchical thesauri in text classification. In A. Jorge, L. Torgo, P. Brazdil, R. Camacho and G. Joao, eds., *Knowledge discovery in databases: PKDD 2005, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases,* Porto, Portugal, 2005, LNCS 3721, pp. 181–192. Springer.

[105] J. Maydt and T. Lengauer. Recco: recombination analysis using cost optimization. *Bioinformatics,* 22(9):1064–1071, February 2006.

[106] K. Mehlhorn, R. Osbild and M. Sagraloff. Reliable and efficient computational geo metry via controlled perturbation. In M. Bugliesi, B. Preneel, V. Sassone and I. Wegener, eds., *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Part I,* Venice, Italy, 2006, *LNCS 4051,* pp. 299–310. Springer.

[107] U. Meyer and N. Zeh. I/O-efficient undirected shortest paths with unbounded edge lengths. In Y. Azar and T. Erlebach, eds., *Algorithms - ESA 2006, 14th Annual European Symposium,* Zurich, Switzerland, September 2006, LNCS 4168, pp. 540–551. Springer.

[108] S. Michel, M. Bender, N. Ntarmos, P. Triantafillou, G. Weikum and C. Zimmer. Discovering and exploiting keyword and attri-bute-value co-occurrences to improve P2P routing indices. In P. S. Yu, V. J. Tsotras, E. A. Fox and B. Liu, eds., *ACM Fifteenth Conference on Information and Knowledge Management (CIKM2006),* Arlington, USA, 2006, pp. 172–181. ACM. Acceptance Ratio: 15

[109] S. Michel, M. Bender, P. Triantafillou and G. Weikum. IQN routing: Integrating quality and novelty in p2p querying and ranking. In Y. Ioannidis, M. H. Scholl, J. W. Schmidt, F. Matthes, M. Hatzopoulos, K. Boehm, A. Kemper, T. Grust and C. Boehm, eds., *Advances in Database Technology - EDBT 2006: 10th International Conference on Extending Database Technology,* Munich, Germany, March 2006, *LNCS 3896,* pp. 149–166. Springer.

[110] S. Michel, P. Triantafillou and G. Weikum. KLEE: A framework for distributed top-k query algorithms. In K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-A. Larson and B. C. Ooi, eds., *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005),* Trondheim, Norway, 2005, pp. 637–648. ACM.

[111] S. Michel, P. Triantafillou and G. Weikum. Minerva1: A scalable efficient peer-to-peer search engine. In G. Alonso, ed., *Middleware 2005, ACM, IFIP, USENIX 6th International Middleware Conference,* Grenoble, France, 2005, *LNCS 3790,* pp. 60–81. Springer.

[112] G. Moerkotte and T. Neumann. Analysis of two existing and one new dynamic programming algorithm for the generation of optimal bushy join trees without cross products. In U. Dayal, K.-Y. Whang, D. B. Lomet, G. Alonso, G. M. Lohman, M. L. Kersten, S. K. Cha and Y.-K. Kim, eds., *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006),* Seoul, Korea, September 2006, pp. 930–941. ACM.

[113] N. Ntarmos, P. Triantafillou and G. Weikum. Counting at large: Efficient cardinality estimation in internet-scale data networks. In L. Liu, A. Reuter, K.-Y. Whang and J. Zhang, eds., *Proceedings of the 22nd International Conference on Data Engineering (ICDE 2006),* Los Alamitos, USA, 2006, pp. 1–10. IEEE.

[114] G. Nürnberger, C. Rössl, F. Zeilfelder and H.-P. Seidel. Quasi-interpolation by quadratic piecewise polynomials in three variables. *Computer Aided Geometric Design,* 22:221–249, 2005.

[115] Y. Ohtake, A. Belyaev and M. Alexa. Sparse low-degree implicit surfaces with applications to high quality rendering, feature extraction and smoothing. In M. Desbrun and H. Pottman, eds., *Eurographics Symposium on Geometry Processing 2005,* Vienna, Austria, 2005, pp. 149–158. Eurographics.

[116] Y. Ohtake, A. Belyaev and H.-P. Seidel. A composite approach to meshing scattered data. *Graphical Models,* 68(3):255–267, 2006.

[117] Y. Ohtake, A. Belyaev and H.-P. Seidel. Sparse surface reconstruction with adaptive partition of unity and radial basis functions. *Graphical Models,* 68(1):15–24, January 2006.

[118] J. X. Parreira, D. Donato, S. Michel and G. Weikum. Efficient and decentralized pagerank

approximation in a peer-to-peer web search network. In U. Dayal, K.-Y. Whang, D. B. Lomet, G. Alonso, G. M. Lohman, M. L. Kersten, S. K. Cha and Y.-K. Kim, eds., *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006),* Seoul, Korea, 2006, pp. 415–426. ACM.

[119] J. X. Parreira, S. Michel and M. Bender. Size doesn't always matter: Exploiting pagerank for query routing in distributed ir. In *P2PIR '06: Proceedings of the international workshop on Information retrieval in peer-to-peer networks,* Arlington, USA, 2006, pp. 25–32. ACM.

[120] S. Pettie. Towards a final analysis of pairing heaps. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005),* Pittsburgh, USA, 2005, pp. 174–183. IEEE.

[121] A. Pnueli, A. Podelski and A. Rybalchenko. Separating fairness and well-foundedness for the analysis of fair discrete systems. In N. Halbwachs and L. Zuck, eds., *Tools and Algorithms for the Construction and Analysis of Systems: 11th International Conference, TACAS 2005, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2005,* Edinburgh, UK, April 2005, *LNCS 3440,* pp. 124–139. Springer.

[122] A. Podelski and A. Rybalchenko. Transition predicate abstraction and fair termination. In J. Palsberg and M. Abadi, eds., *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2005,* Long Beach, CA, USA, 2005, pp. 124–139. ACM.

[123] A. Podelski, I. Schaefer and S. Wagner. Summaries for while programs with recursion. In M. Sagiv, ed., *Programming Languages and Systems: 14th European Symposium on Programming, ESOP 2005, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2005,* Edinburgh, UK, April 2005, *LNCS 3444,* pp. 94–107. Springer.

[124] A. Podelski and T. Wies. Boolean heaps. In C. Hankin and I. Siveroni, eds., *Static analysis, 12th International Symposium, SAS 2005,* London, UK, September 2005, LNCS 3672, pp. 268–283. Springer.

[125] V. Prevosto and S. Boulmé. Proof contexts with late binding. In P. Urzyczyn, ed., *Typed Lambda Calculi and Applications: 7th International Conference, TLCA 2005,* Nara, Japan, April 2005, *LNCS 3461,* pp. 325–339. Springer. To appear.

[126] V. Prevosto and U. Waldmann. SPASS+T. In G. Sutcliffe, R. Schmidt and S. Schulz, eds., *ESCoR: FLoC'06 Workshop on Empirically Successful Computerized Reasoning,* Seattle, WA, USA, 2006, *CEUR Workshop Proceedings,* vol. 192, pp. 18–33.

[127] J. Rahnenführer, N. Beerenwinkel, W. A. Schulz, C. Hartmann, A. von Deimling, B. Wullich and T. Lengauer. Estimating cancer survival and clinical outcome based on genetic tumor progression scores. *Bioinformatics*, 21(10): 2438–2446, 2005.

[128] M. Ralser, M. Albrecht, U. Nonhoff, T. Lengauer, H. Lehrach and S. Krobitsch. An integrative approach to gain insights into the cellular function of human ataxin-2. *Journal of Molecular Biology,* 346(1):203–214, 2005.

[129] S. Ratschan. Efficient solving of quantified inequality constraints over the real numbers. *ACM Transactions on Computational Logic,* 7(4):723–748, 2006.

[130] S. Ratschan and Z. She. Safety verification of hybrid systems by constraint propagation based abstraction refinement. In M. Morari and L. Thiele, eds., *Hybrid Systems: Computation and Control: 8th International Workshop, HSCC 2005,* Zürich, Schweiz, 2005, *LNCS 3414,* pp. 573–589. Springer.

[131] K. Roberts, F.Mücklich and G.Weikum. Untersuchungen zur automatischen Klassifikation von Lamellengraphit mit Hilfe des Stützvektorverfahrens (Examinations on the Automatic Classification of Lamellar Graphite Using the Support Vector Machine). *Praktische Metallographie = Practical metallography: international journal on materialographic preparation, imaging and analysis of microstructures,* 42(8): 396–410, August 2005.

[132] B. Rosenhahn and H.-P. Seidel. A system for articulated tracking incorporating a clothing model. *Machine Vision and Applications,* 18(1):25–40, 2006.

[133] A. Rybalchenko and V. Sofronie-Stokkermans. Constraint solving for interpolation. In B. Cook and A. Podelski, eds., *8th International Conference on Verification, Model Checking and Abstract Interpretation (VMCAI 2007),* Nice, France, 2007, *LNCS 4349,* pp. 346–362. Springer.

[134] C. Sarrazin, U. Mihm, E. Herrmann, C. Welsch, M. Albrecht, U. Sarrazin, S. Traver, T. Lengauer and S. Zeuzem. Clinical significance of in vitro replication-enhancing mutations of the hepatitis C virus (hcv) replicon in patients with chronic HCV infection. *The Journal of Infectious Diseases,* 192(10):1710–1719, 2005.

[135] N. Sauber, H. Theisel and H.-P. Seidel. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. In E. Gröller, A. Pang, C. T. Silva, J. Stasko and J. van Wijk, eds., *IEEE Visualization Conference 2006,* Baltimore, USA, November 2006, *IEEE Visualization,* vol. 12, pp. 917–924. IEEE.

[136] R. Schenkel, A. Theobald and G. Weikum. Efficient creation and incremental maintenance of the HOPI index for complex XML document collections. In *Proceedings of the 21st International Conference on Data Engineering, ICDE 2005,* Tokyo, Japan, 2005, pp. 360–371. IEEE. Acceptance ratio 1:8.

[137] R. Schenkel, A. Theobald and G. Weikum. Semantic similarity search on semistructured data with the xxl search engine. *Information Retrieval,* 8(4):521–545, December 2005.

[138] R. Schenkel and M. Theobald. Feedback-driven structural query expansion for ranked retrieval of XML data. In Y. Ioannidis, M. H. Scholl, J. W. Schmidt, F. Matthes, M. Hatzopoulos, K. Boehm, A. Kemper, T. Grust and C. Boehm, eds., *Advances in Database Technology - EDBT 2006: 10th International Conference on Extending Database Technology*, Munich, Germany, 2006, *LNCS 3896,* pp. 331–348. Springer. Acceptance ratio 1:6.

[139] R. Schenkel and M. Theobald. Structural feedback for keyword-based XML retrieval. In M. Lalmas, A. MacFarlane, S. M. Rüger, A. Tombros, T. Tsikrika and A. Yavlinsky, eds., *Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006,* London, UK, 2006, *LNCS 3936,* pp. 326–337. Springer.

[140] A. Schlicker, F. S. Domingues, J. Rahnenführer and T. Lengauer. A new measure for functional similarity of gene products based on gene ontology. *BMC Bioinformatics,* 7:1–16, June 2006.

[141] G. Schlosser, J. Hesser, F. Zeilfelder, C. Rössl, R. Männer, G. Nürnberger and H.-P. Seidel. Fast visualization by shear-warp on quadratic super-spline models using wavelet data decompositions. In C. T. Silva, E. Gröller and H. Rushmeier, eds., *16th IEEE Visualization Conference (VIS 2005),* Minneapolis, MN, USA, 2005, pp. 351–358. IEEE.

[142] S. Schreiber, P. Rosenstiel, M. Albrecht, J. Hampe and M. Krawczak. Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nature Reviews Genetics,* 6(5): 376–388, 2005.

[143] G. Shegalov and G. Weikum. EOS2: Unstoppable stateful PHP. In U. Dayal, K.-Y. Whang, D. B. Lomet, G. Alonso, G. M. Lohman, M. L. Kersten, S. K. Cha and Y.-K. Kim, eds., *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB 2006),* Seoul, Korea, 2006, pp. 1223–1226. ACM.

[144] G. Shegalov, G. Weikum and K. Berberich. Unstoppable stateful PHP web services. In K. Aberer, Z. Peng, E. A. Rundensteiner, Y. Zhang and X. Li, eds., *Web Information Systems - WISE 2006, 7th International Conference on Web Information Systems Engineering,* Wuhan, China, 2006, LNCS 4255, pp. 132–143. Springer.

[145] S. Siersdorfer and S. Sizov. Automatic document organization in a P2P environment. In M. Lalmas, A. MacFarlane, S. M. Rüger, A. Tombros, T. Tsikrika and A. Yavlinsky, eds., *Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006,* London, UK, 2006, *LNCS 3936,* pp. 265–276. Springer.

[146] S. Siersdorfer and G. Weikum. Automated retraining methods for document classification and their parameter tuning. In A. H. H. Ngu, M. Kitsuregawa, E. J. Neuhold, J.-Y. Chung and Q. Z. Sheng, eds., *Web information systems engineering - WISE 2005, 6th International Conference on Web Information Systems Engineering,* New York, USA, 2005, *LNCS 3806,* pp. 478–486. Springer.

[147] S. Siersdorfer and G. Weikum. Using restrictive classification and meta classification for junk elimination. In D. Losada and J. M. Fernandez-Luna, eds., *Advances in information retrieval, 27th European Conference on IR Research, ECIR 2005,* Santiago de Compostela, Spain, March 2005, *LNCS 3408,* pp. 287–299. Springer. Acceptance ratio 1:4.

[148] T. Sing and N. Beerenwinkel. Mutagenetic tree fisher kernel improves prediction of hiv drug resistance from viral genotype. In S. B, P. J. and H. T., eds., *Advances in Neural Information*

Processing Systems 19, Vancouver, B.C., Canada, 2007, pp. 1–9. MIT.

[149] T. Sing, O. Sander, N. Beerenwinkel and T. Lengauer. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20):3940–3941, October 2005.

[150] K. Smith, G. Krawczyk, K. Myszkowski and H.-P. Seidel. Beyond tone mapping: Enhanced depiction of tone mapped HDR images. In L. Szirmay-Kalos and E. Gröller, eds., *Eurographics 2006 (EG'06),* Vienna, Austria, September 2006, *Computer Graphics Forum,* vol. 25, pp. 427– 438. Blackwell.

[151] V. Sofronie-Stokkermans. Hierarchic reasoning in local theory extensions. In R. Nieuwenhuis, ed., *Automated deduction - CADE-20, 20th International Conference on Automated Deduction,* Tallinn, Estonia, 2005, *LNAI 3632,* pp. 219–234. Springer.

[152] V. Sofronie-Stokkermans. Interpolation in local theory extensions. In U. Furbach and N. Shankar, eds., *Proceedings of IJCAR 2006,* Seattle, USA, 2006, *LNAI 4130,* pp. 235–250. Springer.

[153] I. Sommer, S. Toppo, O. Sander, T. Lengauer and S. Tosatto. Improving the quality of protein structure models by selecting from alignment alternatives. *BMC Bioinformatics,* 7:1–11, July 2006.

[154] A. Steffen, A. Kämper and T. Lengauer. Flexible docking of ligands into synthetic receptors using a two-sided incremental construction algorithm. *Journal of Chemical Information and Modeling,* 46(4):1695–1703, July 2006.

[155] C. Stoll, S. Gumhold and H.-P. Seidel. Visualization with stylized line primitives. In C. T. Silva, E. Gröller and H. Rushmeier, eds., *IEEE Visualization 2005 (VIS 2005),* Minneapolis, USA, 2005, pp. 695–702. IEEE.

[156] C. Stoll, Z. Karni and H.-P. Seidel. Geodesics guided constrained texture deformation. In *The 14th Pacific Conference on Computer Graphics and Applications Proceedings,* Taipei, Taiwan, October 2006, *Pacific Conference on Computer Graphics and Applications Proceedings,* vol. 14, pp. 144–152.

[157] F. M. Suchanek and P. Baumgartner. Automated reasoning support for first order ontologies. In J. J. Alferes, J. Bailey, W. May and U. Schwertel, eds., *Principles and Practice of Semantic Web Reasoning, 4th International Workshop, PPSWR 2006,* Budva, Montenegro, 2006, *LNCS 4187,* pp. 18–32. Springer. Acceptance Ratio 1:5.

[158] F. M. Suchanek, G. Ifrim and G. Weikum. Combining linguistic and statistical analysis to extract relations from web documents. In T. Eliassi-Rad, L. Ungar, M. Craven and D. Gunopulos, eds., *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006),* Philadelphia, PA, USA, 2006, pp. 712–717. ACM. Acceptance Ratio 1:5.

[159] M. Tarini, H. P. A. Lensch, M. Goesele and H.-P. Seidel. 3d acquisition of mirroring objects. *Graphical Models,* 67(4):233–259, July 2005.

[160] K. Telikepalli and K. Mehlhorn. A polynomial time algorithm for minimum cycle basis in directed graphs. In V. Diekert and B. Durand, eds., *STACS 2005, 22nd Annual Symposium on Theoretical Aspects of Computer Science,* Stuttgart, Germany, 2005, *LNCS 3404,* pp. 654–665. Springer.

[161] H. Theisel, J. Sahner, T. Weinkauf, H.-C. Hege and H.-P. Seidel. Extraction of parallel vector surfaces in 3d time-dependent fields and applications to vortex core line tracking. In C. T. Silva, E. Gröller and H. Rushmeier, eds., *IEEE Visualization 2005 (VIS 2005),* Minneapolis, USA, 2005, pp. 631–638. IEEE.

[162] H. Theisel, T. Weinkauf, H.-C. Hege and H.-P. Seidel. Topological methods for 2d timedependent vector fields based on stream lines and path lines. *IEEE Transactions on Visualization and Computer Graphics,* 11(4):383–394, May 2005.

[163] M. Theobald, R. Schenkel and G. Weikum. Efficient and self-tuning incremental query expansion for top-k query processing. In R. A. Baeza-Yates, N. Ziviani, G. Marchionini, A. Moffat and J. Tait, eds., *28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005),* Salvador, Brazil, 2005, pp. 242–249. ACM. Acceptance ratio 1:5.

[164] M. Theobald, R. Schenkel and G. Weikum. An efficient and versatile query engine for topX search. In K. Böhm, C. S. Jensen, L. M. Haas, M. L. Kersten, P.-A. Larson and B. C. Ooi, eds., *Proceedings of the 31st International Conference on Very Large Data Bases (VLDB 2005),* Trondheim, Norway, 2005, pp. 625–636. ACM. Acceptance ratio 1:6.

[165] S. Ulmschneider, U. Mueller-Vieira, M. Mitrenga, R. W. Hartmann, S. Oberwinkler-Marchais, C. D. Klein, M. Bureik, R. Bernhardt, I. Antes and T. Lengauer. Synthesis and evaluation of imidazolylmethylenetetrahydronaphthalenes and imidazolylmethyleneindanes: Potent inhibitors of aldosterone synthase. *Journal of Medicinal Chemistry,* 48(6): 1796–1805, April 2005.

[166] R. Valentonyte, J. Hampe, K. Huse, P. Rosenstiel, M. Albrecht, A. Stenzel, M. Nagy, K. I. Gaede, A. Franke, R. Haesler, A. Koch, T. Lengauer, D. Seegert, N. Reiling, S. Ehlers, E. Schwinger, M. Platzer, M. Krawczak, J. Müller-Quernheim, M. Schürmann and S. Schreiber. Sarcoidosis is associated with a truncating splice site mutation in BTNL2. *Nature Genetics,* 37(4):357–364, 2005.

[167] M. M. Van Duist, M. Albrecht, M. Podswiadek, D. Giachino, T. Lengauer, L. Punzi and M. De Marchi. A new CARD15 mutation in Blau syndrome. *European Journal of Human Genetics,* 13(6):742–747, 2005.

[168] W. von Funck, H. Theisel and H.-P. Seidel. Vector field based shape deformations. In *Proceedings of ACM SIGGRAPH 2006,* Boston, MA, USA, July 2006, *ACM Transactions on Graphics,* vol. 25, pp. 1118–1125. ACM. Proc. of ACM SIGGRAPH '06.

[169] S. Wemmert, R. Ketter, J. Rahnenführer, N. Beerenwinkel, M. Strowitzki, W. Feiden, C. Hartmann, T. Lengauer, F. Stockhammer, K. D. Zang, E. Meese, W.-I. Steudel, A. von Deimling and S. Urbschat. Patients with high grade gliomas harboring delections of chromosomes 9p and 10q benefit from temozolomide treatment. *Neoplasia,* 7(10):883–893, 2005.

[170] T. Wies, V. Kuncak, P. Lam, A. Podelski and M. C. Rinard. Field constraint analysis. In E. A. Emerson and K. S. Namjoshi, eds., *Verification, Model Checking and Abstract Interpretation, 7th International Conference, VMCAI 2006,* Charleston, SC, USA, January 2006, *LNCS 3855,* pp. 157–173. Springer.

[171] H. Yamauchi, S. Gumhold, R. Zayer and H.-P. Seidel. Mesh segmentation driven by gaussian curvature. *The Visual Computer,* 21(8-10):649–658, September 2005.

[172] H. Yamauchi, H. P. A. Lensch, J. Haber and H.-P. Seidel. Textures revisited. *The Visual Computer,* 21(4):217–241, May 2005.

[173] J. Yin, N. Beerenwinkel, J. Rahnenführer and T. Lengauer. Model selection for mixtures of mutagenetic trees. *Statistical applications in genetics and molecular biology,* 5(1): Article 17, 2006.

[174] A. Yoshida, R. Mantiuk, K. Myszkowski and H.-P. Seidel. Analysis of reproducing real-world appearance on displays of varying dynamic range. In L. Szirmay-Kalos and E. Gröller, eds., *Eurographics 2006 (EG'06),* Vienna, Austria, September 2006, *Computer Graphics Forum,* vol. 25, pp. 415–426. Blackwell.

[175] S. Yoshizawa, A. Belyaev and H.-P. Seidel. A moving mesh approach to stretch-minimizing mesh parameterization. *International Journal of Shape Modeling,* 11(1):25–42, June 2005.

[176] R. Zayer, C. Rössl, Z. Karni and H.-P. Seidel. Harmonic guidance for surface deformation. In M. Alexa and J. Marks, eds., *The European Association for Computer Graphics 26th Annual Conference, EUROGRAPHICS 2005,* Dublin, Ireland, 2005, *Computer Graphics Forum,* vol. 24, pp. 601–609. Blackwell.

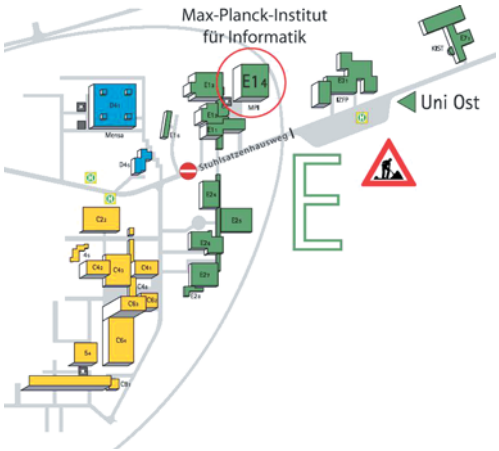
[177] R. Zayer, C. Rössl and H.-P. Seidel. Setting the boundary free: A composite approach to surface parameterization. In M. Desbrun and H. Pottmann, eds., *Symposium on Geometry Processing,* Vienna, Austria, 2005, pp. 91–100. Eurographics/ACM.

[178] H. Zhu, F. S. Domingues, I. Sommer and T. Lengauer. NOXclass: prediction of protein-protein interaction types. *BMC Bioinformatics,* 7:1–15, June 2006.

Wege zum Institut

Das Max-Planck-Institut für Informatik (Gebäude E 1.4) befindet sich auf dem Campus der Universität des Saarlandes etwa 5 km nordöstlich vom Zentrum der Stadt Saarbrücken im Wald nahe bei Dudweiler.

Saarbrücken besitzt einen eigenen Flughafen (Saarbrücken-Ensheim) und ist mit Auto-, Shuttle-Bus- und Zugverbindungen an die Flughäfen Frankfurt und Luxemburg angebunden. Zugstrecken verbinden Saarbrücken im Stundentakt innerhalb Deutschlands. Regelmäßig verkehrende Züge schaffen eine Anbindung an die Städte Metz, Nancy und Paris. Autobahnen führen nach Mannheim/Frankfurt, Luxembourg/Trier/Köln, Strasbourg und Metz/Nancy/Paris.



Sie erreichen den Campus...

von Saarbrücken-Ensheim, Flughafen

mit dem Taxi in ungefähr 20 Minuten

von Saarbrücken, Hauptbahnhof

mit dem Taxi in ungefähr 15 Minuten
mit dem Bus in ungefähr 20 Minuten
Richtung *„Dudweiler-Dudoplatz“* oder *„Universität Campus“*
Ausstieg *„Universität Mensa“* (wegen der Bauarbeiten voraussichtlich erst 2008 wieder möglich)
alternativ Ausstieg *„Universität Campus“*

von Frankfurt oder Mannheim über die Autobahn A6

Abfahrt *„St.Ingbert-West“*
den weißen Schildern *„Universität“* zum Campus folgend

von Paris über die Autobahn A4

Abfahrt *„St.Ingbert-West“*
den weißen Schildern *„Universität“* zum Campus folgend

Wenn Sie mit dem Auto anreisen: Nutzen Sie bitte ausschließlich die Zufahrt *„Universität Ost“* bis voraussichtlich Ende Juli 2007

IMPRESSUM

Herausgeber
Max-Planck-Institut für Informatik
Stuhlsatzenhausweg 85
D-66123 Saarbrücken

Redaktion & Koordination
Benjamin Doerr
Jörn Freiheit
Manuel Lamotte
Karol Myzskowski
Tobias Scheffer
Ralf Schenkel
Ingolf Sommer
Uwe Waldmann
Christoph Weidenbach
Roxane Wetzel

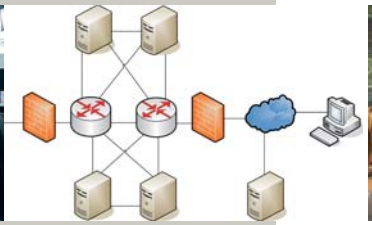
Kontakt
Max-Planck-Institut für Informatik
Telefon +49 681 9325-0
Telefax +49 681 9325-999
Email info@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de

Berichtszeitraum
1. Januar 2005 bis 31. Dezember 2006

Gestaltung
Behr Design | Saarbrücken

Druck
Bliesdruckerei | Blieskastel

max planck institut
informatik


Max-Planck-Institut für Informatik
Stuhlsatzenhausweg 85
D-66123 Saarbrücken

Telefon +49 681 9325-0
Telefax +49 681 9325-999
Email info@mpi-inf.mpg.de
Internet http://www.mpi-inf.mpg.de

max planck institut
informatik